Educational Evaluation and Policy Analysis
March 2016, Vol. 38, No. 1, pp. 88–112
DOI: 10.3102/0162373715587965

Head Start at Ages 3 and 4 Versus Head Start Followed by State Pre-K: Which Is More Effective?

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

As policymakers contemplate expanding preschool opportunities for low-income children, one possibility is to fund 2, rather than 1 year of Head Start for children at ages 3 and 4. Another option is to offer 1 year of Head Start followed by 1 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 prereading 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 prewriting skills or premath skills. This suggests that some impacts of early learning programs may be related to the sequencing of learning experiences to more academic programming.

Keywords: *Head Start, pre-kindergarten, regression discontinuity, propensity scores, early childhood education policy*

Introduction

In light of 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 well-being 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% in 1980 to 40% in 2007, and comprising 63% of first-time Head Start children in 2010 (Aikens, Klein, Tarullo, & West, 2013; Tarullo, Aikens, Moiduddin, & West, 2010).

Expanding ECE programs to include younger children would increase the number of children participating in programs for multiple years. In fact, over half of all 3-year-old entrants now go on to complete 2 years of Head Start (Aikens et al., 2013). Others transition from Head Start at age 3 to state-created and implemented, academically focused pre-kindergarten (pre-K) programs at age 4. 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 2 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 less from a second year in the program relative to 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 participate in a universal pre-K program at age 4? We use data from the study of the Oklahoma pre-K program (OK pre-K) to compare outcomes for two different preschool "pathways" to kindergarten (Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips, Adelstein, & Shaw, 2010; Gormley, Phillips, & Gayer, 2008). 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 (RD) design with a strict age eligibility cutoff for program participation to estimate the effect of these pathways on children's early academic skills at kindergarten. We apply propensity score (PS) weighting to the analyses to address selection into pathways and compare their effects on child outcomes.

This study extends prior findings from these data in several ways. For academic outcomes, Gormley and colleagues estimated two separate RD specifications—one for OK pre-K and one for Head Start—calculated treatment effect sizes, and compared effect sizes descriptively (Gormley, 2008; Gormley & Gayer, 2005; Gormley et al., 2005). One study compares two separately generated RD effect sizes using a basic significance test (a difference in z scores; Gormley et al., 2010; Paternoster, Brame, Mazerolle, & Piquero, 1998). In contrast, this study focuses on comparing the effectiveness of attending OK pre-K and Head Start at age 4 among age 3 Head Start graduates only after pooling both pre-K and Head Start children into the same RD model, addressing differential selection into the programs. As such, this 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, under key assumptions.

We find among children attending Head Start at age 3 that 1 year of Head Start as a 3-year-old followed by OK pre-K at age 4 has better early reading outcomes at kindergarten compared with children who stayed in Head Start at both age 3 and age 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 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 & Phillips, 2008 for reviews).

Because of its use of random assignment, the experimental Head Start Impact Study provides the best evidence on the short-term impacts of Head Start on children's language, literacy, and early writing skills at ages 3 and 4. The end-of-program-year effect sizes average 0.2 *SD* for both the age 3 and age 4 cohorts on early language and literacy skills, and a 0.15-*SD* effect size on early math skills for age 3 cohort participants (Puma, Bell, Cook, & Heid, 2010). Even though short-term gains appear to "fade-out," Ludwig and Phillips (2008) show that the short-term intent-to-treat effects are large enough for Head Start to pass a cost–benefit test. They calculate larger treatment-on-the-treated estimates for some key outcomes (e.g., letter-word [LW] 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*). Strong quasi-experimental evidence on the effects of Head Start shows long-term benefits on academic outcomes, with effect sizes of 0.2 to 0.3 *SD*s (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 with a 1-year Head Start, 1-year pre-K experience.

Pre-K. Pre-K programs are funded locally (i.e., typically by the state) to provide a year or two of education prior to kindergarten for children ages 3 or 4. Nationally, 28% of all 4-year-olds were enrolled in state-funded pre-K across 40 states in 2010 compared with 11% 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 each state creates its pre-K programs independently, and thus, the characteristics of these program vary widely across states (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; Mashburn et al., 2008; Phillips, Gormley, & Lowenstein, 2009; Pianta et al., 2005; Wong, Cook, Barnett, & Jung, 2008). For these reasons, the effects of any particular pre-K program cannot be generalized to state pre-K programs nationwide.

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, which was also confirmed by an RD analysis (Lipsey, Farran, Bilbrey, Hofer, & Dong, 2011). Oklahoma and Boston's pre-K evaluations also use RD designs based on a strict age eligibility cutoff and found large short-term improvements in early reading, writing, math skills, and executive function (ES range = .99–.36; Gormley, 2008; Gormley & Gayer, 2005; Gormley et al., 2005; Weiland & Yoshikawa, 2013). Using a similar RD design, studies of pre-K programs in Arkansas (Hustedt, Barnett, & Jung, 2008) and a five-state pre-K comparison found positive effects for early reading, literacy, and math skills (ES range = .23–.96; Wong et al., 2008).

Other studies of the effects of pre-K programs have used PS methods, finding 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.

Comparing the Effects of Two Types of Programs: Head Start and Pre-K

An important distinction between Head Start and pre-K are the program goals. Head Start mandates a "whole-child" approach that aims to comprehensively support children's development across several outcome domains, whereas pre-K programs—particularly Tulsa's program—often focus on children's early academic skills to prepare children for the academic nature of kindergarten. These differences may result in differential program effects across the broad scope of children's outcomes.

Despite the large body of research on the effectiveness of individual types of ECE programs in improving children's early academic skills, relatively few studies have directly compared the effectiveness of Head Start and different state pre-K programs. Henry and colleagues (2006) use PS matching to address selection and compare Head Start with Georgia's pre-K program, finding that state pre-K participants had statistically significant but only 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 than for Head Start. The effects of Head Start and pre-K vary depending on the comparison treatment condition (Ludwig & Phillips, 2008). Zhai, Brooks-Gunn, and Waldfogel (2011) 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. However, when compared with 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 universal 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 2 years of Head Start would be more beneficial for children than 1 year of Head Start followed by 1 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 evidence from experimental and nonexperimental 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 incremental 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, 2014; 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 produced substantial effects in only 1 year of services (Gormley et al., 2005).

The Head Start duration research is equivocal, with some indication that 2 years are more advantageous than 1, 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 Head Start Family and Child Experiences Survey (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 Head Start participation (ES = 0.27–0.80; 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 over 1 year.

However, 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 2-year participants. Barnett and Lamy (2006) also find no influence of duration in a pre-K program on print awareness and math, with some small effects for vocabulary. Nores and Barnett (2010) 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 *SD*s, 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.

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

Pre-K and Head Start program models differ in several ways. Our study cannot examine which of these components may make a difference in children's outcomes because they are confounded with program type. However, two noteworthy differences are curricula and classroom peer composition.

Curricula. As a part of the Tulsa pre-K study, Phillips et al. (2009) examined classroom characteristics in pre-K and Head Start. A key finding from their study was that the quality ratings for *both* programs were in the good-to-high range based on standard observational measures, higher than the national averages of both program types (Dotterer, Burchinal, Bryant, Early, & Pianta, 2012; Moiduddin, Aikens, Tarullo, West, & Xue, 2012). The only differences that emerged between the two programs were the curricula teachers reported using. Thus, curricula and related instructional practices may be an important distinction between the two programs.

In addition to differences in curricular approaches, the extent to which the curriculum used in Head Start classrooms *differentiates* children's age 3 and age 4 learning experiences would influence both the Head Start dosage effect and the comparative effect of Head Start to OK pre-K (Yoshikawa et al., 2013). 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). Indeed, recent work suggests that kindergarten teachers spending time on math skills students have already mastered has a negative effect on student's math achievement (Engel, Claessens, & Finch, 2013).

We know relatively little about whether Head Start curricula are hierarchical in practice and evolve as children age, because of the variation in curricula and limited support of their efficacy. The Head Start program mandates that program curricula focus on the "whole child," where learning occurs through participating in activities. 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 whole-child 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); Creative Curriculum was also used in the OK pre-K program although the most common curriculum reported by teachers was integrated thematic instruction (Phillips et al., 2009).

Surprisingly, there is little empirical support for High/Scope, none for Creative Curriculum, and neither curriculum—as currently used—has demonstrated effectiveness based on rigorous statistical standards (U.S. Department of Education, 2013). In addition, most ECE practitioners are convinced that whole-child instruction through discovery learning is best for young children based on theoretical models such as Piaget, but limited evidence supports this assumption. Indeed, recent evidence from the Boston Pre-K evaluation suggests the opposite. Boston's highly effective pre-K program uses several domain-specific curricula that focus on presenting lessons that become increasingly complex and build on the inherent hierarchy of skills within that domain (Klein, Starkey, Clements, Sarama, & Iyer, 2008; Weiland & Yoshikawa, 2013). Results from recent studies also indicate that children who receive targeted or content-specific curricula (e.g., literacy or math) during preschool show moderate to large improvements in the targeted content domain (e.g., Clements & Sarama, 2008; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011). Curricula effectiveness also depends on the extent that teachers implement them with fidelity.

This variation in curricula, their limited efficacy, and the unknown degree to which learning activities change as children age highlight the ambiguity of the impact of the second-year Head Start experience. As explained by Reynolds (1995) 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" (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 whole-child curriculum. Curricula packages—including Creative and High/Scope—provide curricular supports to individualize instruction for children within a classroom, but it is unclear whether teachers *use* these resources and adjust their instruction accordingly, especially in mixed-age settings. 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. Head Start programs are available to very low-income 3- and 4-year-old children, whereas the OK pre-K program is universally available to 4-year-old children only, but regardless of income. These two program features create differences in both the distribution of children's ages and the distribution of family income in the classroom, either of which can influence children's outcomes through peer effects.

For practical reasons, Head Start classrooms often combine 3- and 4-year-olds. While child development and educational theorists have supported the use of mixed-age classrooms (Bandura, 1986; Katz, 1990; Montessori, 1917; Vygotsky, 1978), the empirical research in this area is equivocal; some studies show limited positive effects (Blasco, Bailey, & Burchinal, 1993; Urberg & Kaplan, 1986), but several studies find null or negative effects of mixed-age settings (Bailey, McWilliam, Ware, & Burchinal, 1993; Bell, Greenfield, & Bulotsky-Shearer, 2013; Hattie, 2002; Moller, Forbes-Jones, & Hightower, 2008; Winsler et al., 2002).

The more important feature of mixed-age classrooms may be that a 1-year age difference during early childhood can create substantial variation in the classroom's distribution of children's skills. In turn, the skill level of classroom peers can substantially affect children's skill development because teacher-directed activities are often kept to a minimum in ECE. Henry and Rickman (2007) study peer effects in preschool children and find that having peers with higher cognitive skills produced positive effects on children's early math, literacy, and language skills. Others find beneficial peer effects for not only preschool children with low baseline skills (Justice, Petscher, Schatschneider, & Mashburn, 2011) but also preschool children with high baseline skills (Mashburn, Justice, Downer, & Pianta, 2009). Studies also suggest positive peer effects on math and reading achievement for school-age children (Cascio & Schanzenbach, 2012; Chetty et al., 2011; Elder & Lubotsky, 2009; Hanushek, Kain, Markman, & Rivkin, 2003; Hoxby & Weingarth, 2005; 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 first-time Head Start age 4 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; Hoxby & Weingarth, 2005; Lavy, Paserman, & Schlosser, 2012; Moller et al., 2008). In this situation, second year Head Start students who entered at age 3 provide positive peer effects for children entering at age 4 but derive no personal benefit from peer effects. Both mechanisms would reduce the added benefits of children's second year in Head Start.

On the contrary, 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 may be more mixed in terms of children's socioeconomic backgrounds (Reid & Ready, 2013). 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 entering from age-3 HS (Barnett & Belfield, 2006; Hart & Risley, 1995; Henry et al., 2006; Rouse, Brooks-Gunn, & McLanahan, 2005; Schechter & Bye, 2007; Zimmer & Toma, 2000). Still, it is possible that universal pre-K classrooms in economically segregated neighborhoods are not actually socioeconomically diverse (Dotterer et al., 2012).

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 data set, we are not able to estimate the effects of peers in an empirical model, and Phillips et al. (2009) did not explore classroom peer composition in their study of OK pre-K classroom characteristics. However, we do describe some of the conditions likely determining peer effects.

On balance, we judge that prior findings and the likely direction of curricular and peer effects argue that age 3 Head Start graduates will have stronger early 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 because 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.

Method

Research Design and Analysis

Our research question is as follows:

Research Question 1: If children participate in Head Start at age 3, do they have better early 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 treatments. We estimated treatment effects using a RD model. We applied PS weighting to the RD model to 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 Online Appendix 1 and supplemental figures and calculations in Online Appendix 2 (available at http://epa.sagepub.com/supplemental).

Data

Participants. The evaluation focused on the children enrolled in the Tulsa pre-K programs in 2006–2007, using the data from the Tulsa Preschool Study 2006–2007 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 who attended Head Start at age 3 (n = 540). Among these children, the analysis data set includes students who were entering OK pre-K, age 4 Head Start, or OK public school kindergarten in the 2006–2007 school year. The two preschool pathways we created and their sample sizes are (a) 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 (b) participants in Head Start at age 4 *and* age 3 (329 total; 119 kindergarten entrants, 210 Head Start entrants). Ninety-two percent of the OK pre-K children in our sample attended full-day pre-K (6.5 hours) making these participants as similar as possible to Head Start participants, which was a full-day program in Tulsa. Child and family characteristics for both groups are presented in columns 1 and 2 of Table 1.

We also examined whether our analytic sample was representative of the Tulsa kindergarten population. In Online Appendix 2.1, available at http://epa.sagepub.com/supplemental, we present descriptive statistics for kindergarten children who attended OK pre-K or Head Start and other Tulsa kindergarten children in the Tulsa pre-K study file. This table reveals that in general, children attending one of the public preschool programs are more disadvantaged than their nonparticipating peers. They are more likely to be low income, Black, to speak a language other than English in the home, are less likely to have Internet access at home, and to have parents who are married.

Measures. Child academic assessments occurred in August 2006 and included three academic subtests from the Woodcock–Johnson Achievement Tests–III (WJ; Woodcock, McGrew, & Mather, 2001). The LW Identification subtest measures early 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 early writing and spelling skills. The Applied Problems test has children perform simple calculations to solve math problems, which assesses children's early mathematical thinking with respect to counting, cardinality, and early operational skills. The reliability coefficient for the 3- to 5-year-old age group ranges from .97 to .99 (Woodcock et al., 2001). The same subtests of a comparable Spanish test, the Woodcock-Muñoz Batería, were given to Hispanic students capable of being tested in Spanish. The assessment values are in raw scores and are not nationally normed. Further detail regarding the sample, procedures, measurement, and assessments are available in Gormley et al. (2005).

1. Estimating Treatment Effects: RD Design

Our study implements an RD design, a method designed to provide unbiased estimates of 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 2005–2006 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). Therefore in our analysis, child age—measured as distance between their birth date and the cutoff birth date in days—is the assignment variable for the RD specification. This particular RD design is referred to as an "age-cutoff" RD and has been widely adopted for studying the effects of public pre-K programs (Lipsey, Weiland, Yoshikawa, Wilson, & Hofer, 2014; Wong et al., 2008). Figure 1a shows the discontinuity in treatment status by age for the age 4 OK pre-K and age 4 Head Start groups, and 1b plots the density of children near the cutoff for both pathways combined.

TABLE 1

Covariate Balance Between Children Who Attended Age 3 HS + Age 4 OK Pre-K and Age 3 HS + Age 4 HS in Observed Data and in PS 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	Diff
Covariates					
Reduced-price lunch	0.10	0.03	0.06	0.06	0.00
White	0.10	0.09	0.10	0.10	0.00
Black	0.64	0.44	0.54	0.52	0.02
Hispanic	0.17	0.39	0.27	0.30	0.03
Asian/Native/Other	0.08	0.07	0.08	0.07	0.01
Female	0.50	0.55	0.52	0.52	0.00
Below high school	0.08	0.19	0.12	0.15	0.03
High school	0.30	0.28	0.31	0.29	0.02
Some college	0.32	0.32	0.31	0.32	0.10
College+	0.09	0.05	0.07	0.07	0.00
Child had some nonparental care at age 3	0.55	0.46	0.52	0.50	0.02
Internet access in home	0.33	0.29	0.30	0.32	0.02
Number of books in home (1–5 scale)	1.86	1.93	1.93	1.94	0.00
Parent is foreign-born	0.28	0.43	0.36	0.36	0.00
English is home language	0.71	0.59	0.65	0.65	0.00
Child has health insurance	0.79	0.78	0.78	0.80	0.02
Married	0.26	0.36	0.31	0.33	0.02
Child tested in both English and Spanish	0.11	0.31	0.20	0.24	0.04
Father lives in home	0.35	0.44	0.40	0.41	0.01
Full day OK pre-K	0.92	—	0.92	—	—
Parent education missing	0.21	0.16	0.19	0.17	0.02
Nonparental care missing	0.26	0.29	0.27	0.27	0.00
Internet missing	0.22	0.16	0.20	0.17	0.03
Books in home missing	0.21	0.17	0.19	0.17	0.02
Foreign-born parent missing	0.10	0.06	0.08	0.08	0.00
Home language missing	0.16	0.13	0.15	0.13	0.02
Health insurance missing	0.17	0.13	0.16	0.13	0.03
Marital status missing	0.21	0.18	0.19	0.18	0.01
Father status missing	0.22	0.17	0.20	0.18	0.02
Health status missing	0.16	0.13	0.15	0.13	0.02
Medical visit missing	0.16	0.13	0.15	0.13	0.02
Outcomes					
Assessment at kindergarten entry					
WJ letter-word raw score—Cohort 1	10.51	7.98	10.35	8.08	
	(4.06)	(4.06)	(4.14)	(4.01)	
WJ applied problems raw score—Cohort 1	13.15	12.95	13.03	12.62	
	(3.97)	(3.94)	(3.90)	(4.08)	
WJ spelling raw score—Cohort 1	9.06	8.53	9.05	8.46	
	(2.90)	(2.41)	(2.96)	(2.40)	

(continued)

TABLE 1 (CONTINUED)

	(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	Diff
Assessment at age 4 program entry					
WJ letter-word raw score—Cohort 2	4.55	4.81	4.53	4.82	
	(3.14)	(3.14)	(3.12)	(4.03)	
WJ applied problems raw score—Cohort 2	8.39	8.00	8.42	7.86	
	(4.76)	(4.66)	(4.55)	(4.63)	
WJ spelling raw score—Cohort 2	4.25	5.04	4.54	4.86	
	(2.14)	(3.03)	(2.55)	(3.10)	
Observations	211	329	211	329	

Note. Sample restricted to children who are eligible for free and reduced-price lunch. Cohort 1 refers to the group of children who participated in OK pre-K or Head Start during the 2005–2006 school year and are entering kindergarten at the time of the assessment, the start of the 2006–2007 school year. Cohort 2 refers to the group of children who are entering OK pre-K or Head Start in the 2006–2007 school year. Diff refers to differences between the observed and PS weighted proportions or means, where $p < .05$. HS = Head Start; OK pre-K = Oklahoma pre-kindergarten; PS = propensity score; WJ = Woodcock–Johnson Achievement Tests–III.

Using RD to compare the mean outcomes of children who made the cutoff with those who did not provides “pseudo” pre- and posttest 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 to 6 years old and are entering kindergarten at the outcome assessment date, and Cohort 2 children are 4 to 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 2005–2006 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 2006–2007 can serve as the pretest 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 of 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), known as the local conditional independence assumption (Van Der Klaauw, 2008). All other characteristics of these individuals can be considered independent of treatment status, and therefore should be “smooth”—not discontinuous—around the cutoff. One can test this assumption by comparing the means of observed characteristics within a bandwidth around the treatment cutoff. We did this for observations very close to the cutoff (90-day bandwidth) and for the full analysis sample (270-day bandwidth) for each pathway (shown in Online Appendix 2.11, available at http://epa.sagepub.com/supplemental). We find that across all variables included in the models there were very few significant relationships between child cohort and child and family covariates within each pathway when the Inverse Probability of Treatment Weights (IPTW) are applied.

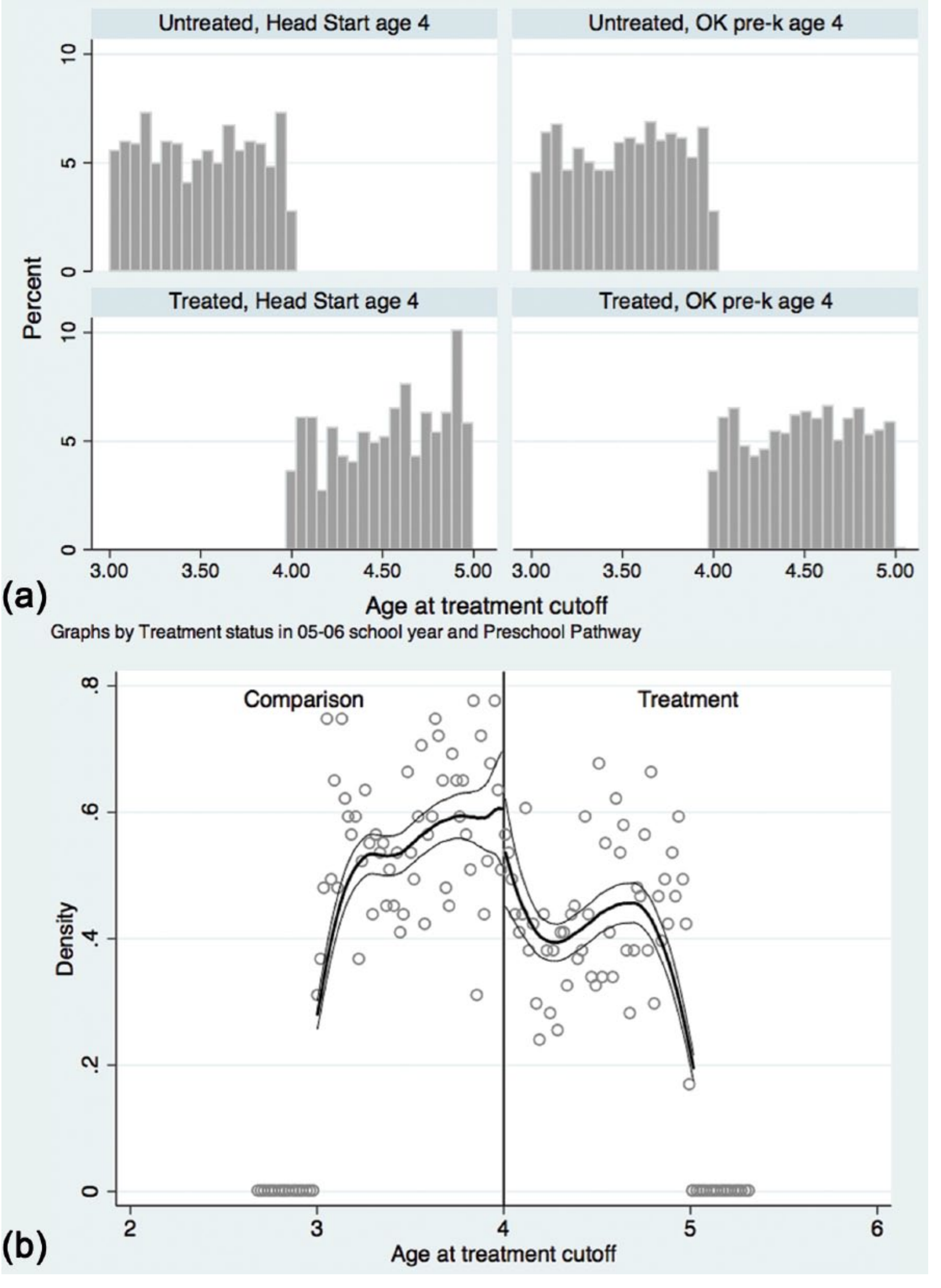

FIGURE 1. *Histogram and McCrary density plot of age by treatment status: (a) Histogram of age by treatment status and preschool pathway; (b) McCrary density plot of age.*

Note. (a) The *x*-axis indicates children's age in years on September 1, 2005 (i.e., the start of the 2005–2006 school year); bars represent the percentage of the sample for each age. These four histograms illustrate that children's treatment status is a function of their ages, which is discontinuous at 4 years. (b) The *x*-axis indicates children's age in years on September 1, 2005. The graph shows the McCrary (2008) test for a discontinuity in the density of children near the birth date cutoff for both pathways combined. Test results confirm no differences in the density of children near the cutoff (Theta = 0.10, *t*-statistic = 0.88, *p* value = .19). OK pre-K = Oklahoma pre-kindergarten.

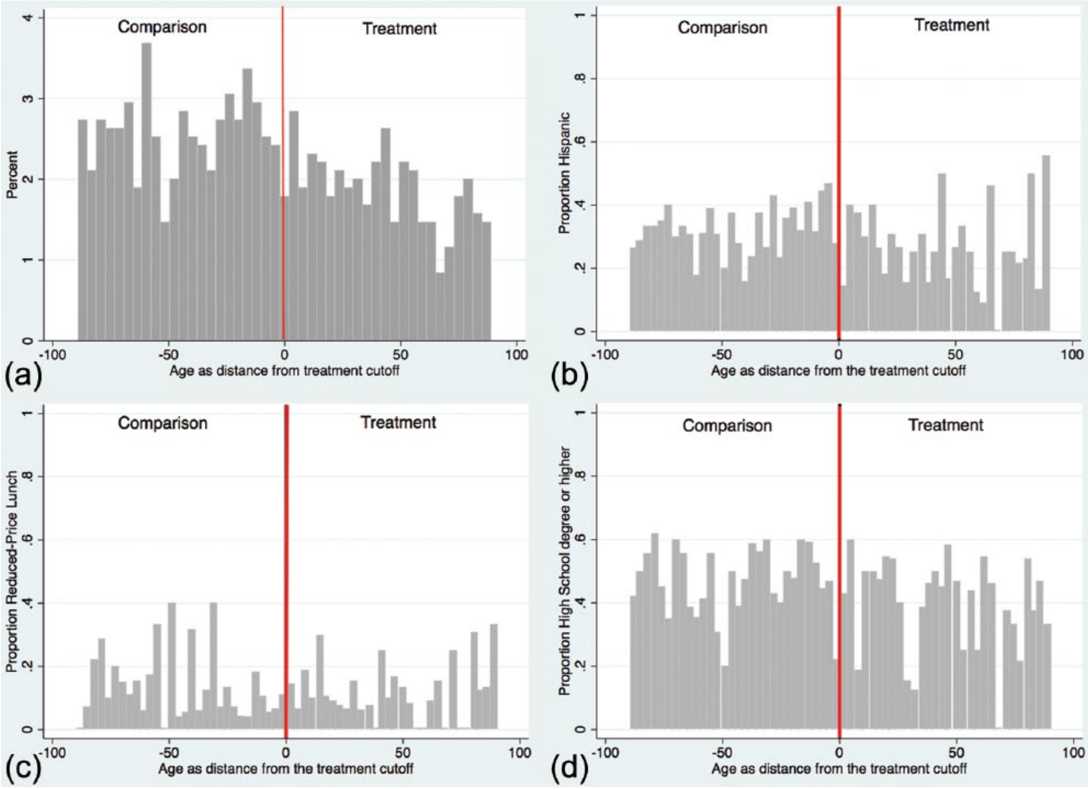

FIGURE 2. *Histograms of the assignment variable and selected covariates within a 90-day bandwidth of the treatment cutoff: (a) Age of study sample relative to the treatment cutoff; (b) proportion of study sample Hispanic relative to the treatment cutoff; (c) proportion of study sample reduced-price lunch eligible relative to the treatment cutoff; (d) proportion of study sample parents with High School degree or higher.*

Note. (a) The *y*-axis indicates the percent of children within a birth date range around the treatment cutoff in the study sample. This figure shows that the distributions of children's ages are similar on both sides of the cutoff (i.e., no clustering at the cutoff). Children in Cohort 1 are shown on the right-hand side of the figure (treatment), and children in Cohort 2 are shown on the left-hand side (comparison). (b–d) The *y*-axes indicate the percent of children within a birth date range around the treatment cutoff with the identified characteristic (Hispanic, reduced-price lunch eligible, and High School degree or higher, respectively) near the cutoff. These figures illustrate that the distributions of children's observable characteristics are similar on both sides of the cutoff. Children in Cohort 1 are shown on the right-hand side of the figures (treatment), and children in Cohort 2 are shown on the left-hand side (comparison).

We also tested for the smoothness of covariate means around the cutoff graphically. In Figures 2b, 2c, and 2d, we show histograms of covariate proportions for Hispanic, reduced-price lunch, and parents with a high school degree or higher, near the cutoff. These figures illustrate that the distributions of children's observable characteristics are similar on both sides of the cutoff. Because the composition of covariates is similar across the cutoff (i.e., cohorts) within each pathway, these two diagnostics also indicate that our sample is not biased by differential attrition between the preschool and kindergarten years, which is central to the smoothness assumption in age-cutoff RDs (Lipsey et al., 2014). We also used the histograms in Figure 2 to ensure that observations were not disproportionately clustered near the cutoff.

Because age—measured as distance from the birth date 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 as these adjusted scores include the effect of the preschool program, they can be used as posttest measures. The younger students within Cohort 2 have their scores adjusted forward to what they are expected to be at their 5th birthday, and as these adjusted scores do not include the effect of the preschool program they are just entering, they can be used as pretest measures.[3] The effect identified in the RD model is an average treatment effect that generalizes to cases closest to the cutoff and are therefore most similar in potential outcomes, also known as a local average treatment effect (Angrist & Pischke, 2008).

Model Specification. We estimated the RD models using ordinary least squares regression with PS weights (described below) to generate local average treatment effects of each pathway and to test for pathway differential effects on outcomes at kindergarten entry. In combining this estimand with that of PS methods, which estimate the average treatment effect for treated cases, we refer to our estimand as a local average treatment effect on the treated. Comparing two different exposures with RD involved a nuanced RD specification. We include an interaction term between the treatment indicator (birth date 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 nonparental care (*yes* = 1), and missing data indicators, presented below:

$$Y_{ijc} = \alpha + \beta_1 \text{Cutoff}_{ic} + \beta_2 \left(\text{Cutoff}_{ic} \times \text{HS}_{ic}\right) + \beta_3 \text{HS}_{ic} + \beta_4 \left(\text{Age}_{ic} - Q\right) + \beta_5 \left(\text{Age}_{ic} - Q\right)^2 + Z_{ic} + e_{ic},$$

where Y is one of three early academic skill outcome measures (j), indexed by child (i) and classroom (c). *Cutoff* is a dichotomous indicator of whether the child's birth date 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—our coefficient of interest—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 (local) average treatment effect for OK pre-K, the reference group. β_4 is the effect of the quantitative assignment variable, age, which is measured in days and is centered at the birth date cutoff Q (September 1). β_5 is a quadratic version of age and **Z** is a vector of control variables. The error term is indexed by child and classroom to reflect our classroom clustered standard errors. An RD specification comparing two separate discontinuities as we do here (β_2) is also referred to as a "difference-in-discontinuities" design (Grembi, Nannicini, & Troiano, 2012).

Because the treatment effect comes from this discontinuity in outcomes at the birth date 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 (0.75 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 Online Appendix 1.2, available at http://epa.sagepub.com/supplemental, for further detail on our RD methodology and robustness tests.

2. *Addressing Selection: PS 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. PS weights induce comparability between Head Start and OK pre-K children, allowing us to 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 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 as we do not know why age 3 Head Start participants would choose pre-K over Head Start at age 4. Another assumption of PS methods in our application is that the relationship between individual characteristics and treatment for both Head Start and OK pre-K children follow the same functional form (i.e., a logistic response function).

One can implement PS methods in a number of ways, with matching methods being most common (Caliendo & Kopeinig, 2008). In this study, we use a method based on 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 calculating the PSs for each age 3 Head Start graduate, we assessed whether there was common support across the age 4 OK pre-K and Head Start groups using the histograms shown in Figure 3. This indicated that there was adequate overlap in PSs, meaning that individuals in both treatment states were comparable with respect to their propensity for treatment (i.e., were exchangeable), allowing us to use PS methods.

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 PS 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. An asterisk in the left column indicates a significant difference in proportions. 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.

Ideally, we would have additional variables in our PS equation to help us further capture a family's preference for pre-K and Head Start (e.g., distance between children's homes and OK pre-K and Head Start program sites). However, we use the same set of covariates that Gormley, Phillips, Newmark, Welti, and Adelstein (2011) use in their PS analysis study, matching children who attended OK pre-K to kindergarten children who did not attend either Head Start or OK pre-K (and analogously matching Head Start participants). These variables provide more detailed information on children and their families than "convenience" variables alone (i.e., age, gender, race, marital status; Shadish, Clark, Steiner, & Hill, 2008). In addition, PS methods are better able to remove bias when comparing cases within the same locality and when study outcomes are short term (proximate to selection), as is the case in our Tulsa sample (Bloom, Michalopoulos, Hill, & Lei, 2002). See Online Appendix 1, available at http://epa.sagepub.com/supplemental, for further detail.

Results

Pathway Effects

Full model results are presented in Table 2, and the main findings are illustrated in Figure 4. 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 gray box at the top of the table that includes the calculated effect sizes shown below the standard error of the estimate.

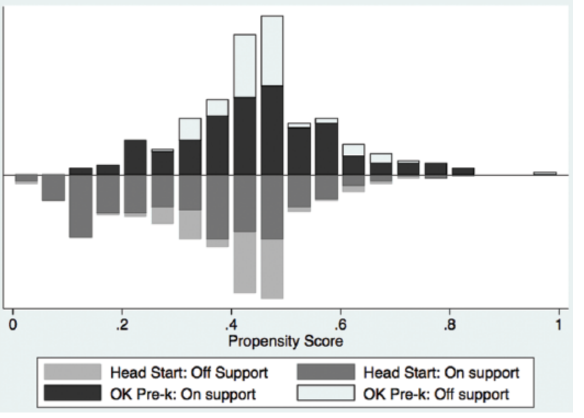

FIGURE 3. *Histogram of propensity scores to assess common support between age 4 treatment states.*
Note. Bar height indicates the proportion of children at each value of the propensity score value for the age 4 OK pre-K and age 4 Head Start groups to assess common support. These overlay histograms show that there is adequate overlap in propensity scores, meaning that individuals in both preschool pathways were comparable with respect to their propensity for treatment. OK pre-K = Oklahoma pre-kindergarten.

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

The effect sizes for the WJ-LW 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.

Another way to test for dosage effects of a second year in Head Start would be to compare the outcomes of children who attended 2 years of Head Start with those that only attended 1 year. We tested this using the OK study data, comparing children who attended Head Start at age 4 with those who attended at both ages 3 and 4. We employed the same methodology as above, combining RD and PS weighting. The results are shown in Online Appendix 2.3, available at http://epa.sagepub.com/supplemental. 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.[4]

Descriptive Comparison of Classroom Peers

In Online Appendix 2.2, available at http://epa.sagepub.com/supplemental, we present the average assessment scores for the age 3 Head Start graduates measured at the beginning of their age 4 programs in 2006–2007 (using the younger cohort) as a proxy for a post–age 3 Head Start assessment.[5] We compare the age 3 Head Start graduates attending OK pre-K with those attending a second year of Head Start and find that the two groups do not have significantly different LW and applied problems scores (p = .45, .50), but that second-year Head Start entrants have higher spelling scores (standardized mean difference [SMD] = 0.27, p = .00). This indicates that the two groups of children were comparable in terms of most academic skills at the start of their age 4 program. However, comparing the ability and characteristics of age 3 Head Start graduates with their classroom peers in their age 4 programs who did *not* attend age 3 Head Start reveals more consistent differences. Age 3 Head Start graduates appear to have stronger early

TABLE 2

Propensity Score Weighted Regression Discontinuity Results for the Effects of Age 3 HS + Age 4 OK Pre-K Versus Age 3 HS + Age 4 HS

	Letter-word		Applied problems		Spelling	
	B (*SE*)	*d*	*B* (*SE*)	*d*	*B* (*SE*)	*d*
Age 4 OK pre-K and age 3 HS effect (cutoff)	3.77*** (1.03)		0.69 (1.10)		2.17*** (0.73)	
Effect size		0.92	0.14			0.68
Age 4 HS and age 3 HS effect (Pathway × Cutoff + cutoff)	1.88* (0.98)		1.36 (1.06)		1.72** (0.72)	
Effect size		0.46	0.27			0.53
Age 4 HS and age 3 HS differential effect (Pathway × Cutoff)	−1.89** (0.88)		0.66 (1.09)		−0.45 (0.69)	
Effect size and direction of difference		−0.46	+0.13			−0.14
p value of difference		.02	.54			.51
Covariates						
Age 4 HS and 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)		0.00000063 (0.0000069)	
Female	0.68* (0.39)		0.45 (0.44)		0.82*** (0.29)	
Child had some nonparental 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)	
Maternal education						
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)	
Child race						
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)	

(continued)

TABLE 2 (CONTINUED)

	Letter-word		Applied problems		Spelling	
	B (*SE*)	*d*	*B* (*SE*)	*d*	*B* (*SE*)	*d*
Missing nonparental care	1.19		1.45**		0.60	
	(0.62)		(0.66)		(0.42)	
Constant	3.60***		7.70***		3.45***	
	(0.82)		(1.10)		(0.69)	
Observations	407		404		391	

Note. Reference group for effect of exposure is age 4 OK Pre-K + age 3 HS. Observations that fall within the 270-day bandwidth from the treatment cutoff are included (Age-birth date cutoff ≤ 270 in absolute value). Outcome variable is a raw score. All models use clustered *SE*s by teacher. OK pre-K = Oklahoma pre-kindergarten.
*significant at .10 level. **significant at .05 level. ***significant at .01 level.

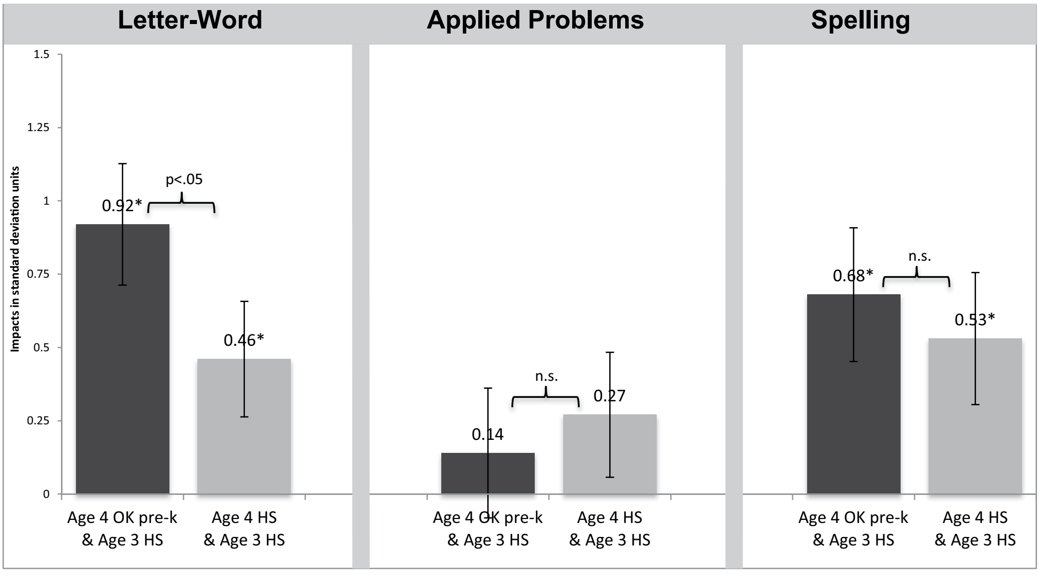

FIGURE 4. *Regression discontinuity-propensity score weighted results comparing age 4 OK Pre-K + Age 3 HS versus Age 4 HS + Age 3 HS.*
Note. Bars represent preschool exposure effect sizes for each outcome. Brackets indicate the significance of the difference in effect sizes between the two preschool pathways. OK pre-K = Oklahoma pre-kindergarten; HS = Head Start.

academic skills relative to their peers in age 4 Head Start, while the skills of those graduates attending OK pre-K are fairly similar to their peers. In addition, the peers of children in the OK pre-K are from higher income families. These comparisons indicate—at least descriptively—the potential for different peer effects for both the OK pre-K entrants and age 4 Head Start entrants (further detail in the online appendix, available at http://epa.sagepub.com/supplemental).

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, universal state pre-kindergarten program at age 4?" There was limited prior research on whether the Head Start program is effective as a 2-year program that builds on 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: (a) age 3 Head Start and age 4 OK pre-K, and (b) age 3 Head Start and age 4 Head Start. We employed a combination of strong quasi-experimental methods, using RD to estimate the effects of both age 4 programs, and PS weighting to address selection into these two pathways into kindergarten.

Our findings suggest that children attending Head Start at age 3 will have stronger early reading skills if they attend a high-quality universal pre-K program 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 LW 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 2 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 early writing and spelling skills (ES = 0.68, 0.53; no significant difference) and neither program significantly improved children's math skills.

Although the only significant differential effect we found in our study was on the LW score, the effect size for the difference was substantial—.46—where children who switched to OK pre-K had twice the estimated effect size of their Head Start peers. Recent estimates of the disparity in reading scores between kindergarten children in the top and bottom deciles of income are 1.25 standard deviations (Reardon, 2011). In terms of the achievement gap, then, the .46 effect we find in our study would represent more than one third of this disparity in early reading 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., 2014; Reynolds, 1995; Reynolds et al., 2011; Schweinhart & Weikart, 1981; Tarullo et al., 2013). There are 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. It may be that the curricula used in Head Start classrooms do 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 differentiated learning experiences that are essential to children's intellectual development (Bronfenbrenner, 1989). Because the OK pre-K advantage was concentrated to early 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. Although there are numerous ways in which these program models differed, our study was not able to assess which of these program characteristics caused the observed difference because they are confounded with program type. However, this is an important avenue for future research.

Furthermore, if programs are not designed to build on gains, they may show lower incremental impacts when measured toward the end of the program relative to children's outcomes measured midprogram. 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 Carolina Approach to Responsive Education (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 Head Start, 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, whereas 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. Although 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.

Overall, our study suggests that these two preschool pathways *may* matter. However the specific reasons for why they may matter, and the extent to which they matter in different states with different programs, must be studied in future research. Indeed, we did not find any differences between the two programs in improving children's early writing skills, and neither improved their early math skills. In other contexts, it is possible that there may be no differences in the sequencing of programs on children's school readiness. Understanding the differences between these two pathways is important for policy, but we could not know the causal effects based on our study alone.

The most substantial limitation of our study is that PS methods assume there is no unobserved confounding, which is not testable, and therefore our estimates do not represent causal effects. We also were not able to assess the specific mechanisms or program features through which OK pre-K produced better reading skills, and this must be addressed by future research. The other study limitations are as follows: (a) The Tulsa programs may not be representative of most state pre-K and Head Start programs because of Tulsa's stringent quality standards and classroom quality ratings that are higher than national averages; (b) children living in Tulsa, Oklahoma, are not representative of the broader population of children in the United States; (c) 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; (d) our sample sizes may not provide sufficient power to detect effects; (e) we cannot know why some parents took their children out of Head Start in the second year; (f) we do not have other neighborhood or school-level information about the representativeness of Head Start and OK pre-K program sites, nor do we have classroom-level information about the teachers' curricular choices and instructional practices to explore our hypotheses about differential instruction; (g) we do not have information about children's summer learning opportunities between their age 3 and age 4 programs; (h) we cannot assess whether our RD estimates would be biased from sample attrition into kindergarten in the younger cohort, and (i) 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 often 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.

Authors' Note

The content is solely the responsibility of the authors and does not necessarily represent the official views of Institute of Education Sciences (IES), the U.S. Department of Education, or the National Institutes of Health.

Acknowledgments

We thank the Center for Research on Children in the United States at Georgetown University for collecting and sharing the data. We would also like to thank Ana Auger, Marianne Bitler, Damon Clark, Robert Crosnoe, Thad Domina, Dale Farran, William Gormley, and Sean Reardon for helpful comments on prior drafts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We are grateful to the Institute of Education Sciences (IES) for supporting this work through Grant R305B120013 awarded to Principal Investigator Greg Duncan and Co-Principal Investigators Farkas, Vandell, Bitler, and Carpenter. 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 No. P01-HD065704.

Notes

1. The Head Start Impact Study did not include an experimental analysis of participating in 1 year versus 2 because children were able to select into receiving Head Start at age 4 after being randomly assigned to treatment at age 3.

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 propensity score, 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. The Dummy Variable Adjustment (DVA) approach is biased if covariates with missing data and without missing data are correlated, but unbiased if uncorrelated with one another (Puma, Olsen, Bell, & Price, 2009). In our sample, these correlations were all below .1. We also tested the robustness of our DVA approach relative to multiple imputation (fully conditional specification; 50 imputed data sets) by estimating our regression discontinuity (RD) models using both methods without weighting by the propensity scores. Both missing data strategies yield very similar coefficients and standard errors, with no major differences in significance on our focal treatment variables (shown in Online Appendix 2.12, available at http://epa.sagepub.com/supplemental).

3. We checked for noncompliance with the age cutoff in the data and found very few children who did not comply with the treatment assignment rule (7 total). These children are omitted from the analysis.

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

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

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 empiricist's companion*. Princeton, NJ: Princeton University Press.

Arteaga, I., Humpage, S., Reynolds, A. J., & Temple, J. A. (2014). 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*, *40*, 221–237. doi: 10.1016/j.econedurev.2013.07.009

Bailey, D. B., McWilliam, R. A., Ware, W. B., & Burchinal, M. R. (1993). Social interactions of toddlers and preschoolers in same-age and mixed-age play groups. *Journal of Applied Developmental Psychology*, *14*, 261–276. doi: 10.1016/0193-3973(93)90036-U

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Barnett, W. S., & Belfield, C. R. (2006). Early childhood development and social mobility. *The Future of Children*, *16*, 73–98.

Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). *The state of preschool 2011*. New Brunswick, NJ: 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*, 108–132. doi: 10.1162/003465304323023714

Bell, E. R., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2013). Classroom age composition and rates of change in school readiness for children enrolled in Head Start. *Early Childhood Research Quarterly*, *28*, 1–10. doi: 10.1016/j.ecresq.2012.06.002

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 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*, 21–26. doi: 10.1016/S0272-7757(99)00022-9

Blasco, P. M., Bailey, D. B., & Burchinal, M. R. (1993). Dimensions of mastery in same-age and mixed-age integrated classrooms. *Early Childhood Research Quarterly*, *8*, 193–206.

Bloom, H. S., Michalopoulos, C., Hill, C. J., & Lei, Y. (2002). *Can nonexperimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* (MDRC Working Papers on Research Methodology). New York, NY: MDRC

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*, 73–97.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*, 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*, 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*, 231–242.

Card, D. (1999). The causal effect of education on earnings. In C. A. Orley & C. David (Eds.), *Handbook of labor economics* (Vol. 3, Part A, pp. 1801–1863). San Diego, CA: 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). doi: 10.3386/w13663

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *The Quarterly Journal of Economics*, *126*, 1593–1660. doi: 10.1093/qje/qjr041

Clements, D., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*, 443–494. doi: 10.3102/0002831207312908

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*, 126–143.

Clifford, R. M., & Crawford, G. M. (Eds.). (2009). *Beginning school: U.S. policies in international perspective*. New York, NY: Teachers College Press.

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *The American Economic Review*, *85*, 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*(1–2), 30–43. doi: 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*, 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*, 704–713. doi: 10.1016/j.ecresq.2013.04.004

Dotterer, A. M., Burchinal, M. R., Bryant, D., Early, D. M., & Pianta, R. (2012). Universal and targeted pre-kindergarten programmes: A comparison of classroom characteristics and child outcomes. *Early Child Development and Care*, *183*, 931–950. doi: 10.1080/03004430.2012.698388

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives*, *27*, 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*, 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*, 641–683.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*, 157–178. doi: 10.3102/0162373712461850

Foster, E. M. (2011). Deployment and the citizen soldier: Need and resilience. *Medical Care*, *49*, 301–312.

Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *The American Economic Review*, *92*, 999–1012.

Gibbs, C., Ludwig, J., & Miller, D. L. (2011). *Does Head Start do any lasting good?* (National Bureau of Economic Research Working Paper Series, No. 17452). doi: 10.3386/w17452

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.

Gormley, W. T. (2008). The effects of Oklahoma's pre-K program on Hispanic children*. *Social Science Quarterly*, *89*, 916–936. doi: 10.1111/j.1540-6237.2008.00591.x

Gormley, W. T. (2011). *Tulsa 2006-07 public use data set*. Washington, DC: Georgetown University.

Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, *40*, 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*, 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*, 397–418. doi: 10.1111/j.1541-0072.2010.00367.x

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, *320*, 1723–1724. doi: 10.1126/science.1156019

Gormley, W. T., Phillips, D. A., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. *Child Development*, *82*, 2095–2109. doi: 10.1111/j.1467-8624.2011.01648.x

Grembi, V., Nannicini, T., & Troiano, U. (2012). *Policy responses to fiscal restraints: A difference-in-discontinuities design*. (CESifo working paper: Public finance). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1852523

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes.

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*, 449–481. doi: 10.1016/S0883-0355(03)00015-6

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, *65*, 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 pre-kindergarten. *Educational Evaluation and Policy Analysis*, *28*, 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*, 730–744. doi: 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*, 27–50.

Hoxby, C. M., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects*. Cambridge, MA: Harvard University.

Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly*, *27*, 33–45. doi: 10.1016/j.ecresq.2011.03.006

Hulsey, L. K., Aikens, N., Kopack, A., West, J., Moiduddin, E., & Tarullo, L. B. (2011). *Head Start children, families, and programs: Present and past data from FACES*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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: 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*, 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*, 147–165.

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, *82*, 1768–1777. doi: 10.1111/j.1467-8624.2011.01665.x

Katz, L. G. (1990). *The case for mixed-age grouping in early education* (Publication No. 333 ed.). Washington, DC: National Association for the Education of Young Children.

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*, 155–178. doi: 10.1080/19345740802114533

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*, 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*, 698–716. doi: 10.1002/jcop.20462

Lipsey, M. W., Farran, D. C., Bilbrey, 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.

Lipsey, M. W., Weiland, C., Yoshikawa, H., Wilson, S. J., & Hofer, K. G. (2014). The prekindergarten age-cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis*. Advance online publication. doi: 10.3102/0162373714547266

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*, 52–66. doi: 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*, 47–65.

Lombardi, J. (2003). *Time to care: Redesigning child care to promote education, support families, and build communities*. Philadelphia, PA: Temple University Press.

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, *24*, 305–337. doi: 10.1007/s11145-009-9214-6

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*, 257–268.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *The Economics of Early Childhood Education*, *26*, 33–51.

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development*, *80*, 686–702. doi: 10.1111/j.1467-8624.2009.01291.x

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*, 732–749. doi: 10.1111/j.1467-8624.2008.01154.x

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*, 698–714.

Moiduddin, E., Aikens, N., Tarullo, L. B., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009*. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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*, 741–753. doi: 10.1037/a0013099

Montessori, M. (1917). *The advanced Montessori method* (Vol. 1). New York, NY: Frederick A. Stokes.

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*, 271–282. doi: 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*, 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*, 213–228. doi: 10.1016/j.ecresq.2009.05.002

Pianta, R., & Howes, C. (Eds.). (2009). *The promise of pre-K*. Baltimore, MD: Paul H. Brookes.

Pianta, R., Howes, C., Burchinal, M. R., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*, 144–159.

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.

Puma, M., Olsen, R. B., Bell, S., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials*. Jessup, MD: National Center for Education Evaluation and Regional Assistance.

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*, 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*, 2–14. doi: 10.1207/s1532480xads0401_1

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity* (pp. 91–116). New York, NY: Russell Sage.

Reid, J. L., & Ready, D. D. (2013). High-quality preschool: The socioeconomic composition of preschool classrooms and children's learning. *Early Education and Development*, *24*, 1082–1111. doi: 10.1080/10409289.2012.757519

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*, 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. *Journal of the American Medical Association*, *285*, 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*, 41–55.

Rouse, C., Brooks-Gunn, J., & McLanahan, S. (2005). School readiness: Closing racial and ethnic gaps. *The Future of Children*, *15*, 5–13.

Schechter, C., & Bye, B. (2007). Preliminary evidence for the impact of mixed-income preschools on low-income children's language growth. *Early Childhood Research Quarterly*, *22*, 137–146.

Schochet, P., Cook, T. D., Deke, J., Imbens, G. W., Lockwood, J. R., Porter, J., & Smith, J. (2010). *What Works Clearninghouse: Standards for regression discontinuity designs*. Washington, DC: 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*, 29–39. doi: 10.1177/105381518100400105

Shadish, W. R., Clark, M. H., Steiner, P. M., & Hill, J. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments [with comments, rejoinder]. *Journal of the American Statistical Association*, *103*, 1334–1350. doi: 10.2307/27640186

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

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*, 42–49. doi: 10.1016/j.ecresq.2010.05.001

StataCorp. (2011). *Stata statistical software: Release 12*. College Station, TX: Author.

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, DC: 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.

Urberg, K. A., & Kaplan, M. G. (1986). Effects of classroom age composition on the play and social behaviors of preschool children. *Journal of Applied Developmental Psychology*, *7*, 403–415. doi: 10.1016/0193-3973(86)90009-2

U.S. Department of Education. (2013). *Early childhood education intervention report: The creative curriculum for preschool* (4th ed.). Washington, DC: Institute of Education Sciences, What Works Clearinghouse.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (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*, 737–756. doi: 10.1111/j.1467-8624.2010.01431.x

Van Der Klaauw, W. (2008). Regression–discontinuity analysis: A survey of recent developments in economics. *LABOUR*, *22*, 219–245. doi: 10.1111/j.1467-9914.2008.00419.x

Vygotsky, L. S. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*, 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*, 684–694. doi: 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*, 305–330. doi: 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*, 122–154.

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

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*, 134–152. doi: 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*, 75–92. doi: 10.1002/(sici)1520-6688(200024)19:1<75::aid-pam5>3.0.co;2-w

Authors

JADE MARCUS JENKINS is an assistant professor in the School of Education at the University of California, Irvine, and was an Institute of Education Sciences (IES) postdoctoral fellow during the study time period. Her research is on early childhood development and public policy.

GEORGE FARKAS is a professor in the School of Education at the University of California, Irvine. His work has made major contributions to educational inequality, low-income and ethnic minority students, and how it can be reduced.

GREG J. DUNCAN is a Distinguished Professor in the School of Education at University of California. He has published extensively on understanding child poverty and the importance of early academic skills, cognitive and emotional self-regulation, as well as health in promoting children's eventual success in school and the labor market.

MARGARET BURCHINAL is a senior scientist at the Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill. She is a widely recognized expert in early child care and education research and applied statistical methodology.

DEBORAH LOWE VANDELL is the founding dean of the School of Education at the University of California, Irvine, where she is a professor of education and of psychology and social behavior. She is a renowned expert on the effects of afterschool and summer programs on children's academic, social, and behavioral development.

Manuscript received March 11, 2014
Revision received April 14, 2015
Accepted April 15, 2015