



Impacts of the New Mexico PreK initiative by children's race/ethnicity

Jason T. Hustedt^{a,*}, Kwanghee Jung^b, Allison H. Friedman-Krauss^b, W. Steven Barnett^b, Gerilyn Slicker^a

^a University of Delaware

^b National Institute for Early Education Research, Rutgers University



ARTICLE INFO

Article history:

Received 17 July 2019

Received in revised form 5 June 2020

Accepted 8 September 2020

Keywords:

State prekindergarten

Preschool

Kindergarten readiness

Academic outcomes

Regression-Discontinuity

ABSTRACT

New Mexico is one of 44 U.S. states offering a public pre-K program for children at age 4. State models for pre-K vary in terms of availability, policies related to classroom quality, and populations of children served. In this study, we pool data from five successive cohorts of children (total N = 5218) using regression-discontinuity models to estimate the impacts of participating in New Mexico's pre-K program on young children's language, literacy, and math skills at kindergarten entry. Positive, statistically significant impacts of pre-K were found for each of these academic domains. Due to the high level of diversity in our sample, it was also possible to examine pre-K impacts separately for White, Hispanic, and Native American children. The largest impacts were found for White and Hispanic children, with less consistent and more modest impacts for Native American children. These findings suggest that while New Mexico's pre-K program generated academic benefits for children, not all groups of children benefited equally, and further information is needed to understand the reasons for these differences.

? 2020 Published by Elsevier Inc.

1. Introduction

One-third of 4-year-olds in the U.S. participate in state-funded prekindergarten (pre-K) programs during the year before entering kindergarten, and these initiatives are now offered by 44 states and the District of Columbia (Friedman-Krauss et al., 2020). Both child enrollment levels and state financial investments in state pre-K have increased dramatically over the past two decades. During the 2018–2019 school year, states collectively invested more than \$8.75 billion in pre-K programs designed to prepare children for success in kindergarten and beyond. Large-scale and rapid growth in the state pre-K sector, variation between states in the pre-K models employed, and population differences across U.S. states create a context where more research is needed about the effectiveness of individual state pre-K programs. This study examines cognitive outcomes associated with the New Mexico PreK initiative, employing child assessment data gathered during the initiative's first five years of operation. Because this program serves large numbers of Hispanic and Native American children in line with the state's population, we are able to examine pre-K impacts by child race and

ethnicity in order to assess the effectiveness of the program for children from diverse backgrounds.

Momentum for the earliest state pre-K initiatives began to grow during the 1960s, the same decade in which the federal Head Start program began. However, the majority of the currently operating state pre-K initiatives were established much more recently, in the 1990s or afterward (Friedman-Krauss et al., 2020). These current initiatives have goals related to school readiness and are informed by a growing evidence base about the impacts of early childhood education programs on young children's development. Many state pre-K programs also aim to close achievement gaps at kindergarten entry in order to level the playing field between ethnic/racial minority children and their non-minority peers. Initial evidence that informed the design of state pre-K initiatives focused on research from relatively small-scale model programs, where longitudinal follow-up data has been gathered well into the adult lives of participating children. Key outcomes identified in such research (e.g., Campbell et al., 2012; Reynolds, Temple, Ou, Arteaga, & White, 2011; Schweinhart et al., 2005) include improved high school graduation rates, lower involvement in crime and lower arrest rates, and higher levels of educational attainment at age 30. However, statewide pre-K initiatives differ in important ways from the programs involved in these seminal studies: State pre-K initiatives are now offered at a much larger scale; the historical context has changed significantly, with preschool participation rates much higher now than they were several decades ago; and the popula-

* Corresponding author at: Department of Human Development and Family Sciences, University of Delaware, 111 Alison Hall West, Newark, DE 19716.

E-mail address: jhustedt@udel.edu (J.T. Hustedt).

tion of children served by such initiatives has changed (Barnett, Friedman-Krauss, & Weiland, *in press*).

1.1. Findings from studies of state Pre-K

As state pre-K programs began to expand, researchers conducted evaluations of these emerging models, though in general the initial state studies suffered from serious methodological problems (Gilliam & Zigler, 2000). One key limitation of early state pre-K studies involved specifying an appropriate comparison group to accurately estimate the impacts of pre-K participation. Children whose parents enroll them in pre-K may differ in a variety of ways from those whose parents do not; thus, the specification of comparison groups is a critical design issue. More recent studies have used techniques including regression-discontinuity designs (RDD) based on age-eligibility cutoffs (e.g., Barnett et al., 2018; Gormley, Gayer, Phillips, & Dawson, 2005), as well as randomized control trials (RCTs; e.g., Lipsey, Farran, & Durkin, 2018), both in an effort to identify more equivalent comparison groups and reduce bias in the estimation of program effects.

A number of methodologically rigorous studies have now examined the overall effectiveness of single-state pre-K models. These individual studies have made important contributions to the development of a national knowledge base about public pre-K impacts in the U.S. (e.g., Phillips et al., 2017; Yoshikawa et al., 2013), given that there are notable differences between pre-K models being used in different states rather than a single nationwide model. In general, single-state studies have shown positive impacts of state pre-K on children's language, literacy, and math skills when measured at kindergarten entry. Domains such as children's health and social-emotional development have been much less frequently examined (Yoshikawa et al., 2013). Regression-discontinuity studies of pre-K programs in states including Arkansas (Hustedt, Jung, Barnett, & Williams, 2015), Georgia (Peisner-Feinberg, Schaaf, LaForett, Hildebrandt, & Sideris, 2014), North Carolina (Peisner-Feinberg & Schaaf, 2011), and Oklahoma (Gormley, Phillips, & Gayer, 2008) each provide evidence of positive academic impacts associated with these models, when children enter kindergarten. In Tennessee, Lipsey et al. (2018) conducted an RCT where researchers assigned children to participate in pre-K or to a control condition; this was possible because demand for pre-K slots greatly exceeded the number of slots available. An RDD component of this study was used to estimate pre-K impacts at kindergarten entry, while the RCT component adds to the study's methodological rigor and was used to provide estimates of possible pre-K impacts through third grade. Results from the RCT study show positive academic impacts of state pre-K measured at the end of the pre-K year, a finding consistent with other state studies. However, estimated pre-K impacts were no longer statistically significant by first grade and in some cases were negative during the elementary school years.

Another set of state pre-K studies has used common analytic strategies to examine pre-K impacts across groups of states where similar data are available. An early RDD study of state pre-K (Wong, Cook, Barnett, & Jung, 2008) incorporated data from five distinct state programs, showing generally positive impacts across emergent literacy, mathematics, and language as measured at kindergarten entry. A common RDD approach and consistent data collection procedures were used in each state. A more recent study of eight state pre-K programs (Barnett et al., 2018) incorporated new data from several of the original five states, as well as data from three additional states. More advanced data analysis strategies including sensitivity analyses were used with these RDD models to better address threats to validity, consistent with methodological advances after the earlier study was completed. Again, a key finding across a broad set of state pre-K programs was that the greatest pre-K impacts related to children's early literacy

skills, with an estimated weighted average effect size of 1.10 standard deviation (SD). Estimated average impacts of 0.44 SD were found for mathematics and 0.24 SD for language, across the eight states. In general, findings from multi-state studies of pre-K are consistent with findings from studies examining pre-K impacts in single states, including the finding that the largest pre-K impacts were in the domain of early literacy.

1.2. Moving beyond average impacts of state Pre-K

The research described previously is consistent in showing that state funded pre-K programs have positive effects on children at kindergarten entry, though the magnitude of the effects does differ across state pre-K models. Typically, evaluations of publicly-funded pre-K have examined overall effectiveness of these programs rather than examining impacts separately for different groups of children. Yet, pre-K programs may not be equally effective for all children. For example, do children with different types of strengths, different prior family or cultural experiences, or experiencing different risks, benefit in similar ways? This is especially relevant in the context of publicly-funded pre-K, where eligibility is frequently determined on the basis of risks including poverty. Additionally, a goal of many pre-K programs is to promote equity for all children. Therefore, of particular interest in the current study is whether there are differences in pre-K impacts across racial and ethnic groups, including White, Hispanic, and Native American children. While differences in race and ethnicity promote diversity and do not themselves connote risk, racism imparts risk, and ethnic minority families experience heightened levels of poverty and attendant risks in comparison to non-minority families. A key goal of public pre-K programs, including New Mexico PreK, involves providing supportive early learning experiences that promote equity for children and families from racial and ethnic minority groups. Thus, understanding the extent to which pre-K programs benefit diverse groups of children is important.

Prior research shows that, on average, Hispanic and Native American children tend to enter kindergarten with fewer of the academic skills necessary to succeed in school compared to their White peers (Lee & Burkam, 2002; National Center for Education Statistics, 2008; Reardon & Portilla, 2016; Reardon, 2011). Pre-K programs have been touted as one method for narrowing achievement gaps linked with race and ethnicity at kindergarten entry (Friedman-Krauss, Barnett, & Nores, 2016). Also, some evidence suggests that impacts of early childhood education programs at kindergarten entry do vary by subgroup (more detail is provided in the following paragraphs; e.g., Bitler, Hoynes, & Domina, 2014; Bloom & Weiland, 2015; Gormley et al., 2005; Yoshikawa, Weiland, & Brooks-Gunn, 2016). Generally, effects tend to be larger for low-income, non-White, and Dual Language Learner groups. The Tennessee pre-K study described earlier (Lipsey et al., 2018) examined differences in pre-K impacts based on whether or not children were native English speakers, and found that non-native English speakers demonstrated the largest academic gains at the end of pre-K, especially those with the lowest initial scores at the beginning of the pre-K year. Also, in a recent study of universal pre-K in Georgia, Early, Li, Maxwell, and Ponder (2019) found that English Language Learners showed greater math gains by third grade than children who spoke English at home.

Gormley et al. (2005) examined impacts associated with participation in sites of Oklahoma's statewide pre-K program within the city of Tulsa, using an RDD approach. The overall results showed positive effects of pre-K on young children's letter-word identification, spelling, and early math scores, but the magnitude of impacts varied by children's race/ethnicity. In addition, unique to the Tulsa study, the Native American subgroup was sufficiently large to estimate separate pre-K impacts for Native American children. Pre-K

impacts were statistically significant for White, Black, Hispanic, and Native American children in each of the measured domains, except in the domain of early math, where outcomes for Native American children approached significance ($p < .10$) and outcomes for White children were nonsignificant. In each domain, the largest effect sizes were found for Hispanic children, followed by Native American children. In comparison to White children, effect sizes for Black children were comparable in letter-word identification, smaller in spelling, and larger in math. In sum, the authors concluded that the pre-K program benefited children from diverse backgrounds.

Weiland and Yoshikawa (2013) also used an RDD approach to evaluate a citywide public pre-K initiative in Boston, demonstrating moderate-to-large positive overall effects within the domains of language, literacy, and math, as well as small but statistically significant effects within the less-frequently assessed domains of emotion recognition and executive function. The effects of the Boston pre-K program also varied by children's race/ethnicity: compared to White children, Hispanic and Asian children each showed significantly greater gains across 8 of the 12 child outcome domains, while Black children showed significantly greater gains for 3 of 12 domains. However, within each racial/ethnic group, not all differences in outcomes remained when sensitivity analyses were conducted. In general, effects were strongest for Hispanic children. Findings for letter-word identification and early math skills can be directly compared to the Gormley et al. (2005) study, as the same assessment tools were used in both studies. For letter-word identification, Weiland and Yoshikawa (2013) found the largest pre-K impacts for Hispanic children, with comparatively smaller but still statistically significant impacts for Black and Asian children. Statistically significant impacts were not found for White children. For mathematics skills, statistically significant impacts were found for children in all racial/ethnic groups, but were largest for Asian and Hispanic children.

Although prior research is very limited, the finding that Hispanic and Black children may show comparatively greater benefits from pre-K participation at kindergarten entry (Gormley et al., 2005; Weiland & Yoshikawa, 2013) is consistent with results from studies using national datasets from children participating in other types of early childhood education programs. In an analysis of data from the Head Start Impact Study, Bitler et al. (2014) found the largest effects for Hispanic and Spanish-speaking children upon completing one year of Head Start and that the positive effects of pre-K remain for Spanish-speaking children in first grade. Bloom and Weiland (2015) also found larger impacts of Head Start for Dual Language Learners and Spanish-speaking children. Bassok (2010) analyzed data from the Birth cohort of the Early Childhood Longitudinal Study (ECLS-B) and found larger preschool impacts (compared to parental care) for Black children than for Hispanic or White children. Among Head Start participants, results suggest that Hispanic children benefited more than White children, compared to children in parental care. This study is especially notable because of efforts to statistically distinguish contributions of family income and race/ethnicity. For low-income children in the Bassok (2010) study, there was a significant main effect of preschool participation, but there were not racial and ethnic group differences. However, a different pattern of results emerged for non-low-income children in the same study. Those who were Black and those from primarily Spanish-speaking Hispanic families benefited more from preschool than those who were White. Finally, in a study using state administrative records to examine longer-term pre-K outcomes from North Carolina's More at Four program (Ladd, Muschkin, & Dodge, 2014), impacts on children's third grade math skills, but not reading skills, were greater for children whose mothers were Hispanic. Statistically significant differential impacts were not found for children whose mothers were Black.

In sum, the evidence suggests that not all children benefit equally from attending publicly-funded early childhood education programs. Patterns of results are somewhat variable across different types of public programs, but this overall finding is concerning, given that equity and narrowing academic achievement gaps are often key goals of such programs. Several studies find larger impacts for children who speak a language other than English as well as for Hispanic children. Only one prior study had an appropriate sample to examine impacts separately for subgroups of Native American children, finding that impacts were slightly larger for Native American than White children but smaller than for Hispanic children (Gormley et al., 2005). The current study is the first to provide estimates of impacts of the New Mexico PreK program for Hispanic and Native American children, and the first pre-K study to examine impacts for Native American children using a statewide sample.

1.3. State-funded prekindergarten in New Mexico

The New Mexico PreK program was first established in 2005 with an initial cohort of about 2000 children and served 4264 children by 2011 (Barnett, Hustedt, Hawkinson, & Robin, 2006; Barnett, Carolan, Fitzgerald, & Squires, 2011). By the 2018–2019 school year, 38% of New Mexico's 4-year-olds were enrolled, with a state investment of \$68.2 million (Friedman-Krauss et al., 2020). Children are served in both public school and community-based pre-K settings. Although the state does not use specific family income criteria to determine eligibility for New Mexico PreK, for each site offering the program there is a requirement that two-thirds of participating children must reside within Title I school attendance zones (Friedman-Krauss et al., 2020). Further, Census data show that New Mexico has persistently high rates of child poverty, and in 2015 the state had the second highest rate of child poverty in the U.S. (Annie E. Casey Foundation, 2016). While it is not exclusively designed as a program for children in poverty, New Mexico PreK serves a primarily low-income population. Census data also show that New Mexico is highly diverse, with no single ethnic group representing a majority of the state's population (Annie E. Casey Foundation, 2016). In 2010, the total population was 46% Hispanic, 40% White non-Hispanic, 9% Native American, and 2% Black, with smaller percentages of individuals representing other races or multiple races. For children age 0–4, even larger proportions of the population were Hispanic (59%) and Native American (12%), and a smaller proportion was White non-Hispanic (24%). Thus, the overall population served by New Mexico PreK would be expected to reflect a statewide context of high poverty as well as high levels of diversity.

State policies hold New Mexico PreK sites to high standards with respect to structural quality, meeting 9 of the 10 quality standards benchmarks that the National Institute for Early Education Research (NIEER) uses to rate state pre-K initiatives annually (Friedman-Krauss et al., 2020). In general, this means that the level of aggregate quality in New Mexico is comparable to required levels of quality in other states where rigorous pre-K evaluations have been conducted. The NIEER benchmarks address state-level requirements related to teacher training, maximum class sizes and staff-child ratios, mandated screenings and referrals, availability of curriculum supports, and presence of a continuous quality improvement system. In New Mexico, pre-K classes must enroll no more than 20 children with teacher-child ratios of 1:10 for 4-year-olds and additional services such as health screenings and referrals and meals are required. The state carries out structured classroom observations and also uses data for program improvement. While other pre-K programs that have demonstrated positive effects require all lead teachers to have a bachelor's degree, New Mexico requires bachelor's degrees only for lead teachers in public schools, but not in community settings where only a high school diploma and annual progress towards a bachelor's degree

is required. Whether bachelor's degrees should be an entry-level requirement for teachers of preschool-age children has been a matter of considerable debate, though a National Research Council panel (Allen & Kelly, 2015) recently recommended that all lead teachers of children from birth to age 8 should have bachelor's degrees.

1.4. The current study

This study extends prior research on large-scale public pre-K programs by examining the overall impacts of the state-funded New Mexico PreK initiative. A novel contribution of our study is that we are able to examine child outcomes across the program's first 5 years. Also, our large and diverse sample with high numbers of Hispanic and Native American participants allows us to investigate potential differences in child outcomes across racial and ethnic groups. Our two primary research questions are: (1) What are the overall impacts of the New Mexico PreK program on children's language, literacy, and math skills at kindergarten entry? (2) To what extent do impacts in language, literacy, and math skills differ based on children's racial/ethnic group membership?

2. Method

2.1. The research model

The current study uses an age-cutoff regression-discontinuity approach (Barnett et al., 2018; Gormley et al., 2005) to estimate the effects of attending a single year of the New Mexico PreK program on young children's cognitive skills. In addition to providing overall estimates of pre-K impacts, we present impacts by child race/ethnicity. We pool data from the first 5 years of the New Mexico PreK program between 2005 and 2010 to increase our sample size and the precision of our estimates. Similar procedures were used for each of the 5 years of the study; exceptions are noted below. The overall RDD approach in this study involves making comparisons of two groups of children in a way that minimizes selection bias, because both groups of children enrolled in New Mexico PreK. This strategy is a variant of the RDD model (Bloom, 2012; Lee & Lemieux, 2010) that takes advantage of the state's use of a strict cutoff, the child's birth date, to determine eligibility for both PreK and kindergarten. Children who reach age 4 by August 31 are eligible to attend PreK, while children who reach age 5 by August 31 attend kindergarten. The treatment group in our study is made up of children who completed New Mexico PreK the prior school year and who were starting kindergarten at the time of assessment. The control group is made up of children just entering the New Mexico PreK program. Both groups of children attended New Mexico PreK, but they did so in subsequent years.

2.2. Sampling and participants

Data for the current study come from children who attended the New Mexico PreK program between the 2005–2006 and 2009–2010 school years. Data were collected at the start of each school year, beginning in fall 2006, after the first cohort of children had completed New Mexico PreK. Two groups of children are represented: One group where children were in kindergarten and completed New Mexico PreK the previous year ("treatment") and one group where children were just entering New Mexico PreK ("control"). An assumption of the RDD model is that these two groups are equivalent on observed and unobserved characteristics other than their birthdate. State administrators provided lists of New Mexico PreK sites and numbers of children enrolled at each site, as well as likely kindergarten destinations for children who had recently completed PreK.

In communities offering New Mexico PreK for the first time in the fall of each year, there were not kindergarten children who had completed PreK; as this is a requirement of our analysis, those communities were excluded from the study until the following year. At each remaining site, we randomly selected a pre-specified number of 4-year-old PreK students to participate in our study, based on the proportion of New Mexico PreK children statewide enrolled at that site. Then, PreK children were randomly selected from class enrollment lists until we had completed the designated number of assessments. A corresponding number of kindergartners who attended each site the previous year were randomly selected using the state's master enrollment list and were tracked to the elementary schools where they were enrolled in kindergarten. Following an IRB-approved protocol, consent materials were distributed to children's caregivers via PreK or kindergarten classrooms, and children whose families did not provide consent were excluded from the selection process.

A total of 5218 children are included in our pooled analysis, including 813 children in fall 2006 (Year 1), 856 children in fall 2007 (Year 2), 1260 children in fall 2008 (Year 3), 1303 children in fall 2009 (Year 4), and 986 children in fall 2010 (Year 5). Target sample sizes for each year of the study varied somewhat, depending on availability of funds and the size of the PreK program. The overall sample is representative of children attending New Mexico PreK statewide.

Table 1 provides descriptive statistics for the overall sample used in our RDD analysis. The sample was 51% female. Children were primarily Hispanic (60%), White (19%), or Native American (16%), with much smaller proportions of children representing other racial groups, and 78% of children spoke English as their home language. About half the children attended New Mexico PreK in public school settings and the other half in community settings.

2.3. Measures

Child outcomes were measured across three cognitive domains using standardized, direct assessments.

Language. Children's language skills were measured using the Peabody Picture Vocabulary Test, Third Edition (PPVT-III; Dunn & Dunn, 1997), which specifically assesses children's receptive vocabulary and is predictive of general cognitive abilities. Reliability of the PPVT is good as judged by either split-half or test-retest reliabilities (published reliability = .95). Spanish-speaking children were tested with both the PPVT and the Spanish-language *Test de Vocabulario en Imágenes Peabody* (TVIP; Dunn, Padilla, Lugo, & Dunn, 1986). The TVIP is appropriate for measuring growth in Spanish vocabulary for Spanish-speaking children. Raw scores were used here.

Literacy. Children's literacy skills were initially measured using the Print Awareness subtest of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgesen, & Rashotte, 2002) in both English and Spanish. The Pre-CTOPPP was subsequently published in English as the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007), with the Print Awareness subtest renamed the Print Knowledge subtest. The TOPEL was used in Years 2 and 3 of the study, but the Spanish version of the Pre-CTOPPP continued to be used to assess Spanish-speaking children. An alternative assessment of literacy was used in Years 4 and 5 of this study, but the measure is somewhat different from the TOPEL and has not been widely used in research; thus, here we report only TOPEL outcomes. The TOPEL has been used with both middle- and low-income samples, and reliability of the subtests is reported to be good to excellent (test-retest reliability = .89; $\alpha = .95$). Print awareness/knowledge items measure whether children recognize individual letters and letter-sound correspondences, and

Table 1
Characteristics of New Mexico PreK Children Included in Analyses.

Variable	Total			Current Pre-K (Control)			Current Kindergarten (Treatment)		
	N	% or M	SD	N	% or M	SD	N	% or M	SD
Female	5218			2660			2558		
Race	2661	51.0		1410	53.0		1251	48.9	
White	1003	19.2		524	19.7		479	18.7	
Black	103	2.0		55	2.1		48	1.9	
Hispanic	3107	59.5		1554	58.4		1553	60.7	
Native American	849	16.3		428	16.1		421	16.5	
Other Race	115	2.2		72	2.7		43	1.7	
Missing	41	0.8		27	1.0		14	0.5	
Home Language									
English	4077	78.1		2105	79.1		1972	77.1	
Non-English	1133	21.7		548	20.6		585	22.9	
Child has an IEP	147	2.8		50	1.9		97	3.8	
Child Outcomes									
Language	5192	54.93	22.01	2647	46.39	20.39	2545	63.82	20.04
Math	5182	12.52	5.08	2630	9.96	4.34	2552	15.15	4.39
Literacy	2904	18.80	11.42	1482	11.64	8.44	1422	26.26	9.11

Note. Language data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw scores. Literacy data represent the percentage correct on the Pre-CTOPPP/TOPEL Print Awareness subtest. Spanish-speaking children were tested using the corresponding Spanish-language version of each instrument.

also whether they can differentiate words in print from pictures and symbols. Raw scores were used here.

Math. Children’s math skills were assessed using the Applied Problems subtest from the Woodcock-Johnson Tests of Achievement, Third Edition (WJ-III; Woodcock, McGrew, & Mather, 2001). For Spanish speakers, the comparable subtest from the *Batería Woodcock Muñoz* (Woodcock & Munoz, 1990) was used. Woodcock-Johnson subtests are reported to have good reliability (reliability = .93 for Applied Problems). Raw scores were used here.

2.4. Procedures

Child assessment data were collected during fall of five successive years beginning in 2006, to examine the impacts of attending the New Mexico PreK program during each immediately prior school year. Children were assessed either when they entered PreK or just after attending PreK, when they entered kindergarten. Child assessors were trained by researchers with extensive assessment experience over two days and demonstrated reliability in using each instrument. The assessments were conducted in the child’s school as early in the school year as possible, and all assessment instruments were administered on the same day. The PPVT was administered first to all children in order to gauge children’s English language skills. Children who spoke Spanish were subsequently assessed using the TVIP. The literacy and math assessments were conducted in either English or Spanish, depending on the language the child’s teacher designated as the child’s dominant language. When visiting children’s schools to conduct child assessments, the researchers also gathered information about child age, home language, race/ethnicity, gender, and IEP status. Additional information about the New Mexico PreK sites children attended was provided by state administrators, including whether the program was offered by a public school or a community organization.

2.5. Data analysis

We used an age cutoff RDD approach to estimate the effects of attending New Mexico PreK on child outcomes (e.g., Barnett et al., 2018; Gormley et al., 2005). This analytic approach takes advantage of New Mexico’s birth date cutoff for PreK eligibility where children were required to be age 4 no later than August 31 to be eligible that fall. Two groups of children – one group that met the age cutoff and a second group that did not – are compared using the RDD approach.

These two groups of children are expected to be similar on observed and unobserved characteristics other than age. All children participated in New Mexico PreK but children in the two groups attended the program during different school years. The treatment group for our analyses consists of current kindergartners who completed PreK during the prior school year. The control group consists of children currently enrolled in PreK, tested at the beginning of their PreK year. Our primary RDD analyses were instrumental variable (IV) models in order to handle “fuzzy” cases that violate the birth date cut-off assignment rule. Although the number of “fuzzy” cases was minimal in this sample (1%), we use age eligibility for New Mexico PreK as an instrument for actual participation. In order to examine the possibility that there was manipulation of child placement relative to the age cutoff, we plotted the density of children’s age relative to the cutoff. Fig. 1 shows that child placement was not manipulated at the New Mexico sites.

The equation that we used to estimate the effect of participation in New Mexico PreK was:

$$Y_{ij} = \beta_0 X_{ij} + \beta_1 (\text{PreK})_{ij} + \beta_2 g(\text{AV})_{ij} + \beta_3 D_k + \beta_4 C_{ij} + \epsilon_{ijk}$$

Where *Y* is the child cognitive test score for child *i* in school *j*, *X* is a vector of child characteristics, *PreK* is a binary variable that equals 1 if the child participated in New Mexico PreK already and 0 if the child just entered New Mexico PreK, *g(AV)* is a smooth function of the age variable, *D* represents school district fixed effects, *C* represents the cohort (i.e., year 1 of New Mexico PreK) and ϵ is the error term. Child covariates include child race/ethnicity, home language, and gender, if the child had an IEP, and language of testing. We also controlled for the number of days between the beginning of the program and testing as well as if the program was in a public school or community organization. This model was re-estimated separately for racial/ethnic subgroups of children. *T*-tests were used to determine if effects sizes differed significantly between White, Hispanic, and Native American children.

Our primary analysis estimated the impacts of New Mexico PreK using a 6-month window around the kindergarten entry date. In addition, we developed models employing samples using both wider (12-month) and narrower (3-month) windows around the cutoff. Further, we estimated the optimal bandwidth of each outcome measure (following Lee & Lemieux, 2010) and our subsequent analyses used models based on the optimal bandwidth selected. We also conducted the McCrary (2008) to explore if there was a discon-

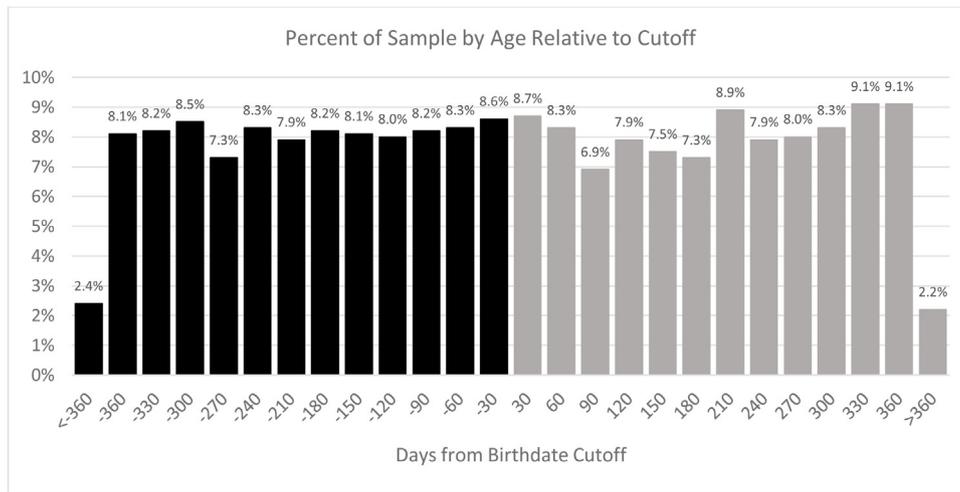


Fig. 1. Density of Observations around the Age Cutoff.

Table 2
Estimated New Mexico PreK Impacts from Regression-Discontinuity Analysis.

	Empirically Identified Functional Form	Linear, 6 months	Quadratic, 6 months	Cubic, 6 months	Truncated at 3 months	Truncated at 12 months
Language	Linear	5.90*** (1.51)	6.31*** (1.55)	6.13*** (1.68)	6.92** (2.23)	4.57*** ^a (1.20)
Math	Linear	1.94*** (0.34)	2.16*** (0.35)	1.92*** (0.38)	2.06*** (0.48)	1.87*** ^a (0.28)
Literacy	Linear	7.99*** (1.10)	7.30*** (1.08)	7.07*** (1.38)	6.68*** (1.56)	9.21*** (0.73)

Note. Robust standard errors are in parentheses. Language data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Literacy data represent increases in the percentage correct on the Pre-CTOPPP/TOPEL Print Awareness subtest. Spanish-speaking children were tested using the corresponding Spanish-language version of each instrument.

^a A cubic functional form was selected rather than a linear functional form.
*** $p < .001$.

tinuity in the assignment variable at the cutoff. These analyses did not show any evidence of manipulation of the assignment variable.

3. Results

We estimated overall impacts of attending one year of New Mexico PreK on children’s language, literacy, and math skills at kindergarten entry, using regression-discontinuity models. In the interest of taking advantage of a large sample size while still providing more precise estimates of pre-K effects around the kindergarten cutoff than 12-month models that are frequently used, our primary findings are the results from IV models employing a 6-month time window around the kindergarten cutoff date. These results, as well as findings from models using 3- and 12-month windows around the cutoff, are depicted in Table 2.

Results from our primary 6-month models show that children who completed one year of New Mexico PreK just prior to entering kindergarten demonstrated statistically significant increases ($p < .001$) in language skills, math skills, and early literacy skills. We explored linear, quadratic, and cubic functional forms to ensure we modeled the correct relationship between the assignment variable and the outcomes. The preferred functional form was linear for all outcome measures, though overall quadratic and cubic models were also statistically significant. We tested additional regression models employing samples with 3- and 12-month windows around the kindergarten cutoff date. Again, we found statistically significant ($p < .001$) increases in children’s language, math, and early literacy outcomes.

In Table 3, we present effect sizes for the overall sample. Again, three sets of separate RDD models representing 6-, 3-, and 12-month windows around the kindergarten cutoff date were used. The estimated effect sizes were similar across these three sets of models. We found the largest effect sizes for the measure of early literacy, consistent with prior research. For the 6-month sample, estimated effect sizes were 0.29 for language ($p < .001$), 0.45 for math ($p < .001$), and 0.95 for early literacy ($p < .001$). Our sensitivity analyses showed that the estimated effects of New Mexico PreK were robust across 6-, 3-, and 12-month bandwidths.

3.1. Impacts by children’s race and ethnicity

To examine potential differences in New Mexico PreK outcomes based on children’s race and ethnicity, we estimated outcomes separately for Hispanic, White, and Native American subsamples of children (see Table 4). These three groups represent 95% of the total sample for this study. While prior studies (e.g., Gormley et al., 2005; Weiland & Yoshikawa, 2013) have also estimated pre-K outcomes for subsamples of Black children, consistent with the overall New Mexico population, this group was not well represented in our sample (only 2%). Fig. 2 shows effect sizes by racial and ethnic subgroup for each of the child outcome measures. All effect sizes are calculated based on the results from models employing a 6-month time window around the kindergarten cutoff date.

In each domain, effect sizes were largest for White children, somewhat lower for Hispanic children, and lowest for Native American children. Effect sizes for Native American children were much more variable than estimates for the other groups. Language

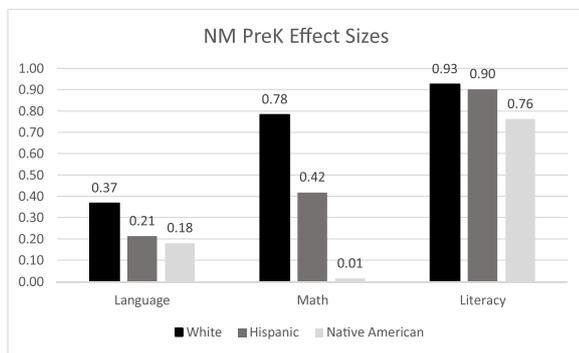


Fig. 2. Estimated New Mexico PreK Effect Sizes. Estimated NM PreK effect sizes for White, Hispanic, and Native American children. Language data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Literacy data represent increases in the percentage correct on the Pre-CTOPPP/TOPEL Print Awareness subtest. Spanish-speaking children were tested using the corresponding Spanish-language version of each instrument.

impacts were modest for all children and statistically significant for White (0.37 SD) and Hispanic (0.21 SD) children but not for Native American children (0.18 SD). Math impacts were also statistically significant for White (0.78 SD) and Hispanic (0.42 SD) children but not for Native American children (0.01 SD). However, unlike language impacts that were all modest, the magnitude of math impacts varied from large for White children to null for Native American children. Also, the difference between the magnitude of math impacts for White and Native American children was statistically significant ($p < .01$). Literacy impacts measured for the first three cohorts of children were uniformly large and statistically significant for all racial and ethnic groups: White (0.93 SD), Hispanic (0.90 SD), and Native American (0.76 SD).

4. Discussion

Our overall findings show that children who participated in the New Mexico PreK program at age 4 benefited with positive impacts on language, literacy, and mathematics outcomes. In general, these results suggest that PreK participants in New Mexico entered kindergarten more academically ready to succeed compared to those who did not attend the program. Although this study was conducted early during the implementation process for a new initiative, child outcomes associated with New Mexico PreK are consistent with outcomes measured in other studies of public prekindergarten programs. Children in our sample are demographically similar to children enrolled in New Mexico PreK today, and program standards today are also similar to those by the end of the data collection period. While the program has now scaled up and expanded, since its inception it has been a statewide program covering geographically diverse regions of the state.

A unique strength of our research design is that it was possible to pool data across five cohorts of PreK participants, resulting in a large sample size. Another strength of the study is that our research design and sampling strategy allowed us to collect data in all New Mexico PreK sites statewide, with participants proportionally added to the overall sample based on the size of their pre-K site. Given that the state PreK program operates in rural and more remote areas of the state as well as larger cities, and in both public and community-based settings, this approach ensured that the broad diversity across program sites was fully represented. Due to the high level of racial and ethnic diversity in both the State of New Mexico and its prekindergarten program, our sample also was sufficiently large for us to estimate impacts separately for Hispanic, White, and Native American PreK subgroups. This approach recog-

Table 3 Estimated Effect Sizes of New Mexico PreK Participation Using Different Bandwidths around the Age Cutoff.

	Effect Size		
	6 month bandwidth	3 month bandwidth	12 month bandwidth
Language	0.29***	0.34**	0.22*** ^a
Math	0.45***	0.47***	0.43*** ^a
Literacy	0.95***	0.79***	1.09***

Note. Effect sizes are calculated using the control group (PreK) sample standard deviations. Effect sizes from the models using a 6-month bandwidth around the age-eligibility cut-off are the preferred models. Language data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Literacy data represent increases in the percentage correct on the Pre-CTOPPP/TOPEL Print Awareness subtest. Spanish-speaking children were tested using the corresponding Spanish-language version of each instrument.

^a A cubic functional form was selected rather than a linear functional form.
 ** $p < .01$.
 *** $p < .001$.

nizes strengths linked with race and ethnicity while at the same time allowing us to examine whether potential achievement gaps are narrowed as a result of attending state pre-K.

Consistent with findings from two key citywide studies (Gormley et al., 2005; Weiland & Yoshikawa, 2013) where similar subgroup estimates were possible, we found positive overall impacts of New Mexico PreK, as well as positive impacts in some academic outcome areas for children in each racial and ethnic group for which analyses were conducted. Although the body of existing research on potential racial and ethnic group differences in preschool impacts is small, this literature suggests that Hispanic children generally experience the most consistently positive outcomes, in comparison to children in other subgroups, including White children. However, somewhat different patterns of results have emerged in different states. For example, Montrosse-Moorhead, Dougherty, La Salle, Weiner, and Dostal (2019) estimated impacts of attending Connecticut’s state pre-K program by racial/ethnic subgroup and found positive impacts at kindergarten entry across 3 of 4 academic domains for Black children, 2 of 4 domains for White children, and no positive impacts for Hispanic children. In addition to public pre-K studies, studies examining impacts by race and ethnicity also include samples from Head Start and other types of early childhood programs (e.g., Bassok, 2010; Bitler et al., 2014; Bloom & Weiland, 2015).

Just as state pre-K initiatives and eligibility criteria differ, states’ populations, primary ethnic groups, and experiences with poverty differ. This may have implications when interpreting the results from pre-K studies conducted in very different contexts. Although children in different states may share a common race or ethnicity, that does not mean that they share common experiences in day-to-day family, school, or community settings, or that they share a common history, particularly in tribal contexts. The Tulsa study (Gormley et al., 2005) is the only prior study of pre-K to separately estimate outcomes for children from Native American backgrounds. While Gormley et al. (2005) found the largest impacts of state pre-K for Hispanic children, they also found consistently positive impacts for Native American children. Effect sizes for White children were comparable to, or in some cases smaller than, those for Native American children.

In New Mexico, we found consistent positive benefits for Hispanic children, although effect sizes in our study were largest for White rather than Hispanic children. Child outcomes within the literacy domain were largest and most consistent across racial and ethnic groups. Child math outcomes were much more variable, with stark differences between groups. In New Mexico, there were relatively large math impacts for White children, more moderate impacts for Hispanic children, and no measured impacts of state

Table 4
Estimated New Mexico Pre-K Impacts by Race/Ethnicity from Regression-Discontinuity Analysis.

	White			Hispanic			Native American		
	<i>b</i>	(SE)	<i>p</i>	<i>b</i>	(SE)	<i>p</i>	<i>b</i>	(SE)	<i>p</i>
Language	7.51	(3.56)	*	4.36	(1.96)	*	3.67	(4.24)	
Math	3.41 ^a	(0.85)	***	1.82	(0.45)	***	0.06 ^a	(0.89)	
Literacy	7.82	(3.05)	*	7.62 ^b	(1.36)	***	6.43 ^b	(2.05)	**

Note. Robust standard errors are in parentheses. Language data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Literacy data represent increases in the percentage correct on the Pre-CTOPPP/TOPEL Print Awareness subtest. Spanish-speaking children were tested using the corresponding Spanish-language version of each instrument.

^a Estimated effects of NM PreK on children’s math are statistically significantly larger for White than Native American children ($p < .01$).

^b A quadratic functional form was selected rather than a linear functional form.

* $p < .05$. ** $p < .01$. *** $p < .001$.

pre-K on Native American children’s math scores. The Tulsa study (Gormley et al., 2005) also found wide disparities in math outcomes, though in that study the largest impacts were found for Hispanic children, with no impacts for White children.

One possible explanation for the differences in results between New Mexico and Tulsa, particularly with regard to the Native American children, is that in our study children were geographically dispersed throughout the state, while in the Tulsa study (Gormley et al., 2005), they all resided in a single city. Therefore, in New Mexico, Native American children may be enrolled in a broader variety of pre-K settings than their counterparts in Tulsa, potentially including less diverse programs and programs of lower quality. It is also important to consider the relationship between poverty and children’s academic skills. We were not able to gather family income data on the children in the current study and this is a limitation. However, on average, the participants in our study likely experienced high levels of economic disadvantage irrespective of their racial or ethnic group membership, since New Mexico has high overall child poverty rates and further targets high-poverty communities for PreK. Also, Native American families may define poverty in different non-material ways than in other cultures, and there are a variety of economic, demographic, and cultural differences across the numerous Native American tribes within New Mexico (Hollis, 2012). Some researchers have argued that between-group differences in areas such as kindergarten readiness can potentially be explained by variations in family income that are associated with racial and ethnic group membership. It is possible that certain children benefit the most from pre-K because they are most economically disadvantaged, and Native American children in New Mexico have higher poverty rates than children in the state’s overall population (Hollis, 2012). It might then be expected that Native American children would derive greater benefits from participating in New Mexico PreK than children from other racial/ethnic groups, but that was not the case in our study. However, Bassok (2010) found that among non-low-income families, Black and Spanish-speaking Hispanic children derived greater benefits from participation in early childhood programs than White children, suggesting that relationships between poverty, race, and children’s academic skills are complex.

When interpreting the results of this study, the characteristics of the New Mexico PreK programs attended by children of different racial/ethnic backgrounds should also be considered. That is, one possible reason for the differences in the size of child outcomes could be that children of different ethnic/racial backgrounds attended different types of programs. Valentino (2017) found that ethnic/racial minority children in state pre-K were more likely to be in low-quality classrooms. In the current sample (Friedman-Krauss, Hustedt, Jung, Barnett, & Slicker, 2019), White children were significantly more likely to attend New Mexico PreK in community organizations than public schools. And Native American children were significantly more likely to attend New Mexico PreK in public schools than community organizations. There were no differences

in types of setting for Hispanic children. It might be expected that impacts are larger for children in public schools where teachers were required to have a bachelor’s degree in early childhood education – and where Native American children also tended to attend PreK. However, data collected in our study show that observed quality was higher in New Mexico PreK programs in community organizations than public schools during the first few years of New Mexico PreK (Friedman-Krauss et al., 2019). This is possibly because community-based programs had a longer history of providing early childhood education programs, prior to the establishment of this initiative.

In addition to differences in the match between student characteristics and program characteristics, it is important to consider the possible role of teacher characteristics. First, though state policy required teachers to have a bachelor’s degree in public school settings, it is likely that during the first few years of New Mexico PreK, teachers were still working towards this goal. Another consideration is whether certain teachers may be particularly attuned to the needs and cultural backgrounds of children in their communities. For example, the teachers in community-based programs may be more likely to share cultural and/or linguistic backgrounds with children in their classrooms. Teachers with more knowledge about their students can individualize instruction and other classroom experiences based on this knowledge. Specific to our study, teachers with greater awareness of Native American children’s home traditions also may be more successful in forming partnerships with families and reflecting those traditions in New Mexico PreK classrooms. While we only collected data about children and programs, data about teachers may make a valuable contribution to understanding our findings about PreK impacts for Native American children. Thus, data about teachers’ educational qualifications and demographics should be collected and analyzed in future studies. Data about teacher experiences and training, including trainings on cultural sensitivity, would also be valuable. Further, there are likely to be changes in teacher skills, effectiveness, and qualifications over time, as preschool initiatives such as the one in New Mexico mature.

It is important to recognize that that our regression-discontinuity approach can only be used to estimate the unique impacts of New Mexico PreK participation after the conclusion of the PreK year, but prior to academic gains linked with subsequent school experiences. Therefore, we cannot determine if the impacts identified here persist as children move through elementary school. Pre-K studies using research designs that permit comparisons further into elementary school (e.g., Early, Li, Maxwell, & Ponder, 2019; Hill, Gormley, Adelstein, & Willemin, 2012; Jung, Barnett, Hustedt, & Francis, 2013; Lipsey et al., 2018) show more mixed results than studies of pre-K impacts at kindergarten entry, though children’s experiences throughout elementary school contribute to their later skills in ways that are difficult to measure. Finally, it is worth noting that data for the current study were collected early during the implementation process, while the New Mexico PreK program was quickly expanding. Although our pooled analyses control for pos-

sible differences in PreK impacts across successive years of data collection, it is possible that impacts of the New Mexico PreK program would be greater if measured during a more mature phase of the program when growth had stabilized and sites and staff members had more experience offering this initiative.

5. Conclusions

A key goal of state-funded pre-K programs is to better prepare children for success when they enter kindergarten. In general, the evidence from our study suggests that New Mexico PreK accomplished that goal for each of the academic outcomes examined. Yet, not all children experienced equally positive outcomes, and the smallest benefits accrued to Native American children. Our results show that for children who participated in New Mexico PreK there were differences in early learning prior to kindergarten entry for Native American children in particular. While these findings suggest a need for additional support for Native American children in this program in order to promote equity, the reasons for the observed racial and ethnic group disparities in PreK impacts need to be better understood. One issue that would benefit from further exploration relates to family socioeconomic status and economic strain (Hurwich-Reiss, Watamura, Raver, & the BTS Consortium Principal Investigators, 2019). While state pre-K initiatives often determine eligibility based on family income, there may be important differences even among families at high risk levels that have implications for family functioning. On average, some racial and ethnic groups may experience greater levels of disadvantage than others, though diversity and strengths within those groups must be recognized. Without more nuanced data on family poverty and associated measures of risk, subgroup analyses based on children's race and ethnicity only tell part of a larger story about the relative benefits of state pre-K programs for different groups of children.

CRedit authorship contribution statement

Jason T. Hustedt: Conceptualization, Methodology, Writing - original draft, Funding acquisition, Project administration. **Kwanghee Jung:** Formal analysis, Data curation, Writing - original draft. **Allison H. Friedman-Krauss:** Writing - original draft, Visualization, Project administration. **W. Steven Barnett:** Methodology, Funding acquisition, Supervision. **Gerilyn Slicker:** Writing - review & editing, Visualization.

Acknowledgements

Funding for this project was provided by the Department of Finance and Administration; the Children, Youth and Families Department; and the Public Education Department through the Legislature of the State of New Mexico. Additional funding was provided by The Pew Charitable Trusts. The funders were not involved in the study design; data collection, analysis, or interpretation; or the writing of this report. The authors are solely responsible for the views expressed here, and do not have competing interests. We extend special thanks to Alexandra Figueras-Daniel, Ellen Frede, Linda Goetze, and Scott Hughes.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ecresq.2020.09.006>.

References

- Allen, L., & Kelly, B. B. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington: National Academies Press.
- Annie E. Casey Foundation. (2016). *KIDS COUNT data center* Retrieved from <https://datacenter.kidscount.org/data/#NM>.
- Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). *The state of preschool 2011: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. S., Friedman-Krauss, A. H., & Weiland, C. (in press). Scale up of effective preschool education: New directions for research. In Graue, M. E., Ryan, S., Gadsden, V., & Levine, F. (Eds.), *Building research capacity in early childhood*. Washington: American Educational Research Association. In press.
- Barnett, W. S., Hustedt, J. T., Hawkinson, L. E., & Robin, K. B. (2006). *The state of preschool 2006: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. S., Jung, K., Friedman-Krauss, A., Frede, E. C., Nores, M., Hustedt, J. T., et al. (2018). State prekindergarten effects on early learning at kindergarten entry: An analysis of eight state programs. *AERA Open*, 4(2), 1–16. <http://dx.doi.org/10.1177/2332858418766291>
- Bassok, D. (2010). Do Black and Hispanic children benefit more from preschool? Understanding differences in preschool effects across racial groups. *Child Development*, 81, 1828–1845. <http://dx.doi.org/10.1111/j.1467-8624.2010.01513.x>
- Bitler, M. P., Hoynes, H. W., & Domina, T. (2014). *Experimental evidence on distributional effects of Head Start*. NBER working paper No. 20434. Cambridge, MA: National Bureau of Economic Research.
- Bloom, H. S. (2012). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, 5, 43–82. <http://dx.doi.org/10.1080/19345747.2011.578707>
- Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. New York, NY: MDRC working paper.
- Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., et al. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology*, 48, 1033–1043. <http://dx.doi.org/10.1037/a0026644>
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test-third edition (PPVT-III)*. Circle Pines, MN: AGS Publishing.
- Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de vocabulario en Imágenes Peabody (TVIP)*. Circle Pines, MN: AGS Publishing.
- Early, D. M., Li, W., Maxwell, K. L., & Ponder, B. D. (2019). Participation in Georgia's Pre-K as a predictor of third-grade standardized test scores. *AERA Open*, 5(2), 1–16. <http://dx.doi.org/10.1177/2332858419848687>
- Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. A. (2020). *The state of preschool 2019: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Friedman-Krauss, A. H., Barnett, W. S., & Nores, M. (2016). *How much can high-quality universal pre-K reduce achievement gaps?* Washington DC: Center for American Progress.
- Friedman-Krauss, A. H., Hustedt, J. T., Jung, K., Barnett, W. S., & Slicker, G. (2019, March). Does auspice matter? Impacts of New Mexico PreK in public and nonpublic settings. In M. McCormick (Ed.), *Where do pre-K programs operate best? Exploring variation in pre-K access and quality in mixed-delivery systems*. Washington: Symposium presented at the annual conference of the Society for Research on Educational Effectiveness.
- Gilliam, W. S., & Zigler, E. F. (2000). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly*, 15, 441–473. [http://dx.doi.org/10.1016/S0885-2006\(01\)00073-4](http://dx.doi.org/10.1016/S0885-2006(01)00073-4)
- Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, 41, 872–884. <http://dx.doi.org/10.1037/0012-1649.41.6.872>
- Gormley, W. T., Jr., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, 320, 1723–1724. <http://dx.doi.org/10.1126/science.1156019>
- Hill, C. J., Gormley, W. T., Jr., Adelstein, S., & Willemin, C. (2012). *The effects of Oklahoma's pre-kindergarten program on 3rd grade test scores*. Washington, DC: Center for Research on Children in the U.S., Georgetown University.
- Hollis, C. (2012). *Native American children and families in New Mexico: Strengths and challenges*. Albuquerque, NM: New Mexico Kids Count, New Mexico Voices for Children.
- Hurwich-Reiss, E., Watamura, S. E., Raver, C. C., & the BTS Consortium Principal Investigators. (2019). Beyond income: Expanding our empirical toolkit to better predict caregiver well-being. *Journal of Child and Family Studies*, 28, 753–764. <http://dx.doi.org/10.1007/s10826-018-01304-5>
- Hustedt, J. T., Jung, K., Barnett, W. S., & Williams, T. (2015). Kindergarten readiness impacts of the Arkansas Better Chance state prekindergarten initiative. *The Elementary School Journal*, 116, 198–216. <http://dx.doi.org/10.1086/684105>
- Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). *Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. *Journal of*

- Policy Analysis and Management, 33, 162–187. <http://dx.doi.org/10.1002/pam.21734>
- Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, D.C: Economic Policy Institute.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48, 281–355. <http://dx.doi.org/10.1257/jel.48.2.281>
- Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45, 155–176. <http://dx.doi.org/10.1016/j.ecresq.2018.03.005>
- Lonigan, C., Wagner, R., Torgeson, J., & Rashotte, C. (2002). *Preschool comprehensive test of phonological and print processing (Pre-CTOPPP)*. Tallahassee: Department of Psychology, Florida State University.
- Lonigan, C. J., Wagner, R. K., Torgeson, J. K., & Rashotte, C. A. (2007). *Test of preschool early literacy (TOPEL)*. Austin, TX: Pro-Ed.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714. <http://dx.doi.org/10.1016/j.jeconom.2007.05.005>
- Montrosse-Moorhead, B., Dougherty, S. M., La Salle, T. P., Weiner, J. M., & Dostal, H. M. (2019). The overall and differential effects of a targeted prekindergarten program: Evidence from Connecticut. *Early Childhood Research Quarterly*, 48, 134–145. <http://dx.doi.org/10.1016/j.ecresq.2019.02.006>
- National Center for Education Statistics. (2008). *Status and trends in the education of American Indians and Alaska Natives: 2008*. Washington, D.C: National Center for Education Statistics. Retrieved from: nces.ed.gov/pubs2008/nativetrends/ind_4_2.asp.
- Peisner-Feinberg, E. S., & Schaaf, J. M. (2011). *Effects of the North Carolina More at Four pre-kindergarten program on children's school readiness skills: Summary of key findings*. Chapel Hill, NC: Frank Porter Graham Child Development Institute.
- Peisner-Feinberg, E. S., Schaaf, J. M., LaForett, D. R., Hildebrandt, L. M., & Sideris, J. (2014). *Effects of Georgia's pre-K program on children's school readiness skills*. Chapel Hill, NC: Frank Porter Graham Child Development Institute.
- Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., et al. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects*. Washington & Durham, NC: Brookings Institute & Duke University.
- Reardon, S. F. (2011). *The widening academic achievement gap between the rich and the poor: New evidence and possible explanations*. In R. Murnane, & G. J. Duncan (Eds.), *Whither opportunity? Rising inequality, schools and children's life chances*. New York, NY: Russell Sage Foundation Press.
- Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, 2(3), 1–18. <http://dx.doi.org/10.1177/2332858416657343>
- 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. <http://dx.doi.org/10.1126/science.1203618>
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry preschool study through age 40 (monographs of the High/Scope Educational Research Foundation, 14)*. Ypsilanti, MI: High/Scope Press.
- Valentino, R. (2017). Will public pre-k really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal*, 55(1), 79–116. <http://dx.doi.org/10.3102/0002831217732000>
- 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. <http://dx.doi.org/10.1111/cdev.12099>
- Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state prekindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122–154. <http://dx.doi.org/10.1002/pam.20310>
- Woodcock, R. W., & Munoz, A. F. (1990). *Bateria Woodcock-Munoz pruebas de aprovechamiento – Revisados*. Itasca, IL: Riverside Publishing.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL: Riverside Publishing.
- Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, 26(2), 21–35. <http://dx.doi.org/10.1353/foc.2016.0010>
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., et al. (2013). *Investing in our future: The evidence base on preschool education*. Ann Arbor, MI & New York: Society for Research in Child Development & Foundation for Child Development.