Benefits of Pre-Kindergarten for Children in Baltimore, MD

Brenda Jones Harden[a,b], Bonnie E. Bretta[,d], Jacquelyn T. Gross[a], Christina Weiland[c], Jordan Berne[c], Elisa L. Klein[a], and Christy Tirrell-Corbin[a]

[a] University of Maryland College of Education, [b] Columbia University, [c] University of Michigan, [d] Colorado Department of Education

Published in *Early Childhood Research Quarterly* 2023

Authors' note: This study was part of a larger evaluation of the Pre-Kindergarten Expansion program in the state of Maryland. Major funding for this study was provided by the MD State Department of Education. Additional funding was provided by the Brady Education Fund and, for Jordan Berne's time, the Institute of Education Sciences Predoctoral Training grant to the University of Michigan R305B200011. Although the sponsors approved the study design prior to funding, as it was included in the proposal for funding, the sponsors were not involved in any aspect of study implementation. Corresponding author: Department of Human Development and Quantitative Methods, University of Maryland College of Education, 3942 Campus Way, College Park, MD 20742, USA. E-mail address: bjharden@umd.edu

Early Childhood Research Quarterly 64 (2023) 1–12

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

journal homepage: www.elsevier.com/locate/ecresq

Benefits of pre-kindergarten for children in Baltimore, MD

Brenda Jones Harden[a,b,*], Bonnie E. Brett[a,d], Jacquelyn T. Gross[a], Christina Weiland[c], Jordan Berne[c], Elisa L. Klein[a], Christy Tirrell-Corbin[a]

[a] University of Maryland College of Education, College Park, MD, USA
[b] Columbia University, School of Social Work, New York, NY, USA
[c] University of Michigan, Ann Arbor, MI, USA
[d] Colorado Department of Education, CO, USA

ABSTRACT

Substantial research has documented that Pre-Kindergarten (Pre-K) programs have important benefits for the school readiness of young children, especially those from low-income backgrounds. Recent studies in this arena have taken advantage of the Regression Discontinuity (RD) design to evaluate the impacts of Pre-K, due to ethical and pragmatic issues. In this study, we use an RD approach to evaluate the effects of the Preschool Development Program in Baltimore, MD, a city with a large proportion of families from socioeconomically and racially marginalized backgrounds who contend with multiple environmental risks. We compared a group of children who had received Pre-K (n = 554) to a group who had not yet participated in Pre-K (n = 542). Results suggest that the Pre-K experience positively affects these children in a variety of domains, including language (d = 0.41–0.74), early literacy (d = 0.99–1.36), and early mathematical problem-solving (d = 0.54–0.71), a pattern of benefits which is consistent with the targets of the Pre-K curriculum in this jurisdiction. However, effects were not robust for executive functioning, a domain that has had less empirical attention in Pre-K evaluations. Importantly, our study also includes several enhancements to the Pre-K RD design that are recommended by experts but that have not been used widely in this literature. Thus, this study contributes both to the literature on the benefits for children, from largely minoritized, low-income backgrounds, of preschool education implemented at scale and to the methodological literature on RD evaluations of Pre-K programs.

1. Introduction

Scholars from a wide array of disciplines have argued that early childhood education (ECE) is a promising mechanism for promoting the school readiness of children from low-income backgrounds (Besharov et al., 2020; Daniel, 2018; Heckan, 2012; McCoy et al., 2017). The results of studies of ECE programs, designed in the middle of the twentieth century, have provided empirical evidence of the benefits of ECE, such as Head Start (Currie & Thomas, 1995), the Perry Preschool (Weikart et al., 1978), and the Abecedarian Project (Campbell & Ramey, 1994). Many studies document that attendance in ECE programs that aim to improve school readiness (e.g., Pianta et al., 2007) increases children's chances of success in both Kindergarten and the early school years (Campbell & Ramey, 1994; McCarton et al., 1997; Reynolds & Temple, 1995; Weikart et al., 1978). Such studies have led to policies that expanded ECE efforts across the United States, including public Pre-Kindergarten (Pre-K) programs.

Recently, there has been increased investment in the accessibility and affordability of Pre-K in the United States. The 2013 Preschool for All (PA) initiative called for federal support to expand access to high quality Pre-K programs for all children in the U.S. (White House, 2013, 2015). Over a four-year period, a total of 167,725 children living at or below 200% of the federal poverty limit and in high-needs communities attended high quality Pre-K classrooms supported by these grants (Office of Early Learning, 2019). Today, all but six states in the United States (including the District of Columbia) offer some form of state-funded Pre-K (Parker et al., 2018). The 2021 report on the state of preschool (pre-pandemic) from the National Institute for Early Education Research (NIEER) reveals that 47% of the nation's 4-year-olds and 16% of 3-year-olds have been provided early learning opportunities across all public programs, including preschool general and special education, as well as federal- and state-funded Head Start (Friedman-Krauss et al., 2021). The most recent NIEER report documents the effect of the COVID pandemic on preschool participation; it had decreased to 29% of 4-year-olds and 5% of 3-year-olds (Friedman-Krauss et al., 2022).

Multiple studies have documented benefits of Pre-K attendance, including enhanced Kindergarten readiness, cognitive development, and physical and mental health outcomes, and in the long term, higher high school graduation rates, reduced engagement in juvenile delinquency and crime, increased participation in the workforce, stable household formation, and higher income in adulthood (Ansari & Winsler, 2016; Campbell et al., 2002; Council of Economic Advisors, 2015; D'Onise et al., 2010; Gormley et al., 2005; Gray-Lobe et al., 2021; Ludwig & Miller, 2007; Schweinhart et al., 2005; for reviews, see Barnett, 1995; Barnett et al., 2018; Currie, 2001; Gormley et al.,

* Corresponding author: Department of Human Development and Quantitative Methods, University of Maryland College of Education, 3942 Campus Way, College Park, MD 20742, USA.
E-mail address: bjharden@umd.edu (B.J. Harden).

https://doi.org/10.1016/j.ecresq.2023.01.011
Received 3 August 2021; Received in revised form 29 December 2022; Accepted 16 January 2023

Table 1
Sample means (standard deviations) or percentages for child outcomes and covariates.

Variable	Full sample (n = 997) *M* (*SD*) or%	Comparison (Cohort2) (n = 529) *M* (*SD*) or%	Treatment (Cohort 1) (n = 468) *M* (*SD*) or%	% Missing
Outcomes				
HTKS score	14.35 (16.84)	6.87 (11.82)	22.75 (17.66)	.30
PPVT raw score	69.96 (25.71)	58.63 (23.23)	82.56 (22.25)	.90
W-J LW Identification raw score	10.44 (7.18)	6.64 (4.46)	14.71 (7.27)	.30
W-J Applied Problems raw score	10.31 (4.39)	8.01 (3.75)	12.90 (3.53)	.30
W-J Passage Comp raw score	6.75 (3.07)	5.64 (2.11)	8.00 (3.48)	.30
W-J Word Attack raw score	3.84 (2.91)	2.28 (1.69)	5.60 (2.99)	.30
Covariates				
Child age (in years) at outcome	5.14 (0.58)	4.67 (0.29)	5.68 (0.30)	.10
Distance (in days) between test date and 9/1/2017 cutoff	55.15 (18.11)	51.27 (17.63)	59.52 (17.66)	.10
Distance (in days) between birthdate and 9/1/2012 cutoff	−4.23 (208.79)	−172.52 (106.38)	185.99 (108.81)	.00
HTKS – child tested on Form B	84.10	72.20	97.40	.30
HTKS – child tested in English	86.70	85.20	88.50	.30
PPVT – child tested on Form B	56.40	55.40	57.50	.20
W-J – child tested on Form B	53.50	53.20	53.80	.30
Child is female	51.30	51.20	51.30	.00
Child is Black	68.80	67.90	69.90	.00
Child is Hispanic	19.60	20.20	18.80	.00
Child usually speaks only English at home	79.50	80.60	78.10	7.92
Parent education is HS or less	64.90	64.00	66.10	9.53
Parent age (in years)	31.48 (7.92)	31.15 (8.11)	31.92 (7.66)	12.34
Parent is mother	84.10	85.30	82.60	7.62
Parent is employed fulltime	48.60	46.40	51.40	9.53
Number of adults in home	2.06 (1.06)	2.03 (1.01)	2.10 (1.13)	10.83
Number of children in home	2.38 (1.25)	2.38 (1.28)	2.38 (1.20)	9.93
Birth mother lives in home	94.10	94.90	93.20	7.52
Birth father lives in home	45.20	44.40	46.20	8.32
Early care – at home with primary caregiver	47.80	48.40	47.00	8.22
Early care – informal (e.g., relative) care/family daycare	23.70	23.20	24.40	8.22
Early care – Early Head Start	27.70	25.50	30.50	8.22
Early care – center-based	15.30	17.30	12.70	8.22

Note. HTKS = Head-Toes-Knees-Shoulders (Executive Functioning assessment); PPVT = Peabody Picture Vocabulary Test IV (Receptive Vocabulary assessment); W-J = Woodcock-Johnson Tests of Achievement IV; LW = Letter Word; HS = High School.

2008; Heckman & Masterov, 2007). In their study of early elementary school students in Baltimore, MD (the setting of the current study), Williams et al. (2019) found that students who had experienced Head Start plus public Pre-K had stronger kindergarten readiness skills than children who had been in informal care and other early childhood programs.

Most Pre-K evaluations document benefits of program attendance in key academic areas, such as language, literacy, and math, at least in the short run (e.g., Barnett et al., 2018; Gormley et al., 2008; Pion & Lipsey, 2021; Lipsey et al., 2011; Weiland & Yoshikawa, 2013). Although the longitudinal evaluation of the Tennessee public Pre-K program suggested that the long-term benefits of the program are either attenuated or reversed (Durkin et al., 2022; Lipsey et al., 2018), there is evidence that Pre-K confers academic benefits on participant children in the long-run (Bai, Ladd, Muschkin, & Dodge, 2020; Barnett & Jung, 2021; Bartik & Hershbein, 2018; Hill et al., 2015). However, there is variability among studies of public Pre-K regarding its impact on specific academic outcomes, including some evaluations with non-significant findings (see supplemental material – S3 Table 1 for summaries of individual Pre-K evaluations with effect sizes for each academic domain). Notably, studies of the Boston, Los Angeles, and Tulsa programs have shown large impacts on language (effect size ranges 0.44–0.76). Impacts on literacy have tended to be larger (effect size ranges 0.40–1.36); math effects have been mostly in the 0.40–0.59 range.

There is less evidence about the impact of Pre-K programs on children's social-emotional and self-regulatory processes. Extant studies have documented limited benefits of Pre-K on children's social skills and behavior problems (Bartik, 2013; Peisner-Feinberg et al., 2014). The findings have been mixed in the few studies that have examined the impact of Pre-K on children's self-regulatory or executive functioning skills, with documentation of both benefits of Pre-K (Diamond et al., 2007; Gaylor et al., 2019; Shah et al., 2017; Weiland & Yoshikawa, 2013) and null effects (Hofer et al., 2018).

Nevertheless, documented academic benefits are particularly important for children in low-income communities, who tend to lag substantially behind their counterparts from higher-income backgrounds in domains such as math and reading as early as Kindergarten (e.g., Garcia, 2015; Reardon, 2013). In fact, research shows that children in low-income communities appear to benefit the most from Pre-K programs, making larger gains in multiple domains than their more advantaged peers (though children from middle-income households also benefit; e.g. Gormley et al., 2005, 2008; Loeb et al., 2007; Magnuson & Shager, 2010; Weiland & Yoshikawa, 2013). Unfortunately, children from families with low incomes are less likely to attend Pre-K than children from families with middle- and high-incomes, making investments in accessibility that much more critical (e.g., Magnuson & Waldfogel, 2016).

Although a few recent evaluations of PDG programs exist (Grindal et al., 2019; Hofer et al., 2018; Montrosse-Moorhead et al., 2019), rigorous tests continue to be necessary to confirm the efficacy of recent public investments, particularly given that effects of these large-scale public programs tend to be smaller than those of research-driven model programs (Camilli et al., 2010; Duncan & Magnuson, 2013). Often, the large scale studies do not include data on children's family backgrounds, nor do they address child outcomes beyond academic achievement. Moreover, expanding the literature on the impact of Pre-K with samples comprised largely of children from minoritized groups (e.g., Black, Hispanic) may be particularly important (Iruka et al., 2017; Meek et al., 2020). Equally as important is to examine Pre-K effects on children from families with low incomes who are exposed to high levels of adversity (Phillips, 2016; Phillips et al., 2017).

Further, there has been much attention to the methodologies that have been utilized to evaluate Pre-K impacts. The age cutoff variant of the regression discontinuity (RD) design has become common when randomized controlled trials (RCTs) are not ethically or logistically possible, such as when evaluating a state-wide program and specifically when estimating the effects of PDG investments (Cook et al., 2008; Hofer et al., 2018; Lee & Lemieux, 2010). Age-cutoff RD is a quasi-experimental design that relies on pre- and post-intervention measures and the use of a predetermined threshold (i.e., the age cutoff) that determines who receives the intervention in a given year. Participants closest to the cutoff (i.e., those who are the closest in age but who experience the program in different years due to their birthdays, relative to the cutoff) are compared to estimate the average causal treatment effect. RD has proven to be a useful tool in the education research arsenal to evaluate the short-term effects of Pre-K programs (e.g., Gormley et al., 2008; Hustedt et al., 2015; Lipsey et al., 2011; Manship et al., 2015; Peisner-Feinberg & Schaaf, 2011; Weiland & Yoshikawa, 2013; Wong et al., 2008).

However, more recently, Lipsey et al. (2015) identified potential pitfalls of prior RD evaluations and called for future RD studies to be implemented with more rigor (see also Barnett et al., 2018). Among other suggestions, these authors argue that RD evaluations of preschool programs should include: 1) a test of baseline equivalence of treatment and comparison groups (i.e., baseline data collection on treatment group); 2) an examination of attrition bias; 3) the same measurement and procedures for treatment and control groups; and 4) the collection of data on the counterfactual condition (e.g., children's prior care experiences). As we detail below, in the present study, we provide additional evidence on the benefits of public Pre-K for students from urban, minoritized backgrounds characterized by multiple environmental risks, employing an RD design that entails enhancements to the more typical methodological approaches.

1.1. The current study

The current study addressed several of the gaps in the literature, including focusing on children from minoritized, high-risk backgrounds, collecting data on family demographics and child executive functioning, and implementing advancements in the RD design. Using Maryland's strict age requirement for Pre-K eligibility (i.e., age 4 by September 1 of the Pre-K year), we examined whether attending Pre-K, in a public school setting that received Maryland's PDG funds, improved children's school readiness in a variety of domains relevant to later school success (Barnett et al., 2018; Phillips et al., 2017). Specifically, we estimated how Pre-K influenced children's abilities in language, early literacy, early numeracy, and executive functioning (i.e., inhibitory control) at Kindergarten entry, all of which are predictive of success in Kindergarten and beyond (e.g., La Paro & Pianta, 2000, Ng et al., 2014).

This evaluation of Baltimore's Pre-K program advances the literature on Pre-K impacts in several important ways. First, given the particular importance of Pre-K for children from families with low-incomes, we elected to conduct the study in Baltimore City public schools, in which 55% of students are currently counted as low-income. This is likely an underestimate because the school district does not collect applications for free and reduced meals, and the estimate is based on enrollment in programs not available to some families due to immigration status (City Schools; D.A. Baltimore City Schools, 2019). Moreover, we were interested in evaluating the impact of Pre-K using a sample that included a larger proportion of children from minoritized, low-income backgrounds than many of the samples in other evaluations of Pre-K.

Specifically, the sample was drawn from a set of Baltimore City public schools in which 75.3% of the students are Black. Thus, we could address the impacts of Pre-K on a group of children who experience inequities in the early childhood arena (Meek et al., 2020). Further, Baltimore City is an urban area where violent crime far exceeds the national average (United States Department of Justice, 2019), which reflects the structural and systemic adversities and inequities prevalent in the contexts of young children in this city (Ellis & Dietz, 2017). Thus, Baltimore is distinct from other cities where RD Pre-K evaluations have been conducted. For example, according to the Uniform Crime Reports (UCR) in 2017 (United States Department of Justice, 2021), the year of the current study, there were 12,430 violent crimes in a total population of 613,217 in Baltimore, MD. In contrast, during the time of the Boston study (Weiland & Yoshikawa, 2013), the UCR in 2009 reported that there were 6192 violent crimes in a total population of 624,222. Similarly, during the time of the Tulsa study (Gormley et al., 2005), the UCR for Tulsa County reported that there were 4879 violent crimes for a total population of 587,353 in 2002. Further, in an examination of a randomly selected subgroup of parents (n = 269) of children participating in the current study, 50% of respondents reported being exposed to community violence as children, 31% reported being exposed to violence and crime in their lifetimes, and 30% reported that their young children had been exposed to violence and crime (Jones Harden, 2021; Lee et al., 2023). These contextual characteristics are critical to consider given the evidence on the negative effects of violence, crime, and other adverse neighborhood characteristics on young children (Burdick-Will, 2016; Kohen et al., 2008). Thus, this study addressed an important research direction for preschool research – examining Pre-K effects on children from families with low incomes who are exposed to high levels of contextual and structural adversity (Phillips, 2016; Phillips et al., 2017).

Further, we elected to expand the domains that are typically examined in Pre-K evaluations, which tend to focus on academic outcomes such as literacy and math. Specifically, we included a measure of executive functioning (EF), defined as higher order cognitive skills that includes such processes as working memory, inhibitory control, and selective attention (Blair, 2002). EF has been theorized and documented to be a critical predictor of school readiness and academic competence writ large (Fuhs et al., 2015; Perry et al., 2018). Moreover, young children's experience of poverty and material hardship has been found to negatively impact the development of EF (Raver et al., 2013; Ursache et al., 2015), both of which are strong predictors of young children's school readiness (Perry et al., 2018). Despite this evidence on EF, few Pre-K evaluations have included it as an outcome. The limited evidence does suggest that Pre-K participation may enhance children's executive functioning (Bai, Ladd, Muschkin, & Dodge, 2020; Diamond et al., 2007; Shah et al., 2017).

We also had a goal of implementing a more rigorous RD design. To achieve this goal, we followed several recommendations delineated by Lipsey et al. (2015). Our methodological enhancements included use of equivalent measures for both cohorts, the inclusion of carefully chosen baseline covariates including consideration of the counterfactual condition (i.e., prior-care experiences, demographic characteristics of families), and the use of baseline measures on the experimental cohort to ensure that groups were equivalent prior to Pre-K entry (Lipsey et al., 2015). Furthermore, we used measures of classroom quality to describe the Pre-K experiences of our sample and facilitate interpretation of our findings within the broader literature. Finally, as is standard in this literature, we conducted sensitivity analyses to increase confidence in our results, including varying the bandwidths around which analyses are conducted and examining multiple functional forms of the relationships under investigation (Lipsey et al., 2015).

Following this approach, our overarching research question was: *What is the effect of Pre-K on students' language, literacy, mathematics, and executive functioning skills?* Consistent with the literature on Pre-K effects, we hypothesized that children in the Pre-K treatment group would perform better on assessments of language, early literacy, and early mathematics problem-solving than children in the comparison group. Because the measurement of executive functioning is less prevalent in the Pre-K evaluation arena and extant findings are mixed, we wanted to examine whether there would be impacts on this important developmental domain (Fuhs et al., 2015). Additionally, we wanted to explore whether the magnitude of effects documented in other Pre-K evaluations would hold for this group of children in Baltimore City, who represented a sample from more diverse and high-risk backgrounds than what is typically documented in such studies. Although our statistical power was limited, we also explored whether effects differed by parent education, child gender, or child race/ethnicity.

2. Method

2.1. Sample and setting

Following the procedures outlined by Lipsey et al. (2015) and Weiland and Yoshikawa (2013), we recruited two groups of students using the age-based cutoff specified by the state of Maryland. The setting for this study was Baltimore, MD, specifically 15 schools that had received Pre-K expansion funds. Students who were born before September 1, 2012 comprised Cohort 1 (the treatment group who received Pre-K services as a part of the expansion in the fall of 2016). Those born between September 1, 2012 and August 31, 2013 comprised Cohort 2 (the comparison group who did not receive Pre-K services until fall of 2017).

Cohort 1 (treatment group) consisted of 554 Pre-K students and Cohort 2 (comparison group) consisted of 542 students, for a total of 1097 students. However, the analytic sample for this study (see results section) was 997 due to the exclusion of children in Cohort 1 (treatment group) who did not have Pre-K fall baseline and Kindergarten fall outcome assessments (n = 57), children who were tested outside of the assessment window (fall of academic year; n = 20), and children whose birthdays were beyond the 365 day cutoff for eligibility for Pre-K (n = 18). In addition, five children attended Pre-K in fall 2016 (Cohort 1), even though their birthdays were after the cutoff; these children were also removed from analyses. In our sensitivity analysis, we refit key models including these students and show that results are stable.

Demographic characteristics of the two cohorts of children are presented in Table 1. Participant children in the two groups were similar across all variables apart from child age, as was expected because the two cohorts were one year apart in age. However, the two cohorts were virtually the same age at the time of Pre-K eligibility (i.e., Cohort 1′s average age = 4.51 years (SD = 0.30); Cohort 2′s average age = 4.53 years (SD = 0.29)). About half of the children were female. The majority of children were Black and spoke English at home. Approximately 1/5 of the children were Hispanic. About 2/3 of the parents had a high school diploma or less.

2.2. Procedures

As in other studies using the age cutoff RD design (e.g., Grehan et al., 2011; Weiland & Yoshikawa, 2013), we collected data from two cohorts of children, one serving as the treatment group and the other serving as the comparison group. Students in Cohort 1 were assessed at the start of their Pre-K year (fall 2016), which reflects the first of our methodological enhancements. Subsequently, both groups were assessed in fall 2017 using the same measures (i.e., at Pre-K entry for the comparison group and at Kindergarten entry for the treatment group). We used a set of standardized norm-referenced measures designed to measure early literacy, language, early numeracy, and executive functioning skills. To improve precision and to test that the two cohorts were equivalent at baseline, we also measured many other characteristics of the participating families via a combination of administrative records and a parent survey (e.g., child gender and age, home language, race or ethnicity, caregiver education, and number of siblings).

There was one important school system policy change between the Pre-K years for the two cohorts. Specifically, the income eligibility requirement was revised for the 2017–2018 school year from 200% to 300% of the federal poverty level or below. To address this potential difference, we only sampled students from Cohort 2 that matched students from the prior cohort in terms of income. Specifically, we only included Cohort 2 students if they would have been eligible in the prior year (i.e., they came from families who reported annual incomes that are no more than 200% of the federal poverty level or below). We return to this policy shift when we test for differences in baseline characteristics by treatment status and in our limitations section.

A total of 15 schools participated in this project (100% of schools who received requests to participate), with a total of 41 Pre-K classrooms across the two cohorts. These schools were selected by the Baltimore City Public Schools (BCPS) Director of Early Learning for inclusion in our study, given their receipt of PDG expansion funds. The principal of each school gave us permission to recruit students and identified a private location for the testing to occur for individual children. Parents were approached before and after school to request their permission for their children to participate in the study. If parents did not transport their children, consent forms were sent home via children's backpacks, with a request for parents to return signed forms to school. Only children whose parents signed consent forms participated in the study. For cohort 1, we were able to obtain consents for 77.93% of children at Time 1, and 83.69% of children at Time 2. For cohort 2, 81.37% of the parents consented to their children's participation.

Data were collected by a team of six full-time post-baccalaureate research assistants, and 20 temporary research assistants (combination of undergraduate and graduate students). The post-doctoral researcher, who coordinated this project, provided intensive training and supervision of all the research assistants throughout data collection in test administration, engaging students, and collaborating with schools. All testing occurred within a three-month window from the start of school in the fall. Testing was administered to individual children over a 60-minute period, with breaks given for children who needed them. Children were given a small toy at the end of testing.

As stated previously, we collected an array of baseline covariates that we used to assess initial differences between the two cohorts (the second methodological enhancement offered by our study). We collected basic demographic data from a parent survey (e.g., gender, age, race or ethnicity, family SES). Parents typically completed the survey at school when they dropped off or picked up their children. If parents did not transport their children to school, they were sent surveys via their children's backpacks, completed them at home, and returned them via the same method (which our research team retrieved from teachers). These surveys were administered around the time of the baseline assessment (i.e., fall/winter of 2016 for Cohort 1 and 2017 for Cohort 2). Parents received $10 for completing this survey. For cohort 1, 79.6% of parents returned the surveys, and 99.3% of parents in cohort 2.

2.3. Assessments and measures

Students from both cohorts were assessed at Pre-Kindergarten entry (and Cohort 1 at Kindergarten entry) using standardized assessments of language, literacy, and numeracy skills, described below. English versions of the language and academic functioning assessments were administered to both English and Spanish speaking students, as was requested by the funder. However, we did provide instruction for the behavioral self-regulation task (i.e., HTKS) in Spanish. Additionally, for the first cohort, we used different forms of the measures at pre- and post-test, so that children would not be exposed to the same test twice. To repeat the same procedure with cohort two, we varied the forms we used with children during the pretest as well. Following recommendations by Lipsey et al. (2015), to reduce the likelihood of differential operationalization of measures across the different age ranges, we began testing on the academic measures with the same item even though, according to the rules of those assessments, students in Kindergarten would start at a later item (i.e., the third methodological enhancement). We provide more detail on the influence of start rules in the robustness check section.

We also assessed the classroom quality of the classrooms in the study for Cohort 1 (children who experienced Pre-K) using the CLASS tool (Pianta et al., 2008; see below). A trained and Pre-K CLASS certified post-doctoral researcher conducted the CLASS assessments during the morning (e.g., 9 AM-12 PM) of the school day. Each classroom was assessed one morning, across four 20-minute cycles of observation. The assessments were conducted during the late fall of the academic year.

2.3.1. Receptive vocabulary

The Peabody Picture Vocabulary Test, Fourth Edition (PPVT; Dunn & Dunn, 2007), a widely used assessment of receptive vocabulary, was administered to participant children. The examiner orally presents a stimulus word with a set of pictures and the child is asked to select the picture that best depicts the word. The PPVT has been found to have excellent psychometric properties, with high reliability (in the 0.90s) and validity (highly associated with tests of intelligence), and has been used widely with children from low-income, minoritized backgrounds.

2.3.2. Early literacy

We administered three subtests that traditionally comprise the Broad and Basic Reading subscales of the Woodcock-Johnson IV (W-J; Schrank et al., 2014). These subtests have good psychometrics (Schrank et al., 2014) and have been widely used with populations of young children from low-income and minoritized backgrounds (Gormley et al., 2005; Weiland & Yoshikawa, 2013). Administered W-J subtests were: Letter-Word Identification (children identify letters or read aloud individual words); Word Attack (children apply phonic and structural analysis skills to pronounce unfamiliar printed words); and Passage Comprehension (children use syntactic and semantic clues to identify a missing word in text).

2.3.3. Early numeracy/mathematics

We also administered the Applied Problems subscale of the W-J, a numeracy and early mathematics assessment that requires children to perform relatively simple calculations to analyze and solve arithmetic problems. The authors report good psychometrics (Schrank et al., 2014) and this subscale has been widely used with young children from low-income and diverse backgrounds (Gormley et al., 2005; Weiland & Yoshikawa, 2013).

2.3.4. Executive functioning

Finally, we administered the Head, Toes, Knees, and Shoulders (HTKS; Ponitz et al., 2008), a direct assessment of children's behavioral self-regulation. This assessment includes components of executive functioning such as ability to focus attention, working memory, and inhibitory control. In this 20-item assessment, children are asked to play a game in which they do the opposite of what the experimenter says. The authors report good construct and predictive validity for this measure (Ponitz et al., 2009). Further, multiple studies have documented its utility in studies of children from low-income and minoritized backgrounds (Caughy et al., 2022; McClelland & Tominey, 2014; Yu et al., 2020).

2.3.5 Classroom quality

Classroom quality was observed and assessed via the Classroom Assessment Scoring System (CLASS) Preschool version (Pianta et al., 2008). A trained post-doctoral level researcher conducted observations of teacher-child interactions and scored them ranging from 1 = *low* to 7 = *high*. There are three domains (and 10 dimensions/subscales) on the Preschool CLASS: Emotional Support; Classroom Organization; and Instructional Support. The CLASS has good internal reliability with dimensions being significantly associated with one another in the expected ways. The CLASS also exhibits construct validity and is related to other measures of the quality of early educational settings (e.g., ECERS, Pianta et al., 2008).

Covariates and care settings before Pre-K. We developed a brief parent survey for this project to gather demographic and other background information on participant children (e.g., parent education, race/ethnicity, language). We also asked about children's prior care experience, specifically the types of settings in which they spent 10 hours or more per week the year before entering Pre-K (e.g., care by parents/caregivers, relative care, family child care, center-based care, and Head Start). The survey was translated into Spanish for Spanish-speaking parents.

2.4. Data analytic strategy

To examine the impacts of this federally funded PDG program, we used Regression Discontinuity (RD) analyses to measure group differences in children's scores on the assessments. We took advantage of the natural division of children into cohorts based on whether their birthdates fell before or after the cutoff for Pre-K eligibility. Cohort 1 (treatment group) children turned 4 years old before the September 1, 2016 cutoff, whereas Cohort 2 (comparison group) children did not. Because the cutoff was a state requirement, we assume that for children immediately on each side of the cutoff, their assignment to attend the Pre-K program that year was unrelated to characteristics of the child, the child's family, or their circumstances that would also influence children's Kindergarten assessment scores. If this assumption is correct and the current study meets the requirements of RD, any differences post-program in assessment scores between children whose birthdates just fell to one side of the cutoff versus the other provide unbiased estimates of the causal impact of the PDG program on children's school readiness. Using the standard sharp RD approach, data from children whose birthdates were further from the cutoff were used to estimate more precisely the value of each outcome near the cutoff (Imbens & Lemieux, 2008; Shadish et al., 2002; Weiland et al., 2017).

To implement the RD approach, we began by fitting plots of the relations between children's age and each outcome variable, using linear and non-parametric regression. These plots helped us determine the functional form for each outcome and allowed us to visually inspect whether there appeared to be a "jump" at the cutoff. We then fit a series of regression models in which each outcome variable was predicted by: (a) a dichotomous variable denoting the child's cohort (i.e., whether the child was eligible for Pre-K entry in Fall 2016 based on birthdate); (b) the child's age centered around the state-determined cutoff point (i.e., the distance in days between when the child turned 4 years old and September 1, 2016); and (c) relevant covariates (i.e., school fixed effects, testing date and form, child sex, whether child is Black, whether child is Hispanic, whether child speaks only English at home, parent age, whether parent graduated high school, whether parent is the mother, whether parent works fulltime, number of adults living at home, number of children living at home, whether birth mother lives in the home, whether birth father lives in the home, and four dummy variables indicating all early care settings in which the child spent 10 or more hours per week before entering Pre-K). The cohort coefficient and its associated *p*-value was the parameter of interest for determining the effect of the PDG program on each outcome.

As stated above, data from children whose birthdates were further from the cutoff were used to estimate more precisely the value of each outcome near the cutoff. To do this, we fit each model with different bandwidths of children (i.e., children with birthdates within 365, 180, and 90 days on either side of the cutoff, common bandwidths used in other RD studies; Gormley et al., 2005; Weiland & Yoshikawa, 2013; Weiland et al., 2017) in order to examine the sensitivity of our results to choice of bandwidth (Lee & Lemieux, 2010). We examined the stability of the parameter of interest (i.e., the cohort coefficient and its associated *p*-value) across the various bandwidths to determine the robustness of the results. This approach balances bias (i.e., influence from children far from the cutoff) with precision (i.e., a larger bandwidth provides a larger sample; Weiland et al., 2017). The results from models using all three bandwidths (BWs) are reported in the results section. In addition, to specify the correct functional form of the relation between each outcome and child age, we fit a series of additional models for

Table 2
Classroom assessment scoring system[a] means, SDS, and ranges for sample classrooms.

Domain	N = 37 classrooms [b]	
Dimensions	M (SD)	Range
Emotional Support	5.02 (0.40)	4.13 – 5.69
Positive Climate	4.92 (0.58)	3.50 – 5.75
Negative Climate	1.43 (0.41)	1.00 – 2.56
Teacher Sensitivity	4.95 (0.59)	4.13 – 5.75
Regard for Student Perspective	3.64 (0.39)	3.17 – 4.25
Classroom Organization	4.77 (0.52)	3.77 – 5.86
Behavior Management	4.81 (0.64)	3.94 – 6.08
Productivity	5.04 (0.59)	4.00 – 6.58
Instructional Learning Formats	4.47 (0.53)	3.38 – 4.92
Instructional Support	2.43 (0.29)	1.96 – 3.04
Concept Development	2.04 (0.33)	1.31 – 2.58
Quality of Feedback	2.54 (0.29)	2.06 – 3.13
Language Modeling	2.72 (0.43)	2.08 – 3.75

Note. All scales range from 1 to 7. *M* = Mean. *SD* = Standard deviation.
[a] Pianta et al., 2008.
[b] Three classrooms not assessed due to failure to obtain consent.

each bandwidth with polynomial specifications and interaction terms. We compared fit statistics across models and over-specified the models as a robustness check (Trochim, 1984). Finally, we conducted extensive sensitivity analyses to determine the sensitivity of findings to alternative model specifications (see the robustness checks section). Our sensitivity analysis also included examining robustness of impact estimates to a data-driven bandwidth selection procedure and non-parametric analytic approach (i.e., rdrobust; Calonico et al., 2017).

In our exploratory subgroup analyses, we extended our analytic approach to examine whether effects varied by parent education (high school or less versus more than high school), child gender, and child race/ethnicity (Black, Hispanic, Other). We included interaction terms that allowed us to test whether treatment effects were statistically significant different by subgroup. We selected these subgroups based on power constraints and alignment with prior literature (Gormley et al., 2005; Weiland & Yoshikawa, 2013). We limited this analysis to bandwidths of 365 and 180, as bandwidths of 90 were too underpowered.

Analyses were conducted using Stata version 16 and SPSS version 24. For all analyses, we used multiple imputation (with 40 imputations) to account for missing data. (See Table 1 for summary statistics, including the percent missing for each outcome). The imputations were conducted using a parametric model that assumes all variables follow a multivariate normal distribution. In all regression models, we conducted inference using the multiple imputation small-sample correction of Barnard and Rubin (1999).

3. Results

3.1. Descriptive statistics

Descriptive statistics, including sample means and standard deviations or percentages for child outcomes and all covariates, are shown in Table 1. As noted earlier, the sample is majority Black from low-income backgrounds. We also present CLASS scores for the preschool classrooms that were part of this study in Table 2. Of the 41 participating classrooms, we were able to obtain teacher consent and availability for 37 classroom assessments (i.e., 90% of classrooms). On average, the domains of Emotional Support and Classroom Organization were scored in the mid-range (i.e., 5.02 and 4.77 respectively). For the domain of Instructional Support, the average scores of observed classrooms were in the low range (i.e., 2.43).

Counterfactual. Parents of children in the comparison group reported the following types of early care settings in which their child spent 10 or more hours per week the year before entering Pre-K: at home with a primary caregiver (48%); informal care (e.g., with a relative or in family child care homes) (23%); Early Head Start (26%); and other center-based care settings (17%) (Note: percentages sum to over 100% due to individual children having multiple child care arrangements). When we limit the sample just to those who missed entry into Pre-K by less than one month (a bandwidth of 30 days), the distribution of early care settings is the following: at home with a primary caregiver (54%); informal care (6%); Early Head Start (35%); and other center-based care settings (19%) (see Table 3 for percentages of early care settings for each bandwidth). Accordingly, across bandwidths, the treatment is being compared to a counterfactual in which about 43–54% of comparison children were enrolled in an alternative center-based preschool option.

3.2. Regression discontinuity analyses

3.2.1. Primary results

As explained in our analytic section, we began our analysis by examining plots of the relations between children's age and each outcome, using both linear regression and non-parametric regression. These plots (see Fig. 1 for two examples and Appendix S1 Figure S2 for all other plots) helped inform our functional form decisions and also allowed us to visually inspect whether there was a "jump" (i.e., a treatment effect) at the cutoff. We then moved to fitting impact models. As shown in Table 4, we found that the PDG program had a statistically significant and positive effect on all assessments when the models included a bandwidth of 365 days (range in effect sizes of $d = 0.36$ to 1.10 standard deviations). However, the cohort (i.e., treatment status) coefficient was only statistically significant and stable across bandwidths for the PPVT-IV (receptive vocabulary), W-J-IV Letter-Word Identification (early literacy), W-J-IV Word Attack (early literacy), and W-J-IV Applied Problems (early mathematics) assessments, but was not for the HTKS (executive functioning) and W-J-IV Passage Comprehension (early literacy) assessments. Effect sizes for the PPVT-IV assessment ranged from 0.41 to 0.74 across bandwidths. For the W-J-IV Letter-Word Identification assessment, effect sizes ranged from 1.10 to 1.36 across all bandwidths. Effect sizes for the W-J-IV Word Attack assessment ranged from 0.99 to 1.28 across all bandwidths. Finally, effect sizes for the W-J-IV Applied Problems assessment were 0.54 to 0.71 across all bandwidths.

3.2.2. Subgroup analysis

For parsimony, we summarize our subgroup findings in Fig. 2; impact estimates and *p*-values for the 365 and 180 bandwidths are in Appendix S2 Table 1. We found that Hispanic and other race/ethnicity children benefitted more on vocabulary and math than Black children ($p < .05$). Other tested differences were not statistically significant and/or were not robust to bandwidth. Again, given our power limitations, we view these findings as limited and suggestive only.

3.2.3 Robustness checks

We conducted extensive robustness checks to examine the internal validity of our results, as is recommended practice for RD designs (e.g., Lee & Lemieux, 2010, and the *What Works Clearinghouse* guidelines; Schochet et al., 2010) and following other Pre-K RD studies (e.g., Weiland & Yoshikawa, 2013). Threats to the internal validity of our results could include: (1) treatment misallocation at the cutoff; (2) non-smooth or discontinuous variation in observed or unobserved child or parent characteristics around the cutoff; (3) discontinuities in the outcomes at points other than the cutoff; (4) incorrect specification of the functional form of the relation between outcome and child age; (5) sensitivity of the results to the choice of bandwidth around the age cutoff; (6) sensitivity of the results to the choice of covariates included in the analytic models; (7) sensitivity of the results to a data-driven bandwidth and parametric versus non-parametric analytic approach; (8) the accumulation of Type I error as a result of multiple tests being conducted; (9) sensitivity of the results to the use of different start rules on the PPVT-IV and W-J-IV subtests; (10) sensitivity of the results to the use of raw

Table 3
Sample means of comparison group children who attended each early care setting by child birthdate (bandwidth around the cut-off).

	BW = 30 (N = 48) M	BW = 90 (N = 153) M	BW = 180 (N = 277) M	BW = 365 (N = 521) M
Early care – at home with primary caregiver	.54	.49	.46	.48
Early care – informal (e.g., relative) care/family child care homes	.06	.18	.20	.23
Early care – Early Head Start	.35	.30	.31	.26
Early care – center-based	.19	.18	.18	.17

Note. BW = bandwidth; column percentages totals are greater than 100% because individual children could have more than one child care arrangement.

scores rather than standardized scores on the PPVT-IV and W-J-IV subtests; and (11) sensitivity of the results to the choice of analytic sample (i.e., the exclusion of Cohort 1 children who did not have Pre-K fall baseline and Kindergarten fall outcome assessments, as well as the exclusion of children who were tested outside of the assessment window).

In addition, our 12th robustness check is, to our knowledge, the first to implement a version of one of the recommendations of Lipsey et al. (2015). Specifically, we compared the average pretest scores in Pre-K fall for both cohorts to examine overall equivalency of the two cohorts. Note that we could not test for balance in pretest scores at the cutoff, as treatment-group children had just turned four at the time of pretest (their Pre-K fall) while comparison-group children had just turned five at the time of pretest (their Pre-K fall, one year later). Despite this limitation, this check was useful in examining the *overall* comparability of the two cohorts.

Across these checks, we found little evidence of threats to the internal validity of our results (see Supplemental Material for details). However, a few findings from these analyses warrant discussion, and are addressed in the discussion section. Specifically: (a) cohorts differed at the cutoff point in terms of testing date and executive functioning assessment Form type, and thus these imbalanced variables were included as covariates in all models predicting those outcomes; (b) results for the W-J-IV Passage Comprehension subtest are sensitive to the use of raw scores versus standardized scores; and (c) pretest scores for the two cohorts are not balanced on average (though again, we could not test whether they were balanced at the cutoff).

4. Discussion

This study contributes to the growing evidence base on the impact of Pre-K programs on young children's school readiness. Specifically, our findings provide further evidence that Pre-K experiences can enhance the academic functioning of children from low-income backgrounds, particularly in the areas of receptive vocabulary, early literacy (i.e., decoding), and early mathematical problem solving. These impacts are noteworthy given that nearly half of the children in the comparison group had been exposed to other center-based programs. This "counterfactual" (i.e., comparison children experiencing other types of center-based programs) has been found to diminish findings of intervention effects in other evaluations of preschool programs, including Head Start (Zhai et al., 2014).

The current study reveals that the PDG program had a statistically significant and positive effect on all assessments. However, the findings were more robust for receptive language (i.e., PPVT), pre-literacy skills (i.e., W-J Letter Word Identification and Word Attack), and pre-numeracy skills (i.e., W-J Applied Problems). Notably, these are skills that were explicitly targeted by the curriculum used in Baltimore City. In particular, the curriculum focused on pre-literacy skills, which showed effect sizes equivalent to about 70–90% of a year's worth of additional learning in early reading, based on empirical benchmarks in the field (Hill et al., 2008). This set of findings align with a preponderance of evidence of the benefits of Pre-K with respect to both early literacy and mathematics domains of school readiness (Phillips et al., 2017; Yoshikawa et al., 2013). Further, these findings are consistent with other studies of public Pre-K which suggest that these programs tend to have larger effects on literacy skills and smaller effects on math and language (see summary of table of Pre-K evaluations in supplemental materials; S3-Table 1).

A relatively novel feature of the current study was to examine Pre-K impacts on children's self-regulation. Our lack of robust findings for executive functioning (i.e., behavioral self-regulation; HTKS) may be due to the validity of the measure for our study population (e.g., 45% of participant children had a score of 0 at baseline; 42% for Cohort 1 children and 48% for Cohort 2 children). Further, the findings in this area may be attributable to its relative lower emphasis in the curriculum, which tended to focus more on literacy development. Extant evidence is inconsistent regarding the impact of Pre-K programs on executive functioning and other regulatory processes (Gaylor et al., 2019; Hofer et al., 2018; Weiland & Yoshikawa, 2013). Self-regulation may be enhanced if there is an explicit focus on children's social-emotional and regulatory skill development (e.g., Moore et al., 2015; Yoshikawa et al., 2013). Notably, Baltimore City Public Schools is currently placing more emphasis on social-emotional skills in their early childhood education programs (D.A. Baltimore City Schools, 2019).

Similarly, the lack of robust findings for Passage Comprehension may be related to our methodological concern about this subtest for our sample. As noted previously, this subtest of the W-J is particularly sensitive to the use of standard scores vs. raw scores (the latter of which were used in the current study). Further, there is research on elementary aged children that suggests that this subtest is especially sensitive to children's decoding skills and developmental phase, both of which are obviously in the lower range for children in Pre-K (Keenan et al., 2008). Additionally, although there were benefits of Pre-K to passage comprehension in the Tulsa study (Gormley et al., 2005), this subtest is not commonly used in Pre-K evaluations. Given the role of Pre-K programs in promoting early literacy (Yoshikawa et al., 2013) and the importance of comprehension, future evaluations should explore including other tests of this critical domain.

Unique to this project, we examined Pre-K impacts on children from very marginalized backgrounds, unlike the Boston and Tulsa samples, which included middle-class children (see Gormley et al., 2005; Weiland & Yoshikawa, 2013). For example, most children in the current sample had mothers with a high school diploma or less, and less than half of the mothers were employed. Although we do not have data on other risk factors for study participants, the PDG funding was only given to schools that were in very low-income and high-risk neighborhoods that were characterized by the aforementioned violent crime rates in the city of Baltimore (Baltimore City Schools Early Learning Division, personal communication). Using the RD design, the current study extends and confirms the findings of Williams et al. (2019), which documented the benefits of Pre-K for children in Baltimore, MD. Such findings suggest that although early childhood education programs are not designed to reduce the contextual risks that some participant children experience (e.g., violence), they may serve as buffers against the adversities that these children experience (Phillips et al., 2017).

Notably, in our exploratory analysis, we found no evidence of heterogeneity by parent education or child gender. The lack of variability in our sample (i.e., 2/3 of the parents in the current study had a

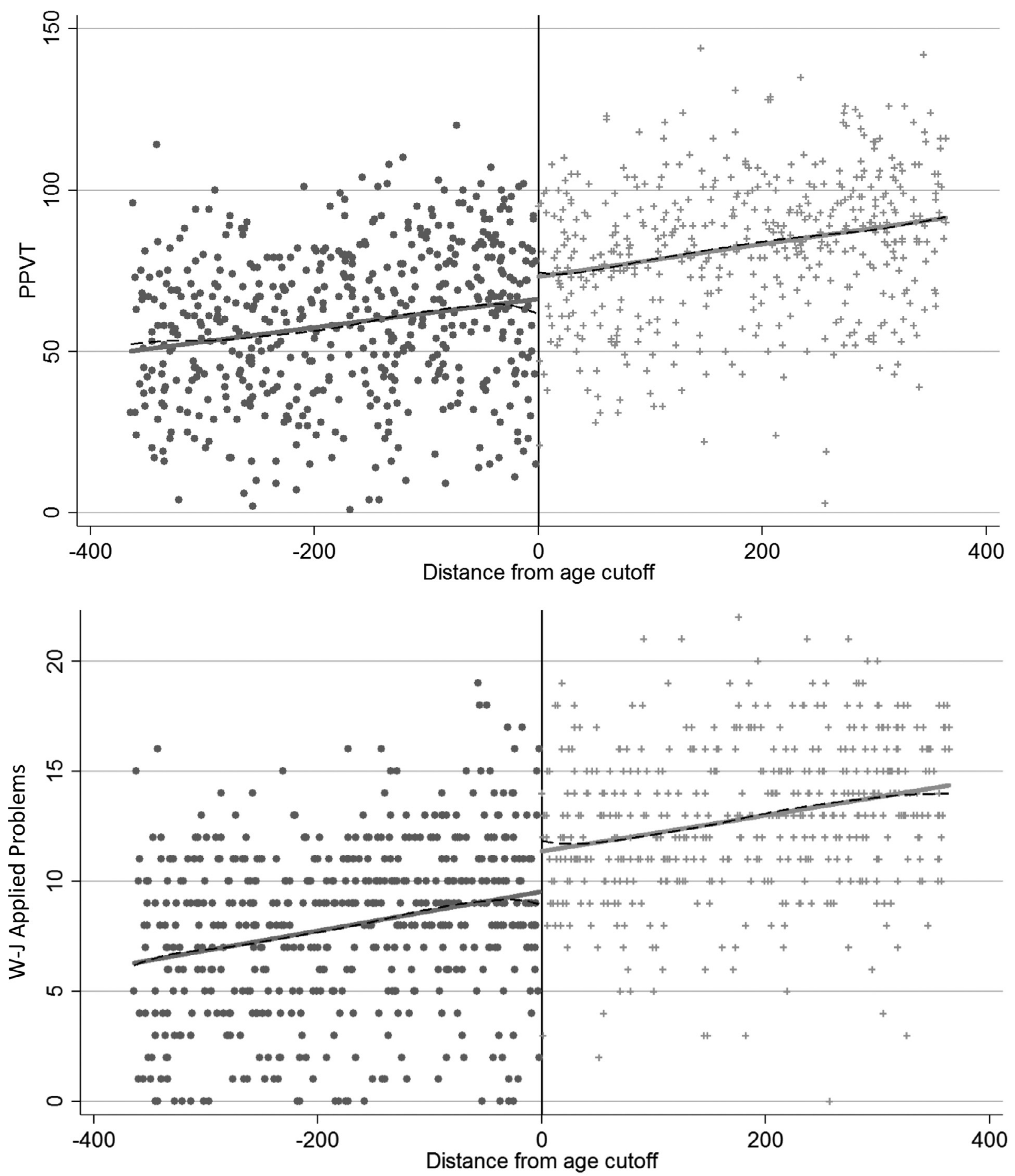

Fig. 1. Example linear and non-parametric plots.
Note. $n = 997$ children for BW365; PPVT = Peabody Picture Vocabulary Test IV (receptive vocabulary); W-*J* = Woodcock-Johnson Tests of Achievement IV. Solid lines = linear regression; dotted lines = non-parametric regression.

high school diploma or less) may have contributed to the lack of heterogeneity regarding parent education. The gender findings align with meta-analytic evidence showing that, on average, boys and girls benefit equally from preschool (Magnuson et al., 2016). However, some studies have documented different long-term outcomes based on gender, favoring boys (Hill et al., 2015) as well as girls (Anderson, 2008). There was heterogeneity by race/ethnicity, in ways that largely also match prior evidence. That is, Hispanic and other race/ethnicity children benefitted more on vocabulary and math than their peers, as they have in many other studies (Mendez Smith et al., 2021). For example, impacts for Hispanic children were greater than for children from other racial/ethnic groups in vocabulary and math when they attended Head Start (Bloom & Weiland, 2015; Puma et al., 2010), in literacy and problem-solving skills when they were enrolled in the Tulsa Pre-K program (Gormley et al., 2005), and in math when they participated in the Boston Pre-K program (Weiland & Yoshikawa, 2013). Although some studies have found

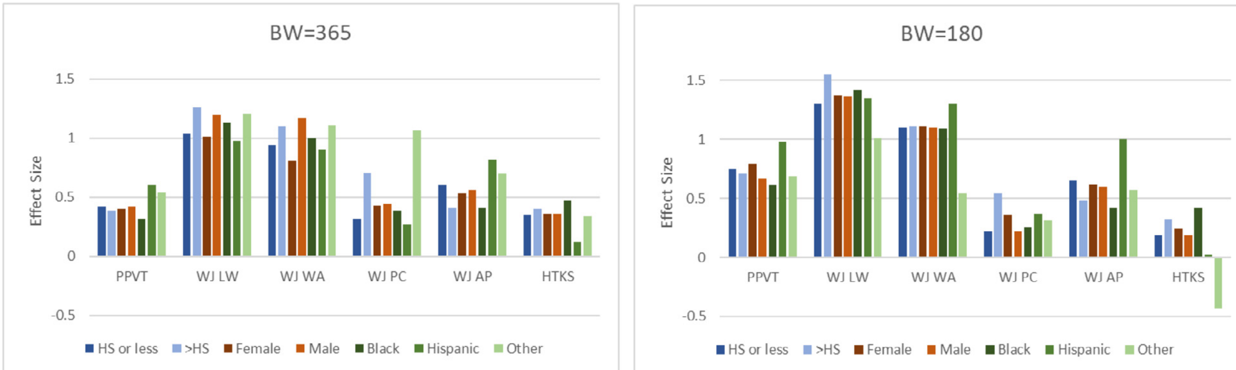

Fig. 2. Heterogeneity in effect size by parent education, child gender, and child race/ethnicity.
Note: HS = High school. Black-Hispanic differences on vocabulary and math was statistically significant and robust across bandwidth ($p < .05$). All other subgroup differences were not statistically significant and/or not robust to choice of bandwidth (see Appendix S2 Table 1).

greater effects for Black children (e.g., Bassok, 2010), there is limited research that focuses on children from backgrounds with similar risks as was the focus of the current study. We underscore again, however, that our subgroup analyses were under-powered and exploratory. More work is needed on the mechanisms behind *why* prekindergarten programs have differential effects (Phillips et al., 2017). Additionally, Pre-K impacts should be assessed with other measures of child functioning that may have more cultural salience for minoritized populations (Iruka et al., 2017).

Another contribution of the current study is our data on children's experiences in Pre-K; other studies did not have these data or only obtained district-wide or control group data on classroom quality (e.g., Weiland & Yoshikawa, 2013). Although imperfect, having classroom quality data from the CLASS allows for comparisons of the effects of this program to other Pre-K programs which have been studied. Specifically, although there was a standard curriculum across the jurisdiction, we found that the quality of the Pre-K classrooms was in the low to middle range based on the CLASS assessment. Our assessment of average classroom quality showed that the Emotional Support and Classroom Organization evident in these classrooms were in the middle range. However, consistent with other research using the CLASS to assess quality of Pre-K classrooms (e.g., LoCasale-Crouch et al., 2007), instructional support tended to be mediocre or low. This finding aligns with other research using the CLASS that suggests that children from low-income backgrounds are exposed to lower-quality classroom practices (LoCasale-Crouch et al., 2007). Although CLASS scores are not consistently related to child outcomes (Weiland et al., 2013), low levels of instructional support have been documented to negatively affect children's cognitive and academic functioning in Pre-K programs (Johnson et al., 2016; Mashburn et al., 2008). Our findings of the early academic benefits of Pre-K, in the context of low instructional support scores, suggests that other measures of classroom quality may be needed, such as those which assess the quality of instruction in specific content areas (e.g., early literacy, early mathematics; Weiland et al., 2013). Future studies should include measures of classroom quality that go beyond global measures to include measures that may be more reflective of the goals of the Pre-K program.

As noted, the current study followed the recommendations of Lipsey et al. (2015) to improve the rigor of RD studies, such as using equivalent measures for both measurement cohorts, collecting baseline measures on the experimental cohort, and conducting multiple robustness checks. Some of our findings illustrate the importance of these enhancements and the practical difficulties of others. For example, as explained above, impacts for passage comprehension were not robust across test score choice. Although we found that there were no differences in our estimates when we applied standard start rules for assessments versus when we did not, we are also able to rule out this potential threat through following guidance about how to improve rigor. Regarding difficulties, we collected data on the treatment group's scores in their Pre-K fall. Although this represented an advancement, it also allowed for the possibility of two opportunities for refusal for cohort 1. Further, we were unable practically to identify and test the comparison group until they entered Pre-K the next fall. Accordingly, we were able to examine *average* differences in baseline test scores between the two cohorts but we were not able to examine balance on baseline scores at the cutoff (i.e., the test needed to examine potential bias at the cutoff). Our findings regarding average differences are nonetheless interesting and suggest some additional caution should be taken in interpreting our results. They also demonstrated that much larger and longer-term efforts, which are likely to be difficult to mount in field-based studies, would be needed to fully follow this important recommendation.

Despite the methodological advances of our study, there are other limitations that should be acknowledged. We discovered that the two cohorts differed at the cutoff point regarding two methodological issues (i.e., testing date, executive functioning assessment form type). We included these imbalanced variables as covariates in our analyses. However, it is important to acknowledge that distinct procedures for the two cohorts could have influenced test scores and to interpret EF findings in particular with more caution. Future studies should ensure to the extent possible that the same procedures are conducted with both cohorts of children.

In addition, although we had basic demographic and prior care data on both cohorts, we were unable to collect other data on family risk and protective factors that may have revealed whether Pre-K experiences were more or less beneficial for specific groups of children. Because the school district was interested in how children fared on English reading skills and due to methodological concerns about using different assessments for children, we did not use Spanish versions of measures (except for the HTKS instructions) for the 19.6% of the children who were from Hispanic households. Clearly, using versions of assessments that are consistent with children's preferred language would be critical for future studies. Further, we only examined outcomes immediately after the Pre-K experience; longitudinal studies have suggested that some of the academic benefits of Pre-K may disappear during the early elementary school years (e.g., Durkin et al., 2022; Lipsey et al., 2018; Weiland et al., 2020). Thus, follow-up studies of Pre-K programs throughout elementary school and beyond are important to conduct.

Additionally, as explained, we found average group differences between the two cohorts on key outcome variables at baseline favoring the treatment group on average. One potential explanation for this finding is that the school district expanded the income eligibility range for Pre-K the year we collected baseline data from Cohort 2 (the comparison group). Although we only recruited children in the sample who were below 200% according to school personnel, it is possible that this policy change resulted in distinct Pre-K recruitment strategies overall for Co-

Table 4
Regression discontinuity analyses: relation between PDG program cohort and outcomes.

Variable	BW = 365	BW = 180	BW = 90
PPVT			
Cohort coefficient	9.61***	17.14***	16.20**
SE	(2.43)	(4.69)	(4.98)
Effect size	.41	.74	.70
W-J LW Identification			
Cohort coefficient	4.92***	6.08***	5.77***
SE	(0.72)	(0.90)	(1.14)
Effect size	1.10	1.36	1.29
W-J Word Attack			
Cohort coefficient	1.67***	1.87***	2.16***
SE	(0.29)	(0.36)	(0.49)
Effect size	.99	1.11	1.28
W-J Passage Comprehension			
Cohort coefficient	.92**	.64	1.12
SE	(0.35)	(0.44)	(0.57)
Effect size	.44	.30	.53
W-J Applied Problems			
Cohort coefficient	2.04***	2.29***	2.66**
SE	(0.43)	(0.61)	(0.88)
Effect size	.54	.61	.71
HTKS			
Cohort coefficient	4.29*	2.59	.47
SE	(1.80)	(2.59)	(3.81)
Effect size	.36	.22	.04

Note. n = 997 children for BW365, n = 501 children for BW180, and n = 280 children for BW90; n = 15 schools for all BWs; RD = regression discontinuity; BW = bandwidth in days; PPVT = Peabody Picture Vocabulary Test IV (receptive vocabulary); W-*J* = Woodcock-Johnson Tests of Achievement IV; LW = Letter Word; HTKS = Head-Toes-Knees-Shoulders (executive functioning). The functional form of the relation between child age and outcomes was specified as linear across bandwidths, with a few exceptions. In accordance with our functional form testing, we included a quadratic child age term for HTKS BW = 365, W-J LW Identification BW = 365, and W-J Word Attack BW = 365, and a cubic child age term for PPVT BW = 180. Covariates for all models included school fixed effects, testing date and form, child sex, whether child is Black, whether child is Hispanic, whether child speaks only English at home, parent age, whether parent graduated high school, whether parent is the mother, whether parent works fulltime, number of adults living at home, number of children living at home, whether birth mother lives in the home, whether birth father lives in the home, and four dummy variables indicating all early care settings in which the child spent 10 or more hours per week before entering Pre-K.

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

hort 2. In interpreting these findings, it is important to underscore again that these are *average* differences across the two cohorts and that the internal validity of our RD estimates rest on a valid quasi-experiment *at the age cutoff*. We did find balance at the cutoff on children's demographic characteristics, as described earlier in the paper, suggesting there was randomization at the cutoff. As the field seeks to improve the application of Pre-K RD studies, these findings help point the way to one additional check and added understanding of this context and our estimates.

These limitations notwithstanding, the current study contributes to the growing literature on the effectiveness of early childhood education interventions that are implemented at scale, specifically Pre-K programs. Building on recommendations to improve the rigor of RD designs through enhanced methodology and extensive robustness checks, this study documents that Pre-K enhances outcomes for children from low-income, high risk urban backgrounds across a variety of academic domains. Most importantly, the current study highlights that Pre-K can confer benefits on children from communities beset by multiple risk factors, suggesting that it may serve as a buffer against some of these adversities. Future research should continue to enhance the rigor of study methodology, examine links between the implementation of the Pre-K program and children's outcomes, and conduct longitudinal studies to explore the potential long-term effects of Pre-K for similar populations of children.

Funding

This study was part of a larger evaluation of the PreKindergarten Expansion program in the state of Maryland. Major funding for this study was provided by the MD State Department of Education. Additional funding was provided by the Brady Education Fund. Although the sponsors approved the study design prior to funding, as it was included in the proposal for funding, the sponsors were not involved in any aspect of study implementation.

CRediT authorship contribution statement

Brenda Jones Harden: Conceptualization, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Bonnie E. Brett:** Supervision, Writing – original draft. **Jacquelyn T. Gross:** Formal analysis, Writing – original draft. **Christina Weiland:** Formal analysis, Methodology, Writing – review & editing. **Jordan Berne:** Formal analysis, Writing – review & editing. **Elisa L. Klein:** Writing – review & editing, Conceptualization. **Christy Tirrell-Corbin:** Writing – review & editing, Conceptualization.

Data availability

The authors do not have permission to share data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecresq.2023.01.011.

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association, 103*(484), 1481–1495.

Ansari, A., & Winsler, A. (2016). Kindergarten readiness for low-income and ethnically diverse children attending publicly funded preschool programs in Miami. *Early Childhood Research Quarterly, 37*, 69–80. https://doi.org/10.1016/j.ecresq.2016.06.002.

Bai, Y., Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2020). Long-term effects of early childhood programs through eighth grade: Do the effects fade out or grow? *Children and Youth Services Review, 112*(104890), 1–10.

Baltimore City Schools. (2019a). *Investing in our future: A world-class education system for Baltimore city students* Retrieved fromhttps://www.baltimorecityschools.org/sites/default/files/2019-01/investinginourfuture.pdf.

Baltimore City Schools. (2019b). *District profile: 2019*. Baltimore City Public Schools Retrieved from https://www.baltimorecityschools.org/sites/default/files/inline-files/999_schoolprofile_0.pdf.

Barnard, J., & Rubin, D. B. (1999). Small-Sample Degrees of Freedom with Multiple Imputation. *Biometrika, 86*(4), 948–955.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25–50. https://doi.org/10.2307/1602366.

Barnett, W. S., & Jung, K. (2021). Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade. *Early Childhood Research Quarterly, 56*, 248–259.

Barnett, W. S., Jung, K., Friedman-Krauss, A., Frede, E. C., Nores, M., Hustedt, J. T., … Daniel-Echols, M. (2018). State Pre-K effects on early learning at Kindergarten entry: An analysis of eight state programs. *AERA Open, 4*(2), Article 2332858418766291. https://doi.org/10.1177/2332858418766291.

Bartik, T. (2013). Integrating enriched learning and protection from toxic stress in early education settings. In *Early childhood programs as an economic development tool: Investing early to prepare the future workforce. In Preparing Wisconsin's Youth for Success in the Workforce [Internet]* (pp. 27–41). Policy Institute for Family Impact Seminars: Madison, WI.

Bartik, T. J., & Hershbein, B. (2018). Pre-K Effectiveness at a Large Scale. *Employment Research Newsletter, 25*(2), 2–6.

Bassok, D. (2010). Do Black and Hispanic children benefit more from preschool? Understanding differences in preschool effects across racial groups. *Child Development, 81*(6), 1828–1845.

Besharov, D. J., Call, D. M., & Scott, J. M. (2020). PROTOCOL: Early childhood education programs for improving the development and achievement of low-income children: a systematic review. *Campbell Systematic Reviews, 16*(3).

Burdick-Will, J. (2016). Neighborhood violent crime and academic growth in Chicago: Lasting effects of early exposure. *Social Forces, 95*(1), 133–158.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal, 17*(2), 372–404.

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

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*(2), 684–698. https://doi.org/10.2307/1131410.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied* Developmental Science*, 6*(1), 42–57. https://doi.org/10.1207/S1532480XADS0601_05.

Caughy, M. O. B., Brinkley, D. Y., Pacheco, D., Rojas, R., Miao, A., Contreras, M. M., et al., (2022). Self-regulation development among young Spanish-English dual language learners. *Early Childhood Research Quarterly, 60*, 226–236.

City Schools at a Glance. (2022). Retrieved October 18, 2022, from https://www.baltimorecityschools.org/district-overview

Ursache, A., Noble, K. G., & Blair, C. (2015). Socioeconomic status, subjective social status, and perceived stress: Associations with stress physiology and executive functioning. *Behavioral Medicine, 41*(3), 145–154.

United States Department of Justice. (2021). 2017 Crime in the United States July 23. Washington, DC: US Department of Justice. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/home.

Blair, C. (2002). School Readiness: Integrating Cognition and Emotion in a Neurobiological Conceptualization of Children's Functioning at School Entry. *American Psychologist, 57*(2), 111–127. https://doi.org/10.1037/0003-066X.57.2.111.

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. *Available at SSRN 2594430*. New York: Manpower Development Research Corporation (MDRC).

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management, 27*(4), 724–750. https://doi.org/10.1002/pam.20375.

Council of Economic Advisers. (2015). *The economics of early childhood investments [Executive summary]*. U.S. Council of Economic Advisers Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/docs/early_childhood_report_update_final_non-embargo.pdf.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives, 15*(2), 213–238. https://doi.org/10.1257/jep.15.2.213.

Currie, J., & Thomas, D. (1995). Does head start make a difference? *American Economic Review, 85*(3), 341–364.

Daniel, J. (2018). Think nationally, act locally: Early childhood education's role in closing the achievement gap. *YC Young Children, 73*(2), 10–13.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science (New York, N.Y.), 318*(5855), 1387–1388.

D'Onise, K., McDermott, R. A., & Lynch, J. W. (2010). Does attendance at preschool affect adult health? A systematic review. *Public Health, 124*(9), 500–511. https://doi.org/10.1016/j.puhe.2010.05.004.

Duncan, G. J., & Magnuson, K. A. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*(2), 109–132. https://doi.org/10.1257/jep.27.2.109.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test: PPVT-4; manual*. Pearson Clinical Assessment.

Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology, 58(3), 470-484*.

Ellis, W. R., & Dietz, W. H. (2017). A new framework for addressing adverse childhood and community experiences: The building community resilience model. *Academic Pediatrics, 17*(7), S86–S93.

Friedman-Krauss, A., Barnett, W. S. ., Garver, K., Hodges, K., Weisenfeld, G., & Gardiner, B. A. . (2021). *The state of preschool 2020*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University Graduate School of Education.

Friedman-Krauss, A., Barnett, W. S. ., Garver, K., Hodges, K., Weisenfeld, G., Gardiner, B. A., et al., (2022). *The state of preschool 2021*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University Graduate School of Education.

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2015). Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. *Journal of Educational Psychology, 107*(1), 207.

Garcia, E. (2015). *Inequalities at the starting gate: Cognitive and noncognitive skills gap between 2010 and 2011 kindergarten classmates*. Economic Policy Institute.

Gaylor, E., Golan, S., Chow, K., Grindal, T., Mercier, B., Williamson, C., et al., (2019). *Virginia preschool initiative plus (VPI+): Final evaluation report. prepared for the virginia department of education*. Menlo Park, CA: SRI International.

Gormley, W. T., Jr, & Phillips, D. (2005). The effects of universal pre-k in Oklahoma: Research highlights and policy implications. *Policy Studies Journal, 33*(1), 65–82.

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science (New York, N.Y.), 320*(5884), 1723–1724. https://doi.org/10.1126/science.1156019.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal Pre-K on cognitive development. *Developmental Psychology, 41*(6), 872–884. https://doi.org/10.1037/0012-1649.41.6.872.

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2021). The long-term effects of universal preschool in boston *(No. w28756)*. National Bureau of Economic Research.

Grehan, A., Cavalluzzo, L., Gnuschke, J., Hanson, R., Oliver, S., & Vosters, K. (2011). Participation during the first four years of tennessee's voluntary prekindergarten program. issues & answers. *Rel 2011-No. 107*. Regional Educational Laboratory Appalachia.

Grindal, T., Gaylor, E., Chow, K., & Golan, S. (2019). The Implementation and Impacts of a Large-Scale Public Preschool Program: Evidence from Virginia Preschool Initiative Plus*the Society for Research in Child Development Biennial Meeting, Baltimore, MD* MarchPaper presented at Paper presented at.

Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics, 29*(3), 446–493. https://doi.org/10.1111/j.1467-9353.2007.00359.x.

Heckman, J. J. (2012). Invest in early childhood development: Reduce deficits, strengthen the economy. *The Heckman Equation, 7*, 1–2.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Hill, C. J., Gormley, W. T., Jr, & Adelstein, S. (2015). Do the short-term effects of a high--quality preschool program persist? *Early Childhood Research Quarterly, 32*, 60–79.

Hofer, K. G., Checkoway, A., Goodson, B., & Nichols, A. (2018). *Massachusetts preschool expansion grant (PEG) impact evaluation report (Research report)*. Cambridge, MA: Abt Associates, Inc.

Hustedt, J., Jung, K., Barnett, W. S., & Williams, T. (2015). Kindergarten readiness impacts of the Arkansas better chance state Pre-K initiative. *Elementary School Journal, 116*(2), 198–216. https://doi.org/10.1086/684105.

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

Iruka, I. U., Curenton, S. M., & Durden, T. R. (Eds.). (2017). *African american children in early childhood education: Making the case for policy investments in families, schools, and communities*. Emerald Group Publishing.

Johnson, A. D., Markowitz, A. J., Hill, C. J., & Phillips, D. A. (2016). Variation in impacts of Tulsa pre-K on cognitive development in Kindergarten: The role of instructional support. *Developmental Psychology, 52*(12), 2145–2158.

Jones Harden, B. (2021). *Racial discrimination and parenting: Implications for intervening with African American infants and families. Zero to three scientific meeting* April 27. ZERO TO THREE.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300.

Kohen, D. E., Leventhal, T., Dahinten, V. S., & McIntosh, C. N. (2008). Neighborhood disadvantage: Pathways of effects for young children. *Child Development, 79*(1), 156–169.

La Paro, M. K., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*(4), 443–484. https://doi.org/10.3102/00346543070004443.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281.

Lee, K., Jones Harden, B., Jaramillo, M., Norman, T., & Jimenez Parra, L. (2023). Conventional and expanded adverse childhood experiences (ACEs) and maternal functioning among low-income African American mothers. *Child and Family Social Work, 28(1), 37-52*.

Lipsey, M. W., Farran, D. C., Bilbrey, C., Hofer, K. G., & Dong, N. (2011). *Initial results of the tennessee voluntary pre-k program* April. Peabody Research Institute, Vanderbilt University Retrieved from https://my.vanderbilt.edu/tnPre-Kevaluation/files/2013/10/April2011_PRI_Initial_TN-VPK_ProjectResults.pdf.

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Pre-K Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45*, 155–176.

Lipsey, M. W., Weiland, C., Yoshikawa, H., Wilson, S. J., & Hofer, K. G. (2015). The Pre-K age-cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis, 37*(3), 296–313. https://doi.org/10.3102/0162373714547266.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., et al., (2007). Observed classroom quality profiles in state-funded Pre-K programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3–17.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review, 26*(1), 52–66. https://doi.org/10.1016/j.econedurev.2005.11.005.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics, 122*(1), 159–208. https://doi.org/10.1162/qjec.122.1.159.

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly, 36*, 521–536.

Magnuson, K., & Shager, H. (2010). Early education: Progress and promise for children from low income families. *Children and Youth Services Review, 32*(9), 1186–1198. https://doi.org/10.1016/j.childyouth.2010.03.006.

Magnuson, K., & Waldfogel, J. (2016). Trends in income-related gaps in enrollment in early childhood education: 1968 to 2013. *AERA Open, 2*(2), 1–13. https://doi.org/10.1177/2332858416648933.

Manship, K., Quick, H., Holod, A., Mills, N., Ogut, B., Chernoff, J. J., et al., (2015). *Impact of California's transitional kindergarten program, 2013–14*. American Institutes for Research Retrieved from https://files.eric.ed.gov/fulltext/ED563818.pdf.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., et al., (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732–749.

McCarton, C. M., Brooks-Gunn, J., Wallace, I. F., Bauer, C. R., Bennett, F. C., Bernbaum, J. C., et al., (1997). Results at age 8 years of early intervention for low-birth-weight premature infants: The Infant Health and Development Program. *Journal of the American Medical Association, 277*(2), 126–132.

McClelland, M. M., & Tominey, S. L. (2014). The development of self-regulation and executive function in young children. *Zero to Three, 35*(2), 2–8.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., … Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher, 46*(8), 474–487.

Meek, S., Iruka, I. U., Allen, R., Yazzie, D., Fernandez, V., Catherine, E., et al., (2020). *Fourteen priorities to dismantle systemic racism in early care and education*. The Children's Equity Project etrieved fromR https://childandfamilysuccess.asu.edu/cep.

Mendez Smith, J., Crosby, D., & Stephens, C. (2021). Equitable access to high-quality early care and education: Opportunities to better serve young Hispanic children and their families. *The ANNALS of the American Academy of Political and Social Science, 696*(1), 80–105.

Montrosse-Moorhead, B., Dougherty, S. M., LaSalle, 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.

Moore, J. E., Cooper, B. R., Domitrovich, C. E., Morgan, N. R., Cleveland, M. J., Shah, H., et al., (2015). The effects of exposure to an enhanced preschool program on the social-emotional functioning of at-risk children. *Early Childhood Research Quarterly, 32*, 127–138.

Ng, F. F.-Y., Tamis-LeMonda, C., Yoshikawa, H., & Sze, I. N.-L. (2014). Inhibitory control in preschool predicts early math skills in first grade. *International Journal of Behavioral Development, 39*(2), 139–149. https://doi.org/10.1177/0165025414538558.

Office of Early Learning. (2019). *Starting strong: Increasing preschool quality and access. The preschool development grants program summary report* December. U.S. Department of Education Retrieved from https://www2.ed.gov/programs/preschooldevelopmentgrants/pdg-summary-rpt-final-12202019.pdf.

Parker, E., Diffey, L., & Atchinson, B. (2018). *How states fund pre-k: A primer for policymakers*. Education Commission of the States Retrieved from https://www.ecs.org/wp-content/uploads/How-States-Fund-Pre-K_A-Primer-for-Policymakers.pdf.

Perry, R. E., Braren, S. H., & Blair, C. (2018). Socioeconomic risk and school readiness: Longitudinal mediation through children's social competence and executive function. *& Family Life Project Key Investigators Frontiers in psychology, 9*, 1544.

Peisner-Feinberg, E. S., & Schaaf, J. M. (2011). *Effects of the North Carolina more at four pre-k program on children's school readiness skills*. Chapel Hill: University of North Carolina, FPG Child Development Institute.

Peisner-Feinberg, E., Buysse, V., Fuligni, A., Burchinal, M., Espinosa, L., Halle, T., & Castro, D. C. (2014). Using early care and education quality measures with dual language learners: A review of the research. *Early Childhood Research Quarterly, 29*(4), 786–803.

Phillips, D. A., Johnson, A. D., Weiland, C., & Hutchison, J. E. (2017). Public preschool in a more diverse America: Implications for next-generation evaluation research. *Working paper series, no. 2-17*. Poverty Solutions, University of Michigan.

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., … Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-k effects, a consensus statement*. Brookings Institution Retrieved from https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf.

Phillips, D. (2016). Integrating enriched learning and protection from toxic stress in early education settings. In S. Jones, & N. Leseau (Eds.), *Leading edge in early childhood education* (pp. 7–28). Cambridge, MA: Harvard University Press.

Pianta, R. C., Cox, M. J., & Snow, K. L. (2007). *School readiness and the transition to kindergarten in the era of accountability*. Paul H Brookes Publishing. https://doi.org/10.1080/15240750701831966.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system™: Manual K-3*. Paul H Brookes Publishing.

Pion, G. M., & Lipsey, M. W. (2021). Impact of the Tennessee Voluntary Prekindergarten Program on Children's Literacy, Language, and Mathematics Skills: Results From a Regression-Discontinuity Design. *AERA Open, 7*, Article 23328584211041353.

Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*, 141–158.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to Kindergarten outcomes. *Developmental Psychology, 45*(3), 605–619.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., et al., (2010). *Technical Report*. Administration for Children & Families.

Raver, C. C., Blair, C., & Willoughby, M. (2013). Poverty as a predictor of 4-year-olds' executive function: new perspectives on models of differential susceptibility. *Developmental psychology, 49*(2), 292.

Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership, 70*(8), 10–16.

Reynolds, A. J., & Temple, J. A. (1995). Quasi-experimental estimates of the effects of a preschool intervention. *Evaluation Review, 19*(4), 347–373. https://doi.org/10.1177/0193841x9501900401.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., et al., (2010). Standards for regression discontinuity designs. *What Works Clearinghouse, 1-8. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.*.

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Riverside.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The highscope perry preschool study through age 40*. Ypsilanti: HighScope Press.

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

Shah, H. K., Domitrovich, C. E., Morgan, N. R., Moore, J. E., Cooper, B. R., Jacobson, L., et al., (2017). One or two years of participation: Is dosage of an enhanced publicly funded preschool program associated with the academic and executive function skills of low-income children in early elementary school?. *Early Childhood Research Quarterly, 40*, 123–137.

Trochim, W. M. (1984). Research design for program evaluation. *The regression-discontinuity approach*: 6. Thousand Oaks, CA: Sage.

United States Department of Justice. (2019). *Maryland U.S. attorney announces support to baltimore city as part of the national public safety partnership to combat violent crime* June 3 Washington, DC: US Department of Justice. Retrieved from https://www.justice.gov/usao-md/pr/maryland-us-attorney-announces-support-baltimore-city-part-national-public-safety.

Weikart, D. P., Bond, J. T., & McNeil, J. T. (1978). *The Ypsilanti Perry preschool project: Preschool years and longitudinal results through fourth grade*. Yipsilanti, MI: High/Scope Educational Research Foundation.

Weiland, C., McCoy, D. C., Grace, E., & Park, S. O. (2017). Natural window of opportunity? Low-income parents' responses to their children's impending Kindergarten entry. *AERA Open, 3*(1), Article 2332858416681509.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly, 28*(2), 199–209.

Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2020). The effects of enrolling in oversubscribed Pre-K programs through third grade. *Child Development, 91*(5), 1401–1422.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a Pre-K program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112–2130. https://doi.org/10.1111/cdev.12099.

White House. (2013). *Fact sheet: President Obama's plan for early education for all Americans* February 13. Office of the Press Secretary Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2013/02/13/fact-sheet-president-obama-s-plan-early-education-all-americans.

White House. (2015). *Fact sheet: Helping all working families with young children afford child care* January 21 Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2015/01/21/fact-sheet-helping-all-working-families-young-children-afford-child-care.

Williams, S. C., Barajas, C. B., Milam, A. J., Olson, L., Leaf, P., & Connolly, F. (2019). Preparing students for success: Differential outcomes by preschool experience in Baltimore City, Maryland. *Prevention Science, 21*(4), 467–476. https://doi.org/10.1007/s11121-019-01069-3.

Wong, V., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state Pre-K programs. *Journal of Policy Analysis and Management, 27*(1), 122–154. https://doi.org/10.1002/pam.20310.

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. Society for Research in Child Development Policy Brief, 1-21. Retrieved from http://www.srcd.org/policy-media/policy-updates/meetings-briefings/investing-our-futureevidence-base-preschool.

Yu, D., Caughy, M. O. B., Smith, E. P., Oshri, A., & Owen, M. T. (2020). Severe poverty and growth in behavioral self-regulation: The mediating role of parenting. *Journal of Applied Developmental Psychology, 68*, 1–11.

Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2014). Head Start's impact is contingent on alternative type of care in comparison group. *Developmental Psychology, 50*(12), 2572–2586. https://doi.org/10.1037/a0038205.