MASSACHUSETTS PRESCHOOL EXPANSION GRANT (PEG) Impact Evaluation Report

July 26, 2018

Prepared for:
Massachusetts Department of Early Education and Care
51 Sleeper Street, 4th Floor
Boston, MA 02210

Submitted by:
Kerry Hofer, Amy Checkoway, Barbara Goodson, and Austin Nichols
Abt Associates
10 Fawcett Street
Cambridge, MA 02138

MASSACHUSETTS PRESCHOOL EXPANSION GRANT (PEG)
Impact Evaluation Report

Recommended Citation for This Report:

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.

Contact Us:

Abt Associates, Inc.

10 Fawcett Street

Cambridge, MA 02138-1168

(617) 492-7100

Acknowledgements

Abt would like to acknowledge the thoughtful contributions to this work by Dr. Jocelyn Bowne, Director of Research and Preschool Expansion Grant Administration at the Massachusetts Department of Early Education and Care. Abt is also greatly appreciative of the PEG district and program staff who supported and helped to facilitate this research effort. The child assessment data collection effort was coordinated by Abt staff Michele de Mars, Alanah Hall, Jackie Mendez, Renee Lamoreau, Rachel Luck, and Katie Murphy, with contributions from Michelle Blocklin.

Executive Summary

In late 2014, the Massachusetts Department of Early Education and Care (EEC) was awarded a four-year federal Preschool Development Grant to support the expansion of high-quality early childhood education to high-needs communities, with particular focus on serving children from low- and middle-income families.

The Massachusetts Preschool Expansion Grant (PEG) model is built around a collaborative public-private delivery system. PEG requires shared governance between local school districts and EEC-licensed programs, with classrooms run by the community-based programs. The 48 PEG classrooms provide free prekindergarten for low income four-year-olds (i.e., age four as of September 1 of the incoming school year) who will be eligible for kindergarten in the upcoming fall and who, with some exceptions, have not yet attended a formal child care program (licensed center-based or family child care).

The PEG model is intended to achieve a high level of quality in instructional and emotional supportiveness, classroom organization, and learning resources, while also being responsive to local needs. Each PEG community was encouraged to design a program that adhered to certain quality requirements, with a goal of ensuring consistently high quality learning environments while also allowing for local variation (see Exhibit E.1).

Exhibit E.1: PEG Model Quality Elements

1	A collaborative local governance structure designed to oversee implementation and work on systems coordination for all children in the community;
2	Full-day, full-year programming (at least 8 hours/day, 12 months/year);
3	A maximum class size of 20;
4	A maximum child-teacher ratio of 10:1;
5	A curriculum/a aligned with the MA Preschool Standards and Guidelines (curriculum/a may vary by grantee);
6	The use of Teaching Strategies Gold® as a formative assessment tool;
7	One educator in each classroom with a bachelor's degree in a relevant field;
8	Salaries for lead educators commensurate with comparable positions in public schools within the respective community;
9	Joint professional development training and coaching for teaching staff, and other supports for planning and implementation of curriculum, in collaboration with the LEA;
10	Family engagement activities, including support for kindergarten transition and resources about child development;
11	Comprehensive services including services addressing health, mental health, and behavioral needs for all families;
12	Inclusion of students receiving special education support; and
13	Efforts to build linkages with services for children from birth to age 3 as well as connections with elementary schools.

Source: Massachusetts Department of Early Education and Care

By the end of the grant period (2018–19), PEG centers are also expected to attain the highest rating (Level 4) in the Massachusetts Quality Rating and Improvement System (QRIS) or Level 3 and National Association for the Education of Young Children (NAEYC) accreditation.

To study the impacts of PEG on children's school readiness, a rigorous impact evaluation, using an age cutoff regression discontinuity design (RDD), was conducted to examine whether children who had attended a PEG program had greater skills at kindergarten entry compared with similar children who did not attend PEG. This type of study design involves comparing the skills of children who are very close to one another in age and development and differ only in their exposure to the PEG program.

The impact evaluation answers the following research questions:

- What is the impact of the PEG program on children's early academic skills (literacy and math)?
- What is the impact of the PEG program on children's language development (vocabulary)?
- What is the impact of the PEG program on children's executive function skills?

The study compared the early academic and executive function skills for students who attended PEG classrooms in the 2016-17 school year versus the skills of students who had missed PEG's age cutoff and had not spent the year in PEG (and were just entering PEG classrooms in the 2017-18 school year). A total of 1,107 children were included in the analysis sample: 582 in the treatment group (PEG enrollees in the 2016-17 school year) and 525 in the control group (children who subsequently enrolled in PEG in the 2017-18 school year). Both groups were similar in terms of gender and home language.

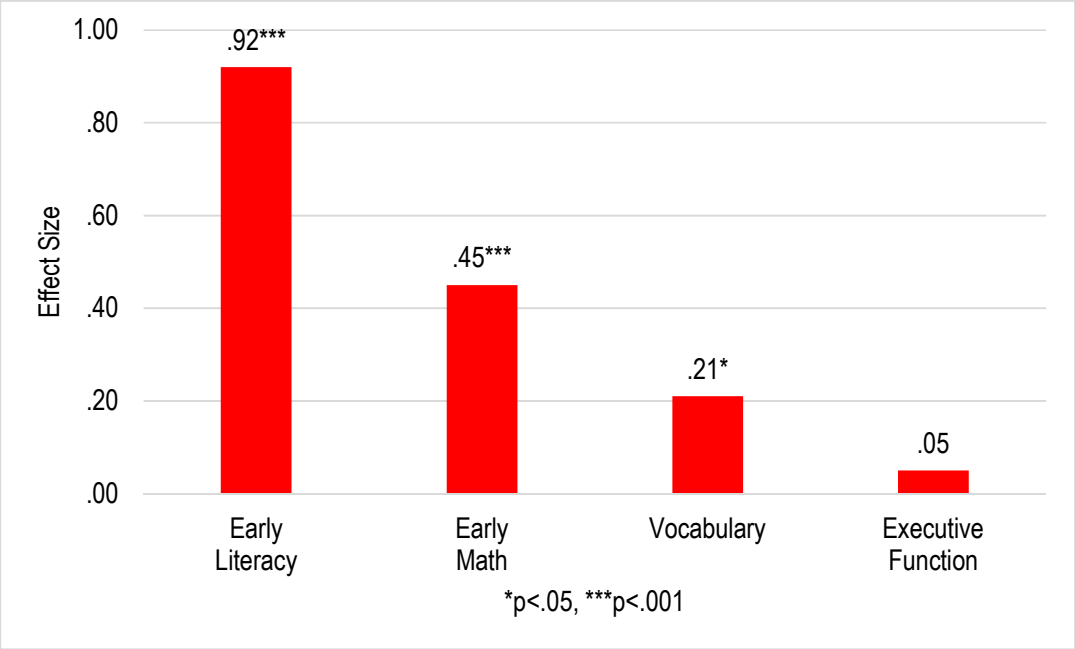
Children were assessed individually by trained assessors, typically in a single assessment visit lasting no more than 45 minutes. All assessments included were administered to children in English, regardless of the students' home language or English proficiency. The study used standardized measures to assess children's early literacy and early math skills, and early vocabulary, and a nonstandard but widely used measure assessed children's executive function skills. Assessors used the following battery of measures:

- *Early Literacy.* Children's early literacy skills were measured with the Woodcock-Johnson III Tests of Cognitive Abilities: Letter-Word Identification Subtest (Woodcock, McGrew, & Mather, 2001; WJ-III).
- *Early Math.* Children's early mathematics skills were measured using the Woodcock-Johnson III Tests of Cognitive Abilities: Applied Problems Subtest.
- *Vocabulary.* Children's receptive vocabulary knowledge was measured with the Peabody Picture Vocabulary Test, Fourth Edition (Dunn & Dunn, 2007).
- *Executive Functioning.* Children's executive functioning was measured with the Hearts & Flowers Task (previously called the Dots Task; Davidson et al., 2006; Diamond et al., 2007), which measures children's ability to remember rules and to inhibit their response when applying those rules under different contexts.

To estimate the effect of PEG, the study ran regression models that predicted children's scores from PEG participation controlling for child age relative to the birthdate cutoff, the interaction of treatment and child age relative to the cutoff (both critical in age-cutoff RDD models), child gender, home language, and prior child care exposure and that accounted for the clustering of children in PEG classrooms.

The study found impacts on children's early literacy and early math achievement (effect sizes of .92 and .45 standard deviation units, respectively) and on their vocabulary development (effect size of .21 standard deviation units). The effect sizes (impact estimate) and statistical significance of the effects are presented in Exhibit E.2 below, arranged in descending order of impact size.

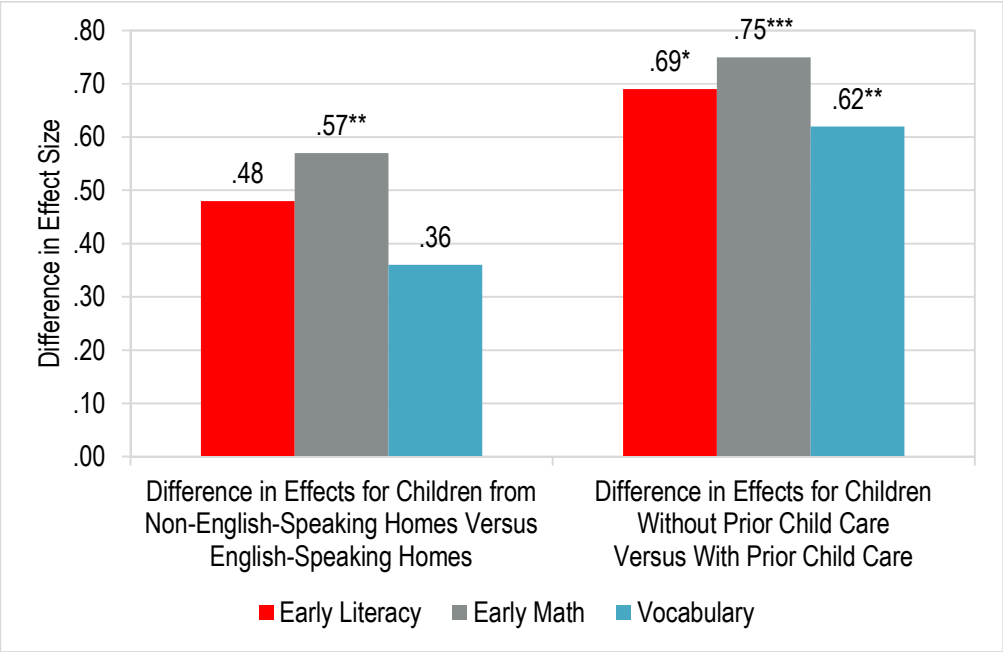
Exhibit E.2: Impact of the PEG Program on Children's Skills



On all three measures of early academic performance, PEG had a positive and statistically significant impact on children's achievement. The largest impact was seen for early literacy skills, and the smallest effect was for vocabulary. On executive function, the children who attended PEG scored higher than the children who had not yet attended PEG, but the impact was not significant.

Exploratory analyses indicated that the impact of PEG was stronger for children in homes where English was not the primary language and for children who had not had prior child care exposure. PEG did not appear to be more or less effective for children of either gender. Exhibit E.3 shows the difference in PEG impact on each academic outcome for different subgroups of children compared to one another.

Exhibit E.3: Difference in PEG Impact by Child Demographic Subgroup



*p<.05, **p<.01, ***p<.001

In sum, PEG had positive impacts on children's early academic skills, with the strongest impacts on the most vulnerable children. This research provides the field with important information about the feasibility of implementing high-quality preschool through collaborations between public schools and private early education programs and provides additional evidence about the benefits of high-quality prekindergarten for children from disadvantaged backgrounds. This study also provides evidence of the impact of a model implemented in community-based preschool programs, which is not often addressed in the existing research on early education effectiveness.

As is true for most other preschool models, the Massachusetts PEG program delivered a combination of programmatic features that alone or together might drive impacts on children, including but not limited to standardized curricula aligned with learning standards, teacher coaching and professional development, and improved teacher compensation. Further exploratory research is underway to try to better understand the relationship of the implementation of particular program features to children's outcomes to try to disentangle which levers may be associated with the observed impact.

1. Introduction

Taken together, the past 40 years of research on the impacts of early education on children's development makes a strong case for its benefits, particularly for children from low-income homes (Leak et al., 2010; Larsen & Robinson, 1989). For example, a recent meta-analysis of evaluations of 84 diverse early childhood programs that were conducted between 1965 and 2007 reported a substantial positive average program effect (Duncan & Magnuson, 2013). The meta-analysis included evaluations of small demonstration programs, such as Perry Preschool, and evaluations of large preschool programs such as Head Start. Combining across outcome domains, including outcomes in cognition (e.g., IQ), language (e.g., expressive and receptive vocabulary) and achievement (e.g., early reading and mathematics skills), the average program impact was estimated to be about .35 standard deviations, although when the precision of the evaluations was taken into account, the average effect size dropped to .21 standard deviation units. Most of the studies included in this meta-analysis focused on programs that served low-income children. However, more recent research focusing on universal preschool programs without income eligibility requirements has shown that middle-class children also can benefit substantially from early education. Two recent evaluations of at-scale urban prekindergarten programs, in Tulsa and Boston, found large effects (between one-half and a full year of additional learning) on language, literacy and math (Gormley, Phillips, & Gayer, 2008; Weiland & Yoshikawa, 2013).

The effects of early childhood programs on children's socio-emotional development have been measured less frequently than early academic outcomes. Across evaluations that have examined this domain, the findings are inconsistent (Gormley, Phillips, Newmark, Welti, & Adelstein, 2011; Raver et al., 2009; Riggs, Greenberg, Kusche, & Pentz, 2006). Perry Preschool was found to reduce children's externalizing behavior problems (such as acting out or aggression) in elementary school (Heckman, Pinto, & Savelyev, 2012). However, more recently, the National Head Start Impact Study found no effects in the socio-emotional domain for four-year-old children, although problem behavior, specifically hyperactivity, was reduced after one year (Puma, 2010). An evaluation of the Tulsa prekindergarten program found the children less timid and more attentive, suggesting greater engagement in the classroom, compared to children who had not attended prekindergarten or Head Start (Gormley, Phillips, & Gayer, 2008). However, there were no differences among children in their aggressive or hyperactive behavior. In contrast, the Boston evaluation found that the public school program increased children's skills on most measures of executive functioning and one measure of emotional control; the effects were much smaller than the impacts on early academic outcomes (Weiland & Yoshikawa, 2013). A recent meta-analysis of early childhood programs indicates that significant reductions in children's externalizing behavior problems were related to the intensity of the program focus on social and emotional development (Schindler et al., 2015). Programs without a clear focus on socio-emotional development showed no significant effects. Among the programs that did focus on this domain, the size of the effects was related to the intensity with which the program targeted socio-emotional development; the largest effects were from child social skills training programs.

The literature also suggests that the quality of early education programs likely relates to the size of their impact. A secondary data analysis of eight studies of preschool children in center-based programs examined the extent to which program quality predicted gains in children's language, literacy, mathematics, and social skills. It found that increases in the quality of instruction were related to gains in children's language and literacy outcomes, but only in higher-quality classrooms (Zaslow et al., 2016). Domain-specific and interaction-specific measures of quality were more strongly related to children's outcomes than were more global measures.

Though structural features of quality (such as group size, ratio, and teacher qualifications) help to create the conditions for positive "process quality," they do not ensure it (Burchinal et al., 2008; Burchinal,

Vandergrift, Pianta, & Mashburn, 2010; Early et al., 2007). Process quality features—children's immediate experience of positive and stimulating interactions—appear to be the most important contributors to children's gains in language, literacy, mathematics and social skills. Research suggests that two aspects of process quality that appear to be most important to children's gains during the preschool years are: (1) interactions explicitly aimed at supporting learning, that foster both higher-order thinking skills in general and learning of content in specific areas such as early math and language, are related to gains; and (2) warm, responsive teacher-child relationships and interactions that are characterized by back and forth conversations—"serve and return"—to discuss and elaborate on a given topic (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000).

There is increasing evidence of the benefits of evidence-based curricula targeting specific teacher behaviors and student-teacher interactions. Whereas evaluations of more global curricula show little or no gains associated with their use (Bierman et al., 2008; Clements & Sarama, 2007; Preschool Curriculum Evaluation Research Consortium, 2008); recent experimental evaluations of math, language, and literacy curricula resulted in moderate and large gains in the targeted domains of children's development (Clements & Sarama, 2008a, Clements & Sarama, 2008b; Fantuzzo, Gadsden, & McDermott, 2011; Gormely, Gayer, Phillips, & Dawson, 2005, Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011; Wasik, Bond, & Hindman, 2006).

1.1 Federal Preschool Development Grant Program: Expanding Access to High Quality Preschool

Recognizing the strong and consistent evidence that participation in high quality early learning programs can lead to both short- and long-term positive outcomes for disadvantaged children,[2] the U.S. Departments of Education (ED) and Health and Human Services (HHS) jointly sponsored the Preschool Development Grant program to support state and local efforts to develop and/or expand high-quality prekindergarten programs to increase access for children from low- and moderate-income families so that they can enter kindergarten ready to succeed. Eighteen states, including Massachusetts, have received grants totaling more than $226 million.

States receiving grants are expected to (a) provide voluntary, high-quality prekindergarten programs for eligible children through subgrants to two or more high-need communities; (b) increase the number of children in high-quality prekindergarten programs by creating new slots for underserved and high-needs children in high-quality programs or by increasing slots in existing state prekindergarten programs; and (c) deliver these prekindergarten programs through a mixed-delivery system of providers that includes schools, licensed child care centers, Head Start programs, and community-based organizations.

Aligned with the research on the features of high-quality programs, the Preschool Development Grant program also specifies that programs should have high staff qualifications, low child-staff ratios and small class sizes, a full-day program, and comprehensive services for children. Additionally, programs should have in place early learning and development standards; a comprehensive early learning assessment system, including screening measures, formative assessments, measures of environmental quality, and a kindergarten screening assessment; comprehensive services, including health screenings, family engagement activities, and nutrition services; and services coordinated with school districts and other organizations providing services for children with special needs.

1.2 Massachusetts PEG Program

In late 2014, the Massachusetts Department of Early Education and Care (EEC) was awarded a federal Preschool Development Grant focused on expansion (referred to in this report as the Massachusetts Preschool Expansion Grant or PEG) in the amount of $60 million over four years to expand high-quality early education to four-year-old children whose families earned under 200 percent of the federal poverty

level. The PEG program provided the Commonwealth with a unique opportunity to increase access to high-quality prekindergarten through collaborative partnerships between local school districts and community-based agencies. It also allowed EEC to pilot a model that, if successful, could be replicated.

The grant has supported PEG classrooms in five underserved communities across Massachusetts. In each community, local education agencies (LEAs) are granted the funds and subcontract with EEC-licensed providers (ELPs) for the direct services to preschool children and families. Participating LEAs and ELPs are following a model (described in Chapter 2) that is intended to deliver the ingredients and supports that research has shown can lead to improved child outcomes.

As part of the PEG program, EEC invested in a rigorous multi-year evaluation. The PEG evaluation is being conducted by an independent research firm, Abt Associates Inc. The evaluation has four main components:

- Implementation study of the PEG quality components in PEG communities and programs[1];
- Longitudinal study of outcomes for PEG children and families;
- Impact study of effects on PEG children and families; and a
- Cost study.

This report describes the results of the evaluation's impact study which compares the effects of PEG on the cohort of children who entered PEG in the fall of 2016 (Year 2) versus those who entered PEG in the fall of 2017 (Year 3). All children were assessed at the same point in time, during the fall of 2017 (the beginning of the kindergarten year for the Year 2 PEG cohort and the beginning of the PEG preschool year for the Year 3 PEG cohort). The evaluation, described in-depth in this report and its Appendix, produced results that generalize to children right around the cutoff (i.e., children who are very similar to one another in terms of age and development) and compared skills for children who had PEG versus children who had not yet attended PEG but who were expected to be similar in all ways but age.

This report is organized into the following chapters:

- Overview of the Massachusetts PEG program (Chapter 2);
- Overview of the impact evaluation design (Chapter 3);
- Results including the main effects on children's development and learning and effects for subgroups of children (Chapter 4); and
- Discussion of the implications of the findings (Chapter 5).

The Appendix provides detailed information about the analyses and findings from multiple analytic models.

[1] The *Year 1 Massachusetts PEG Evaluation Report*, which focuses on the implementation of PEG, can be found at: https://www.abtassociates.com/insights/publications/report/year-1-massachusetts-preschool-expansion-evaluation-report. The *Year 2 Evaluation Report*, which also focuses on implementation, can be found at: https://www.abtassociates.com/insights/publications/report/year-2-massachusetts-preschool-expansion-grant-peg-evaluation-report-0.

2. Overview of the Massachusetts PEG Program

This chapter provides an overview of the Massachusetts PEG program, expectations for participating preschool programs and rationale for the state-level program model, and characteristics of participating children.

2.1 Structure of the PEG Program

Massachusetts used its PEG grant to fund 48 classrooms in five high-need communities—Boston, Holyoke, Lawrence, Lowell, and Springfield—to expand access to free full-day, full-year prekindergarten for four-year-old children through public-private partnerships between the local school district (referred to as LEAs, for local education agency) and EEC-licensed early learning providers (ELPs).

To determine local PEG fund allocations, the state used the Chapter 70 foundation per child allocation for preschool as a baseline and then adjusted upwards to account for the PEG program's extended hours per day and increased services. The design of the funding mechanism ensured a minimum investment in the smallest community (Holyoke) and a corresponding ceiling—adjusted for the high cost of living—for the largest community (Boston). Exhibit 2.1 shows the amount awarded per community, along with the number of ELPs, centers, classrooms, and preschool slots per year.

Exhibit 2.1: Number of PEG Participating Organizations and Classrooms by Community, 2016-17

Public School District	Grant Award	# of ELPs	# of PEG Centers	# of PEG Classrooms	# Preschool Slots/Year
Boston Public Schools	$4,061,250	8	12	15	280
Holyoke Public Schools	$1,425,000	2	4	4	76
Lawrence Public Schools	$2,351,250	2	2	10	130
Lowell Public Schools	$2,850,000	2	1[b]	8	156
Springfield Public Schools	$3,562,500	3	4[c]	11	195
Overall	**--**	**16[a]**	**24**	**48**	**837**

[a] One ELP operated PEG classrooms in two communities (Springfield and Holyoke).
[b] In Lowell, two ELPs jointly operated one center.
[c] In Springfield, three ELPs jointly operated one of the four centers.

Beginning in September 2015, ELPs began to operate PEG classrooms, although full enrollment was not required until December 2015. Most PEG classrooms were managed by a single ELP, though two communities (Springfield and Lowell) established new centers in which multiple ELPs shared space. Prior to the PEG grant, all participating ELPs had experience administering preschool classrooms and managing the licensing of facility space.

In four of the five communities (except Boston), the PEG classrooms were new classrooms. These four PEG communities targeted and primarily served children who had never been enrolled in licensed early education (including both center-based programs and licensed family child care homes) in the prior year.

In Boston, PEG funding was used to support existing preschool classrooms that implemented the PEG operating schedule (i.e., extending the programs to offer full-day, full-year care in Head Start sites) and all elements of the PEG instructional model. As a result, the majority of the PEG children in Boston classrooms had already experienced formal early education prior to their PEG experience, often in the same program.

EEC staff actively collaborated with the designated LEAs and ELPs in the planning and early implementation, especially in the local planning for professional development activities during the first year of implementation (2015–16). The designated ELPs worked together with their LEA around the

selection and implementation of curriculum, coordination and provision of comprehensive services, family engagement supports, and inclusive services for special populations, as well as joint professional development.

To be eligible for PEG, children were required to meet several criteria:

- The child must have reached his/her fourth birthday by the beginning of their preschool year and not yet have turned five years of age;
- The child must be eligible for kindergarten in the following September;
- Their family must reside within the boundaries of the public school district;
- The family income must be less than 200 percent of the federal poverty level; and

In four of the five communities (except Boston), the programs prioritized children who had not previously been enrolled in a licensed early learning setting.

2.2 PEG Program Model and Rationale

The PEG model is intended to achieve a high level of quality in instructional and emotional supportiveness, classroom organization, and learning resources, while also being responsive to local needs. Each PEG community was encouraged to design a program that adhered to certain quality requirements, with a goal of ensuring consistently high quality learning environments while also allowing for local variation (see Exhibit 2.2).

Exhibit 2.2: PEG Model Quality Elements

1	A collaborative decision-making structure designed to oversee implementation and work on systems coordination for all children in the community
2	Full-day, full-year programming (at least 8 hours/day, 12 months/year)
3	A maximum class size of 20
4	A maximum child-teacher ratio of 10:1
5	A curriculum/a aligned with the MA Preschool Standards and Guidelines (curriculum/a may vary by grantee)
6	The use of Teaching Strategies Gold® as a formative assessment tool
7	One educator in each classroom with a bachelor's degree in a relevant field
8	Salaries for lead educators commensurate with comparable positions in public schools within the respective community
9	Joint professional development training and coaching for teaching staff, and other supports for planning and implementation of curriculum, in collaboration with the LEA
10	Family engagement activities, including support for kindergarten transition and resources about child development
11	Comprehensive services including services addressing health, mental health, and behavioral needs for all families
12	Inclusion of students receiving special education support
13	Efforts to build linkages with services for children from birth to age 3 as well as connections with elementary schools

Source: Massachusetts Department of Early Education and Care

By the end of the grant period (2018–19), PEG centers are also expected to attain the highest rating (Level 4) in the Massachusetts Quality Rating and Improvement System (QRIS) or QRIS Level 3 and National Association for the Education of Young Children (NAEYC) accreditation.

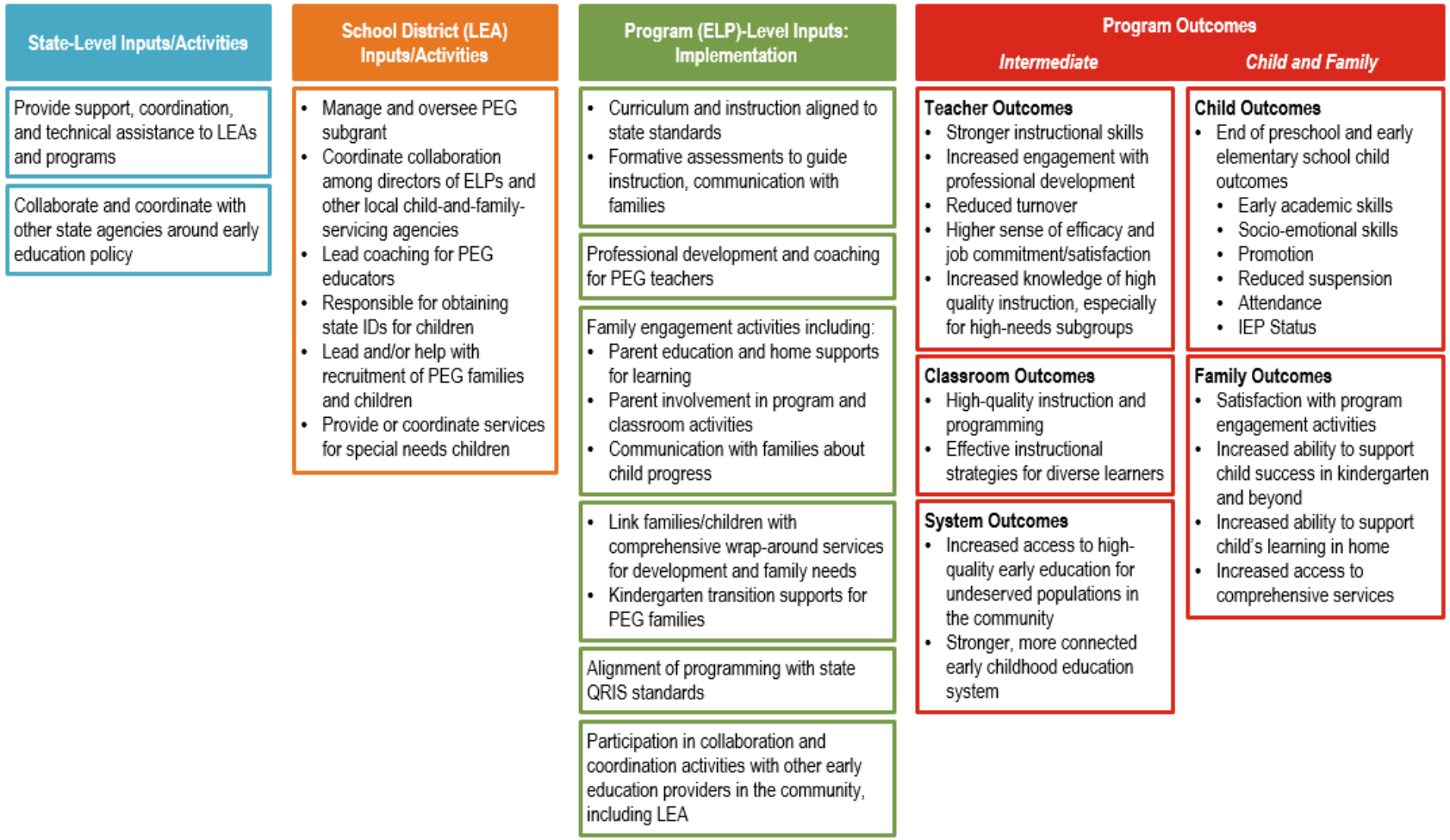
Within the PEG model framework, LEAs and ELPs had flexibility regarding the specific approaches they take to implement each quality element. As a result, PEG communities implemented each component in a variety of ways; for example, communities (and sometimes programs within communities) used different curricula and located services differently (some ELPs co-locate all PEG classrooms within one center, whereas others provide services in centers across the community).

Despite the freedom to develop different models, PEG programs showed some consistency in how they addressed three key components of the grant:

- Collaborative decision making structures (Quality Element 1):
  - Shared governance was established through regularly meeting steering committees and executive boards with representation from all partner agencies.
  - Steering committees planned the program and implemented ongoing course adjustments to ensure quality and alignment.
  - Data collected on an ongoing basis as part of the evaluation was used to support continuous quality improvement.
  - Communities developed enrollment processes that ensured both access and choice for families, often incorporating the public school kindergarten enrollment office in a referral role.
- Investment in educators (Quality Elements 8, 9):
  - Salaries recognized high levels of teacher qualification and were commensurate with public school salaries.
  - Each community planned training and coaching offerings to ensure high quality and aligned supports for educators in all PEG classrooms.
  - Coaching and job-embedded professional supports were provided. These included joint trainings across PEG classrooms and with public school educators.
  - Most communities found a three teacher per classroom structure facilitated consistent teacher participation in professional learning. In a full day program, educators do not have time outside of teaching hours to engage in professional learning; three teachers assigned to each classroom allowed more scheduling flexibility for activities outside of the classroom, such as coaching meetings, trainings and regular time for curriculum planning.
- Supports for vulnerable families (Quality Elements 10, 11, 12, 13)
  - Most programs determined they needed a dedicated family engagement staff member to coordinate the work with families, particularly case management.
  - The family engagement staff were available to provide case management and referrals to mental health and other social services.
  - Extensive outreach was necessary to identify and enroll eligible families, often requiring door-to-door outreach.
  - Most communities also offered home visits to families, generally as a relationship building tool early in the school year or case management opportunity throughout the year.
  - Programs also worked to message the importance of both enrollment in prekindergarten and regular attendance.

The requirements guiding the PEG program model were intended to ensure the delivery of high quality ingredients and supports that research has shown will improve child outcomes, especially for children at risk for academic failure. It also included goals beyond those pertaining to program quality and outcomes for educators, parents, and children. For example, the model had an explicit focus on systems building, as represented in the public-private and cross-agency collaboration that was expected to be developed among the key stakeholders in the early education system in each community.

Exhibit 2.3: Theory of Change for Massachusetts Preschool Expansion Grant



The teacher-focused supports that PEG LEAs and ELPs provided were expected to lead to greater job satisfaction and improved self-efficacy for teachers, and the ability to better recruit and retain high-quality educators. The educator supports were believed to lead to sustained improvements in classroom quality and thus child outcomes. The family engagement activities and comprehensive services were expected to lead to improved parent and child outcomes, including greater family stability, better child behavior and attendance, and less need for services in elementary school. The links between the required ingredients and both short- and long-term outcomes are shown in the PEG program theory of change (in Exhibit 2.3 below).

2.3 Children Enrolled in PEG

As per grant requirements, the children enrolled in PEG came from low-income families; in fact, the majority of families earned well below the poverty threshold. For example, 66 percent of the 2016-17 PEG families reported incomes below 100 percent of the 2016 federal poverty level for a family of four ($24,300); the average family income was $19,203 per year.

In addition to growing up in a low-income household, almost all PEG children were from racial and/or ethnic minority groups; in 2016-17, more than 90 percent were from racial minority groups and more than half of the children were Hispanic. Furthermore, almost half (44 percent) of the 2016-17 PEG children lived in households where English was not the primary language spoken (see Exhibit 2.4).

Exhibit 2.4: Demographic Characteristics of PEG Children Overall and by Community, 2016-17

	Number and Percentage of Children											
	Overall PEG		Boston		Holyoke		Lawrence		Lowell		Springfield	
Race/Ethnicity												
Non-Hispanic White	51	6%	5	2%	0	0%	0	0%	29	18%	15	9%
Hispanic	412	52%	70	27%	59	92%	130	99%	47	29%	106	63%
Black	219	28%	161	61%	0	0%	1	1%	20	12%	37	22%
Asian-American	68	9%	9	3%	0	0%	0	0%	59	36%	6	4%
Two or more races	20	3%	7	3%	0	0%	0	0%	7	4%	6	4%
Other	18	2%	9	3%	5	8%	0	0%	0	0%	0	0%
Primary Home Language												
English	444	56%	184	70%	50	78%	26	20%	49	30%	135	80%
Spanish	218	28%	40	15%	14	22%	104	79%	31	19%	29	17%
Khmer	39	5%	0	0%	0	0%	0	0%	39	24%	0	0%
Other[a]	87	11%	38	15%	0	0%	1	1%	43	27%	5	3%

Source: Data obtained from the Massachusetts Department of Early Education and Care for all 48 PEG classrooms during Fall 2016. Percentages may not add up to 100 because numbers are rounded to the nearest whole.

[a] Other common languages included (primarily in Boston) Cape Verdean, Chinese, and Haitian Creole, and (primarily in Lowell) Portuguese, Vietnamese, and Arabic.

PEG classrooms served a small population of children with Individualized Education Program (IEP) plans, formal plans developed by public school special education staff to guide special education services received by eligible children. The goal was to target enrollment so that at least seven percent of the children in each PEG classroom have an IEP; at the end of the 2016-17 PEG year, almost six percent of children had one in place.

3. PEG Impact Study Design

3.1 Introduction

This study of the impacts of the PEG program was part of a multi-year evaluation being conducted for the Massachusetts Department of Early Education and Care by Abt Associates over the four years of program implementation. The evaluation looked annually at the implementation of PEG and the outcomes for children, parents and staff. The study of the impact of PEG on children focused on a single cohort of children in one year of PEG.

The research questions for the impact study are about the effects of PEG on three domains of child development:

- What is the impact of the PEG program on children's early academic skills (literacy and math)?
- What is the impact of the PEG program on children's language development (vocabulary)?
- What is the impact of the PEG program on children's executive function skills?

The study used an age-cutoff regression discontinuity design (RDD), a methodology popular for evaluating the impact of preschool programs where true randomization (i.e., randomly assigning children to different preschool programs or to preschool versus no preschool) is not feasible. RDDs can be used to estimate the impact of preschool programs that have a strict age requirement for admittance, such that children who fall on either side of the age cutoff form groups that come close to randomly assigned groups in terms of their assumed similarities. When done correctly, RDDs are now generally recognized as superior to other quasi-experimental (i.e., non-randomized) designs for addressing questions related to program impact. Because PEG has a strict age cutoff for eligibility, an RD design can be used.

The first use of a RDD to study the impact of an early childhood program was the evaluation of the Tulsa, Oklahoma public preschool program (Gormley, Gayer, Phillips, and Dawson, 2005). In that landmark RDD, authors reported large statistically significant effects on the children in the program—an effect of .79 standard deviation units on early literacy skills and .38 standard deviation units on early math skills. Since the Tulsa study, there have been several RDD studies of preschool programs across the country, most examining publicly-funded prekindergarten programs operated by school districts (Bartik, 2013; Lipsey, Farran, Bilbrey, Hofer, & Dong, 2011; Peisner-Feinberg, Schaaf, LaForett, Hildebrandt, & Sideris, 2014; Weiland & Yoshikawa, 2013). Across these evaluations, similar positive and statistically significant impacts on children's early academic skills were found.

3.2 Methods

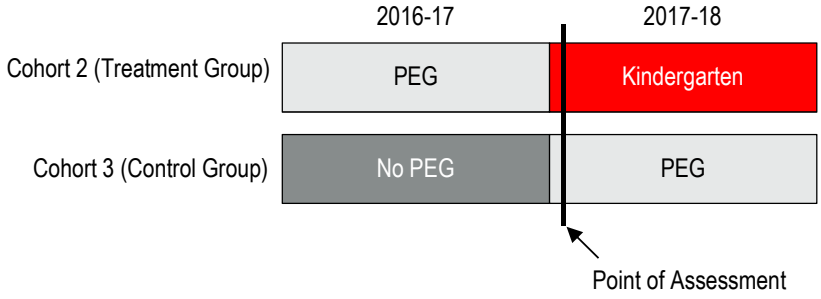
3.2.1 Design

The study of the impacts of the Massachusetts PEG program uses a RDD that takes advantage of the fact that PEG requires that children have reached their fourth birthday by September 1st of the enrollment year and are not yet five years of age. The study contrasts the performance of a cohort of PEG children whose birthdays fall just before the September 1 cutoff date for enrollment in 2016-17 (Cohort 2, the treatment group) versus the performance of a cohort of children with birthdays just after the cutoff date; that is, they were too young to enroll in PEG that year and instead enrolled in PEG in 2017-18 (Cohort 3, the control group).

To understand the age cut-off RDD approach, imagine two children, one who turns four years old on September 1st and is eligible for PEG, and one who turns four a day later, on September 2nd, and thus is not eligible for PEG until the following year. These two children progress through the 2016-17 year having two different experiences—the former gets PEG and the latter does not. In all other observed and unobserved ways, the two children are assumed to be essentially identical. It is this assumption that allows for an age cutoff RDD to produce an estimate of program impact similar to that produced by a randomized study—the RDD compares children who receive the intervention versus very similar children who have not yet received it. Where a random assignment study would randomly determine which students were in those two groups, an RDD study capitalizes on the existing age cutoff as the method of assignment.

In this study, parents of all enrolled children (treatment and control cohort) were contacted for their consent at the time of their children's enrollment in their respective years; the treatment group at the beginning of 2016-17 and the control group at the beginning of 2017-18. Exhibit 3.1 displays the timeline for the RDD.[2]

Exhibit 3.1: Timeline for PEG RDD



Source: Figure adapted from Lipsey et al., 2015, Figure 1.

The analysis sample included 1,107 children, 582 in the treatment group and 525 in the control group, which represents 81 percent of the consented children (see Exhibit 3.2). The analysis sample includes children from all 48 PEG classrooms. On average, each classroom was represented in the analysis sample by 23 children across treatment and control groups. The number of treatment children per classroom ranged from three to 20 with an average of 12; the number of control children per classroom ranged from five to 18 with an average of 11. There were at least three treatment and three control students in the analysis sample from each classroom.[3]

Exhibit 3.2: Analysis Sample

	Treatment Group N (% of consented)	Control Group N (% of consented)	Total N (% of consented)
Total Enrollment	788	783	1571
Total Consented	703	670	1373
Total Analysis Sample	525 (75%)	582 (87%)	1,107 (81%)

Note: Some of the consented children were removed from the analysis sample because they were determined to be ineligible for a variety of reasons: failure to meet PEG age-eligibility criteria (n=8); late enrollment or early withdrawal (n=62); receipt of consent after the assessment window had closed (n=67), or inability to assess (repeated absences, ultimate parent refusal, unable to locate kindergarten placement, etc. (n=129). Further description and justification for the exclusions from the analysis sample based on different eligibility requirements is provided in the Appendix.

2 Additional details about the implementation of the RDD are in the Appendix.

3 The Appendix shows analysis sample numbers by classroom and community for both groups.

3.2.2 Outcomes

The study used standardized norm-referenced measures to assess children's early literacy and math skills and vocabulary; a nonstandard but widely-used measure to assess executive function skills. The battery of measures is described below.

Vocabulary. Children's receptive vocabulary knowledge was measured with the Peabody Picture Vocabulary Test—Fourth Edition (Dunn & Dunn, 2007). The test measures children's receptive (listening) vocabulary skills, and is often thought of as an indicator of overall cognitive performance. The child is shown a card with four pictures on it, and selects the picture that best illustrates the meaning of a stimulus word spoken by the assessor.

Early Literacy. Children's early literacy skills were measured with the Woodcock-Johnson III Tests of Cognitive Abilities: Letter-Word Identification Subtest (Woodcock, McGrew, & Mather, 2001). The subtest measures early letter and word reading skills, specifically. The child is asked to identify individual letters and read individual words of increasing difficulty.

Early Math. Children's early mathematics skills were measures using the Woodcock-Johnson III Tests of Cognitive Abilities: Applied Problems Subtest. The subtest measures the ability to count and solve problems related to numeracy and space. The child hears a story problem and is asked to recognize the mathematical procedure that should be used and to perform the appropriate calculation.

Executive Functioning. Children's executive functioning was measured with the Hearts & Flowers Task (previously called the Dots Task; Davidson et al., 2006; Diamond et al., 2007), which measures children's ability to remember rules and to inhibit their response when applying those rules under different contexts. Its three types of tasks range in difficulty (congruent tasks, which are the easiest; incongruent tasks; and mixed tasks, which are the most difficult). Using a tablet, the child is shown either a picture of a heart or a flower on either the left or right side of the screen. The assessor instructs the child to push a button, sometimes on the same side of the screen as the picture and sometimes on the opposite side of the screen as the picture. The rules change as the game progresses.

The impact analyses used raw scores from each of the measures—that is, scores that are not age-adjusted.[4] The three academic measures each produce a single overall score. The Hearts and Flowers measure produces three raw scores; this analysis used only the score for the mixed task, the most difficult of the three.

3.2.3 Assessment Procedures

Children's skills were assessed over a three-month period in fall of 2017 by testers who were trained and certified as meeting required reliability thresholds. Most children were assessed within a single assessment visit lasting no more than 45 minutes. All assessments included in the main analyses were administered to children in English, regardless of the child's home language or English proficiency, so as to obtain the same score(s) on all children in the analysis sample.[5]

---

[4] Raw scores were used for the Peabody Picture Vocabulary Test, and W-scores were used for the two Woodcock-Johnson III subtests. W-scores are provided as part of the technical manual. These scores are a linear transformation of the raw score; they are not adjusted for age but provide greater variation than just the raw score distribution.

[5] A portion of non-English-speaking children were also assessed with Spanish and bilingual versions of some of the measures, and those data are being analyzed as part of the longitudinal study component of the PEG evaluation.

3.2.4 Analysis

The glossary of terms in the textbox lists common terms used to describe the analytic approach in this section.

Pre-Analysis Data Examination

Prior to conducting the impact analyses, the data were examined in multiple ways to confirm essential RDD assumptions and guide choices of impact analysis models. This examination had three primary steps: (1) graphing the relationships between age and outcomes at the age cutoff to check for visual discontinuity at that point (suggesting a program impact) and no other visual discontinuities at other points (suggesting an RDD might not be appropriate); (2) visually checking for the appropriate functional form for the relationship between age and outcome (guiding how this relationship was modeled in main effects models); and (3) testing the distribution of children in the two conditions and the five communities on the three key child demographic covariates (gender, home language, and prior care) to look for evidence of differences in demographic make-up by condition overall and by community (suggesting that the RDD assumption of equality on everything except age and exposure to the program might not be supported).[6]

RDD Glossary of Terms

Global: a regression model that includes all students in the analysis sample

Bandwidth: the time frame (number of days) around the cutoff within which students are selected for the analysis

Limited bandwidth: a regression model that focuses on only those students whose birthdays fall within a given bandwidth

Functional form: the form of the relationship (linear or quadratic, here) of children's skills and their age relative to the cutoff

Fixed effects: the inclusion of a set of dummy codes in the regression models that represents each PEG classroom, included to control for variation in the outcome due to between-classroom differences

Primary Impact Models

The primary impact model to test the overall impact of PEG on each of the four child outcomes used a linear global regression model that included three child covariate controls (gender, home language English or not, prior child care or not) and classroom fixed effects.[7] The analysis sample included all children, regardless of how far away they fell by age from the age cutoff. By including all children in the analysis, the primary impact models represent the best-powered analyses for the study and therefore are the results that can be reported with the most confidence.

Sensitivity Analyses

The study conducted an initial set of analyses to continue to test the assumptions required for a valid RDD model. These analyses examined the effect of attrition and missing data on the sample overall and examined the density of ages across the age span. These analyses found no evidence of differential attrition or missingness.[8]

Subsequently, the study conducted an extensive set of analyses that examined the robustness of the main effects to various analytic decisions, in line with recommendations by the Department of Education's What Works Clearinghouse.[9] Sensitivity analyses included comparing the results from linear and quadratic regressions and also varying the models as follows: (1) assessing the difference in effects obtained when using an analysis sample made up of children weighted differently depending on their distance from the age cut-off, with children close to the age cutoff given the greatest weight; (2)

---

6 The various forms of data examination are described in the Appendix.

7 The Appendix describes these parameters in more detail.

8 The details of these analyses are included in the Appendix.

9 See *What Works Clearinghouse™ Standards Handbook Version 4.0* (2018).

comparing effects obtained when applying bandwidths of different shapes; (3) using instrumental variables to compare effects using samples with and without the eight PEG-ineligible cases (children who were enrolled in the treatment or control group, but were too young or old according to their date of birth) and the 62 children who enrolled in PEG too late; and (4) comparing effects from models with and without covariate controls.

Subgroup Analyses

In addition to estimating the main effect of the PEG program, the study conducted analyses to test for whether the impact differed for different subgroups of children defined by gender, home language, and prior child care. These analyses were exploratory, given that they compared smaller subgroups of children whereas the study was powered to reliably detect effects only for analyses that used the entire sample of children. Linear and quadratic regression models were run with terms for the interaction of treatment with each child covariate [gender (boys, girls), home language (English, not English), and prior child care (any prior care, no prior care)]. The models that were run alternated the child subgroup reference category. The study also performed sensitivity analysis of the child subgroup differences.[10]

Missing Data

Low percentages of data on children's outcomes or baseline characteristics were missing. Variables for which data were missing included gender (missing for three children including two in the treatment group and one in the control group), home language (missing for three children, all in the control group), prior child care exposure (missing for four children including one in the treatment group and three in the control group), early literacy outcomes (one child in the treatment group), and executive function outcomes (one child in the treatment group, across the three constructs). Because of the paucity of missing data, imputation was not done and case-wise deletion was employed when appropriate.

[10] The sensitivity analyses conducted are detailed in the Appendix.

4. Results

4.1 Descriptive Statistics

Exhibit 4.1 presents descriptive information about the demographic characteristics of both the treatment and control groups in the RDD at the time of their enrollment in PEG. On all variables except prior care, the treatment and control samples were nearly identical.[11] Unadjusted scores on outcome measures for both groups are included in the Appendix.

Exhibit 4.1: Demographics by Condition at Study Enrollment

	Full Sample (n=1107)	Birthday Before Cut-off (Treatment Group; Attended PEG in 2016-17; n=582)	Birthday After Cut-off (Control Group; Attended PEG in 2017-18; n=525)
Demographics	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Age at Cutoff (in months)	47 (6.86)	53 (3.47)	41 (3.51)
Female (%)	50%	50%	50%
English Home Language (%)	59%	59%	60%
Black (%)	22%	22%	22%
Hispanic (%)	61%	60%	62%
White (%)	5%	6%	5%
% With Prior Child Care Exposure: 4 Communities that Targeted Those Without Prior Care	7%	3%	12%
% With Prior Child Care Exposure: All 5 Communities	28%	23%	33%

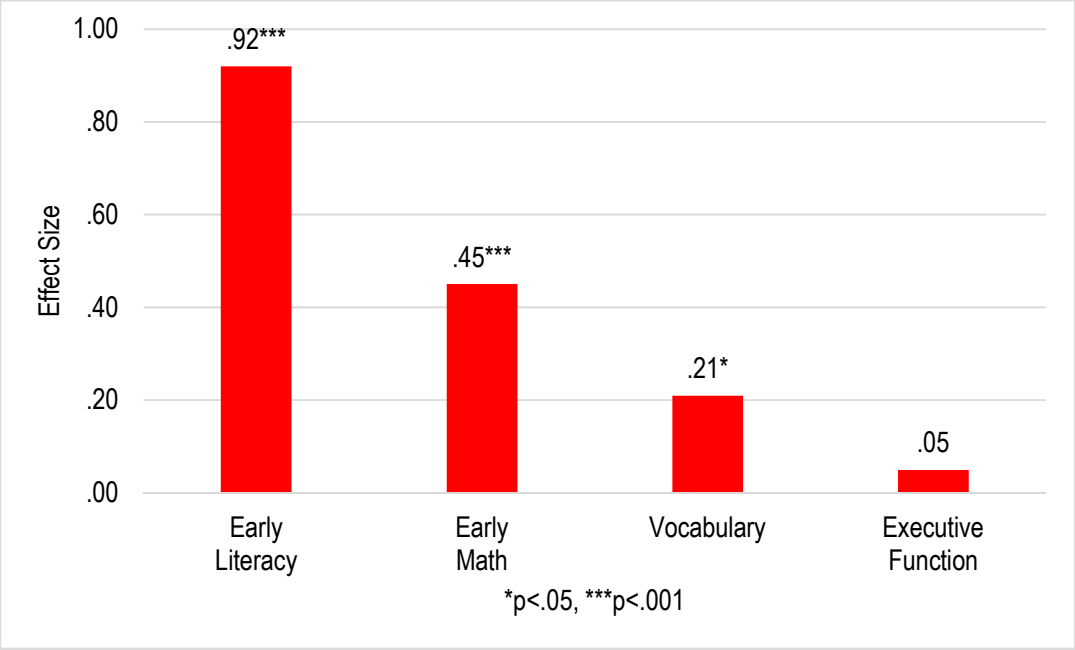
4.2 Main Effects

For all outcomes, positive effect sizes mean that treatment children had higher performance than control children. The standardized effect sizes are presented graphically in Exhibit 4.2, in descending order of impact size. Full model results can be found in the Appendix.

- On the three measures of early academic performance, PEG had a positive and statistically significant impact on children's achievement. The largest impact was seen for early literacy skills; the smallest effect was for vocabulary. Effects on early literacy and early math skills were large enough to be robust to variations in the analytic model; effects on vocabulary were smaller and less robust but still statistically significant in the main effects model. For these skills, there was a significant benefit of participating in PEG.
- On the executive function task, the effect of PEG was not statistically significant.

[11] In 2017-18, a change in state policy led to a slightly higher percentage of families with prior care enrolled in PEG.

Exhibit 4.2: PEG Impact across Child Outcomes (in Standard Deviations)

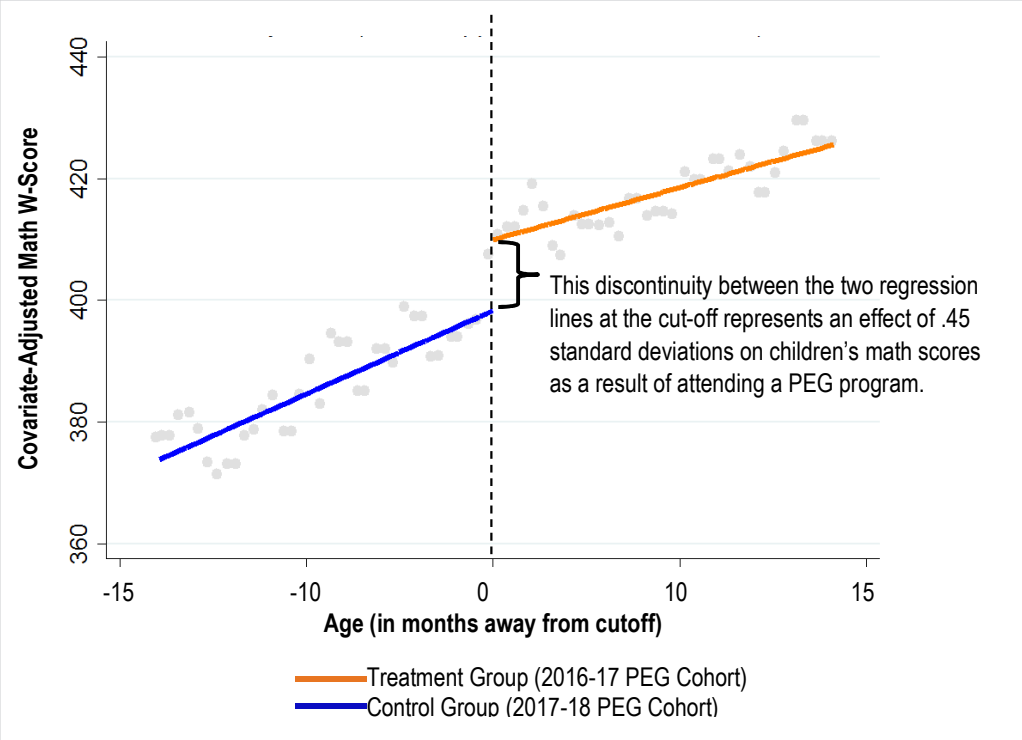


*p<.05, **p<.01, ***p<.001

To illustrate the effect of PEG in the RDD context, Exhibit 4.3 shows the relationship between age and predicted early math scores for the full analysis sample. The 'jump' in the regression line at the cutoff demonstrates the effect of PEG.[12]

[12] Similar graphs for the other three key outcomes are in the Appendix.

Exhibit 4.3: Demonstration of the Discontinuity (PEG Effect) on Early Math Scores



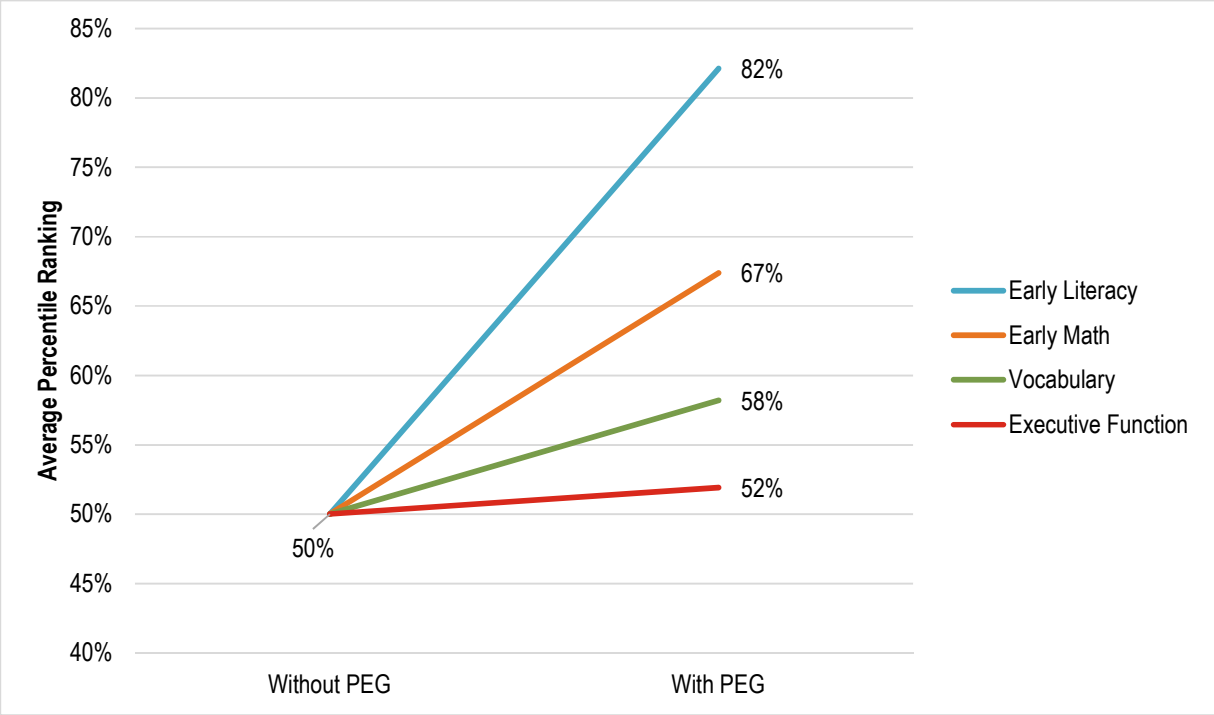
Contextualizing the Effects

Effect sizes are useful because they allow for the valid comparison of impacts across studies regardless of variation in participants, treatment, and outcome scale. However, they often do not provide the context within which to situate the meaningfulness of the impact. To that end, below are three methods of conceptualizing the main effects of the PEG RDD.

Improvement Indices

The What Works Clearinghouse translates effect sizes into "improvement index" values to help contextualize the size of the findings. Exhibit 4.4 below shows the calculated improvement index associated with each of these effect sizes. The improvement index can be interpreted as the expected change in percentile rank for an average control group student if the student received PEG. For example, the improvement index for early literacy is 32.12, which means that PEG moved the performance of the average student from the 50th to the 82nd percentile; in other words, the average student would score better than 50 percent of his/her peers on the early literacy assessment if he/she did not experience PEG, but that same student would score better than 82 percent of his/her peers if he/she did attend a PEG program.

Exhibit 4.4: PEG Effect on Average Student Percentile Ranking



Comparison to What Works Clearinghouse Effects

The What Works Clearinghouse (WWC) reports effect sizes from the research it reviews on various education-related programs. Compared to the average effects reported for 165 studies of early childhood interventions for children age two to six years, the effect sizes for PEG impacts could be considered large. The effect size for the impact of PEG on early literacy (.92) is larger than 88 percent of WWC impacts; the PEG effect size for the impact on early math (.45) is larger than 77 percent of WWC impacts and the impact on children's vocabulary scores (.21) is larger than 61 percent of WWC impacts.

Comparison to Other Findings

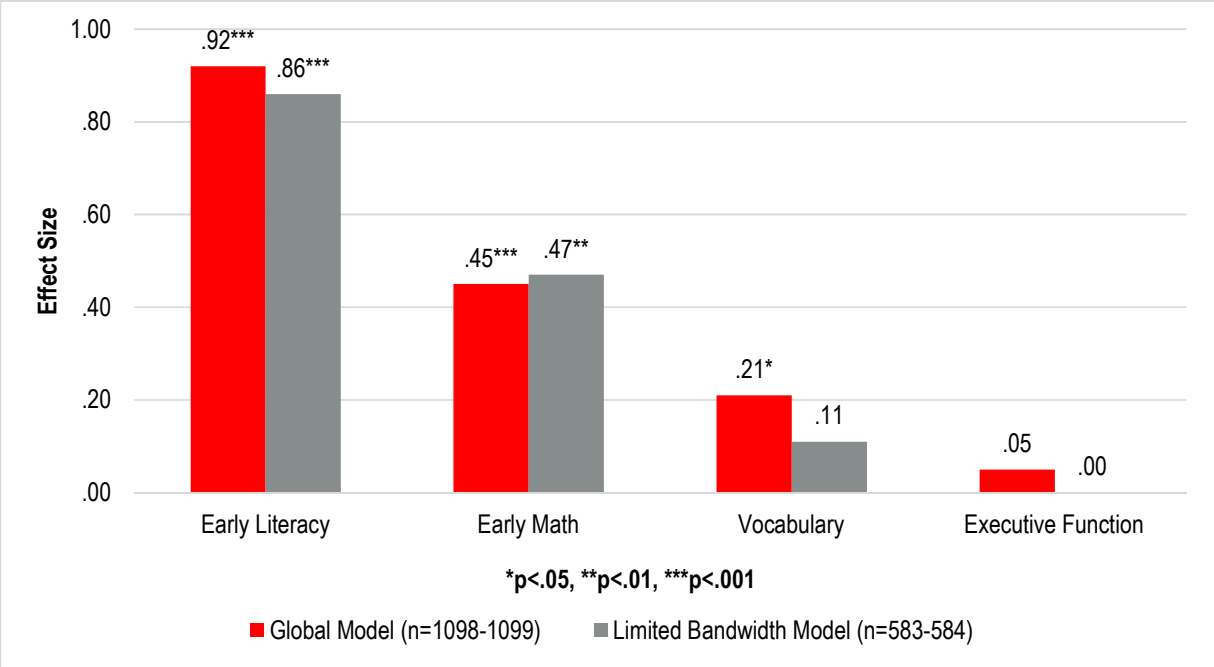
The results of this study can also be compared to effect sizes reported in a meta-analysis of over 300 effect sizes from 38 evaluations of center-based early childhood education programs serving children ages 3 to 5 in the United States, conducted between 1960 and 2007 (Bowne et al., 2017). The authors of that meta-analysis reported an average effect size of program impacts on children's socioemotional outcomes of 0.17, and an average effect on cognitive/achievement outcomes of 0.31. The effect sizes for the impact of PEG on children's early math and literacy skills are considerably larger than what the Bowne et al. study reports, whereas the effect sizes for the impacts on vocabulary and executive function skills are lower.

4.3 Stability of the Effects: Results of Sensitivity Analyses

Sensitivity analyses compared the PEG effects in the main analysis using the full analysis sample versus the effects obtained with the same models but using samples representing different bandwidths around the age cutoff—for example, a sample of children whose age was within 190 days before or after the cutoff. These are the children who are likely to be more similar to one other than the groups that include the full age range.

These analyses showed that some of the PEG effects were sensitive to bandwidth. For the sample of children whose birthdates were within the 190-day bandwidth, the effects on early literacy and early math were similar in size and statistically significant, the effect on vocabulary was smaller and was no longer significant (Exhibit 4.5). The robustness of the effects on early literacy and early math to variations in the model warrants more confidence in the program impact on those skills.

Exhibit 4.5: Comparison of PEG Impacts in Full Sample and Limited Bandwidth Sample



*p<.05, **p<.01, ***p<.001

4.4 Comparison of PEG Effects to Other Early Childhood RDD Studies

Because other pre-kindergarten RDD studies measured the same early academic skills as were measured for this evaluation, the Massachusetts PEG results can be compared to results from similar studies reported in the literature. The impacts of PEG and the other pre-kindergarten programs studied using RDDs were very similar in size on children's early literacy and math achievement (Exhibit 4.6). The impact of PEG on vocabulary achievement was similar to the effect from a recent analysis across eight states, yet smaller than the effects reported in the RDD studies in Boston, Tulsa and Tennessee.

Exhibit 4.6: Effect Sizes on Children's Outcomes in Other PreK RDD Studies

Study	Early Literacy	Early Math	Vocabulary
MA PEG	**.92**	**.45**	**.21**
Boston [a] (Weiland & Yoshikawa, 2013)	.62	.49-.58	.45
Tulsa [b] (Gormley, Phillips, & Gayer, 2008)	.79	.38	n/a
Tennessee [c] (Lipsey, Farran, Bilbrey, Hofer, & Dong, 2011)	.82	.48-.50	.48
Eight State PreK Analysis [d] (Barnett et al., 2018)	1.02	.53	.25

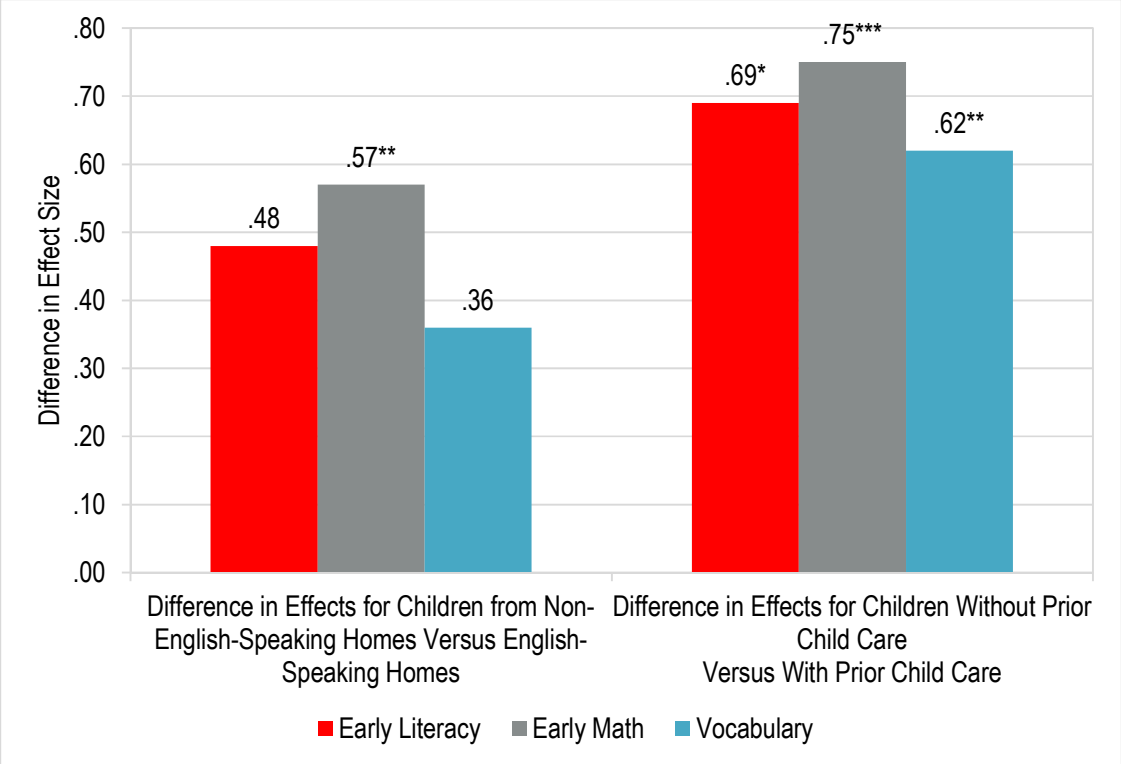
[a] Sample included 2018 students; 69% of the sample qualified for free/reduced-price lunch; 50% of the sample spoke a language other than English.

[b] Sample included 4716 students; 65% of the sample qualified for free/reduced-price lunch; 11-18% were Hispanic.
[c] Sample included 1358 students; majority were from low-income families; 10-14% were English Language Learners.
[d] Sample included over 4,000 students; majority were from low-income families; 10-14% were English Language Learners; income and ethnicity varied widely across the eight states.

4.5 PEG Subgroup Effects

The exploratory analyses examining differential program effects by child demographics suggested that PEG was more effective for subgroups defined by home language and prior care, but not by gender (Exhibit 4.7).[13] Across the three academic outcomes, PEG impacts were larger for children whose home language was not English than for those whose home language was English. Although the differences were apparent on all of the outcomes, the difference was only statistically significant for early math (p=.007). Across all three academic outcomes, PEG impacts were larger for children who did not have any parent-reported formal care before entering the PEG program and the differences were significant for all three outcomes.

Exhibit 4.7: Difference in PEG Impact by Child Demographic Subgroup



*p<.05, **p<.01, ***p<.001

[13] The Appendix includes tables with all model parameters and impact estimates for each of the child subgroup analyses, including gender.

5. Discussion

The Massachusetts Preschool Expansion Grant (PEG) had positive statistically significant impacts on children's academic skills. Effects on early literacy (.92) and early math skills (.45) were large enough to be robust to variations in the analytic model. Effects on vocabulary were smaller but still statistically significant (.21). The evaluation did not find evidence of a significant effect of PEG on children's executive function skills. PEG successfully increased children's kindergarten readiness skills related to early math and early literacy such that, at kindergarten entry, they were much closer to where they could be expected to score given their age than they would have been had they not experienced PEG. Exploratory analyses considering differential program effects by child demographics suggested that PEG was more effective for some of the children most at-risk in the formal educational system: those whose primary home language was not English and those without formal prior early childhood education.

The lack of impact of PEG on children's executive function skills is not completely surprising, given the inconsistent findings on socio-emotional skills from other evaluations of prekindergarten programs. Though the study of the Boston prekindergarten program reported an effect of .20 on children's inhibitory control, other quasi-experimental (and non-RDD) studies reported mixed findings (Gormley, Phillips, Newmark, Perper, & Adelstein, 2011; Magnuson, Ruhm, & Waldfogel, 2007). Though none of these programs, including the Massachusetts PEG program, focused explicitly on building children's regulatory skills, the authors of the Boston study hypothesized that the structured literacy and math curricula used in all of the Boston classrooms had a spillover effect on children's regulatory skills (Weiland & Yoshikawa, 2013).

The PEG model was ambitious in the scope of its vision, and implementation data indicate that participating LEAs and ELPs were able to quickly implement multiple quality components in order to provide a supportive environment for both educators and families, as well as a rich learning environment for children. The educator supports developed and offered as part of the local collaborative partnerships in the PEG communities built the instructional capacity of PEG educators through multiple job-embedded professional learning opportunities, including training and coaching, and paid release time for instructional planning and collaboration. Over the course of the PEG grant, LEAs and ELPs also increased the alignment across the different forms of professional learning (i.e., training and coaching) and the coherence of the professional learning, classroom curriculum, and assessments. Another notable component of the PEG model was the employment of well-educated staff who were provided with levels of compensation that maintained parity with the local school districts.

The combined set of supports for educators were hypothesized to support teacher retention. Over the first three years of the PEG program, retention improved; about 75 percent of PEG lead teachers remained in classrooms between years one and two and about 90 percent remained between years two and three.

The average statewide PEG classroom quality, as measured by the Classroom Assessment Scoring System (CLASS), reflected moderate to high levels of quality. The average scores statewide for two of the CLASS domains—Emotional Support and Classroom Organization—reflected a level of quality that was close to "high" as defined by the developers of the measure (scores of 5.9 and 5.7, respectively). The average score for the domain Instructional Support reflected "moderate" quality, and compares favorably to other national samples. Importantly, progress has been made in bringing up PEG classroom quality

ratings for classrooms that were initially at the lower end of distribution. More details about the implementation of the PEG program in year two are available in a separate report.[14]

It is notable that the impacts for PEG on children's early literacy and math are similar in size to the impacts found in RDDs of primarily public school district-operated preschool programs, given that the PEG classrooms were operated by community-based agencies. Furthermore, PEG programs use a variety of curricula and offer a range of professional development supports to teachers, as well as supports for families.

This research provides important information to the field about the feasibility of implementing high-quality preschool through a mixed delivery system and potential effects of the model. As is true for most other preschool models, the Massachusetts PEG program delivered a combination of programmatic features that alone or together might drive impacts on children, including but not limited to standardized curricula aligned with learning standards, teacher coaching and professional development, and improved teacher compensation. The evaluation was not able to rigorously disentangle which levers caused the detected impacts, although further exploratory research is underway to try to better understand the relationship of the implementation of particular program features to children's outcomes.

In sum, this evaluation provides additional evidence about the benefits of high-quality prekindergarten for children from disadvantaged backgrounds. The federal PEG grant gave the Commonwealth of Massachusetts a unique opportunity to test the feasibility of providing high-quality prekindergarten through local collaboration across a mixed delivery system and, after two years of implementation, yielded substantial impacts on children's academic school readiness.

[14] The *Year 2 Annual Evaluation Report* can be found at: https://www.abtassociates.com/insights/publications/report/year-1-massachusetts-preschool-expansion-evaluation-report.

References

Barnett, S., Jung., K. Friedman-Krauss, A., Frede, E.C., Nores, M., Hustedt, J. T., & Howes, C., and Daniel-Echols- M. (2018). State Prekindergarten Effects on Early Learning at Kindergarten Entry: An Analysis of Eight State Programs, *AERA Open,* 4 (2), 1–16.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 179,* 1802–1817.

Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis, 39* (3), 407-428.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12,* 140-153.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25,* 166-176.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica, 82* (6), 2295-2326.

Clements, D. H., & Sarama, J. (2007b). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. Journal for Research in Mathematics Education, 38, 136-163.

Clements, D. H., & Sarama, J. (2008). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. Early Childhood Research Quarterly, 19, 99-120.

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

Duncan, G. and Magnuson, K. (2013). Investing in preschool programs. Center for Economics and Public Policy. Discussion Paper Series. Irvine, CA. University of California-Irvine.

Duncan, G.J., Magnuson, K., and Votruba-Drzal, E. Moving beyond correlations in assessing the consequences of poverty. *Annul Rev. Psychol*., 2017, 68: 10.1-10.22. https://cloudfront.escholarship.org/dist/prd/content/qt4h261200/qt4h261200.pdf

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78,* 558-580.

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. *American Educational Research Journal, 48,* 763-793.

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

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

Gormley, W.T., Phillips, D.A., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. *Child Development*, 82, 2095-2109.

Heckman, J., Pinto, R., & Savelyev, P. A. (2012). *Understanding the mechanisms through which an influential early childhood program boosted adult outcomes* (NBER Working Paper No. 18581). Cambridge, MA.

Kamat, V. (2018). On nonparametric inference in the regression discontinuity design. *Econometric Theory, 34* (3), 694-703.

Larsen, J.M., & Robinson, C.C. (1989). Later effects of preschool on low-risk children. *Early Childhood Research Quarterly*, 4, 133-144.

Leak, J., Duncan, G.J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010, March). Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program. Paper presented at the Biennial Meeting for the Society for Research in Child Development.

Lipsey, M.W., Hofer, K. G., Dong, N., Farran, D.c., & Bilbrey, C. (2013). *Evaluation of the Tennessee voluntary prekindergarten program: End of Pre-K results from the randomized control design*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

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

Magnuson, K.A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review, 26(1),* 33-51.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P, Jenkins, F., Fletcher, P., Quinn, L., Friedeman, J., Ciarico, J., Rohacek, M., Adams, G., & Spier, E. (2010). *Head Start Impact Study: Final report.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Preschool Curriculum Evaluation Research Consortium (2008). Effects of preschool curriculum programs on school readiness (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Zhai, F., Metzger, M., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 77,* 302–316.

Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science, 7,* 91-102.

Schindler, H.S., Kholoptseva, J., Oh, S.S., Yosikawa, H., Duncan, G.J., Magnuson, K.A., & Shonkoff, J.S. (2015). Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis. *Journal of School Psychology, 53* (3), 243-263.

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

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*, 2112–2130.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W., Ludwig, J.O., Magnuson, K.A., Phillips, D.A., & Zaslow, M.J. (2013). Investing in Our Future: The Evidence Base on Preschool Education. New York: Foundation for Child Development and Ann Arbor, MI: Society for Research in Child Development. Available at: http://fcd-us.org/sites/default/files/Evidence%20Base%20on%20Preschool%20Education %20FINAL.pdf.

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Davieri, P., Green, K., Cavadel, E., Tarullo, L., Burchinal, M. and Martinez-Beck, M. (2016). Quality thresholds, features, and dosage in early care and education: Introduction and literature review. In Burchinal, M., Zaslow, M., and Tarullo, L (Eds.). Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes. *Monographs of the Society for Research in Child Development*.

Appendix

Global Regression Results: Main Effects

Model Parameters

All main effects models included the following parameters:

Treatment indicator: In RDD models, there is one key variable that measures the effect of the treatment, which is age eligibility (an indicator for age of at least 4 at the cutoff). Together with other age variables (either linear or quadratic terms on each side of the cutoff), that indicator for age eligibility models the effect of the treatment in the context of this type of design. For example, in a linear model, indicators were included for participation (measuring the jump at cutoff), distance from cutoff in age (measured in days away from the cutoff), and the interaction of the jump at cutoff and the distance from the cutoff (which measures the differences in slopes).

Key child-level demographics: A key assumption in an RDD is that children in the treatment and control group, particularly very close to the cutoff, are similar to one another in all measured and unmeasured ways except for age and exposure to treatment. Under this assumption, it is unnecessary to adjust for covariates, but adjusting for covariates can improve precision. Therefore, all three child covariates were included to account for any variation not controlled for by the design. Analyses routinely checked for bias in the impact estimate related to the inclusion of child-level covariates and did not find evidence of meaningful bias.

Classroom-level nesting: Classroom-level fixed effects were included for each of the 48 classrooms. These do not address bias in RDD models, but serve to increase precision, to the extent that mean achievement differs systematically across classrooms. Further, this classroom-level nesting accounts for ELP- and LEA-level differences even without including terms for those levels which would only introduce collinearity issues into the models.

Results of the Main Effects Model

The results shown below use a global regression model, meaning the full analytic sample. Under each estimate, the exhibit shows the parameter estimate for the test that that the coefficient is zero, robust to clustering at the classroom level. In each model, the coefficient on linear time (age) is positive, indicating the natural growth in test scores with age, which is exactly why one would not want to compare raw test scores in the treatment group (who are uniformly older) to the control group (younger) without controlling for age. Exhibit A.1 also shows the standard error of the treatment estimate. The associated t-statistic can be obtained as the ratio of the coefficient on the treatment to the standard error; where the resulting t-statistic is greater than 2.0 means that the null hypothesis that the coefficient is zero should be rejected. In the models in Exhibit A.1, in addition to each parameter shown in the table, the model also controlled for the fixed effects of classroom with a series of dummy codes. Also of note is the interaction of time and the treatment indicator, which often has a negative but statistically insignificant estimate. This interaction captures the regression to the mean of effects at the cutoff, though the interpretation of this coefficient does not have the sharp causal interpretation supported by comparisons at the cutoff in an RD design.

Exhibit A.1. Results of Main Effects Models (Parameter Estimate, Standard Error, and Indication of Significance)

Parameter	Early Literacy	Early Math	Vocabulary	Executive Function (Mixed Trials)	Executive Function (Congruent Trials)	Executive Function (Incongruent Trials)
Treatment	**24.54*** **	**11.33*** **	**4.93***	**.01**	**.00**	**-.04**
	(3.83)	**(2.47)**	**(2.21)**	**(.03)**	**(.02)**	**(.04)**
Age (Distance from Cut-off)	.05***	.07***	.07***	.00***	.00***	.00***
	(.01)	(.01)	(.01)	(.00)	(.00)	(.00)
Treatment by Age Interaction	-.02	-.03*	-.01	.00	-.00**	-.00
	(.02)	(.01)	(.01)	(.00)	(.00)	(.00)
Female	1.87	4.65**	4.86**	-.00	.04*	.03
	(1.39)	(1.53)	(1.40)	(.01)	(.02)	(.02)
English as Home Language	3.73*	9.98***	16.52***	.01	.00	-.00
	(1.51)	(2.03)	(1.66)	(.01)	(.02)	(.02)
Prior Childcare Exposure	7.18**	5.64*	5.14*	-.01	-.02	-.06*
	(2.63)	(2.51)	(2.50)	(.02)	(.02)	(.03)
Constant	315.00***	386.00***	48.21***	.60***	.82***	.67***
	(2.06)	(2.30)	(1.76)	(.02)	(.02)	(.03)

*p<.05, **p<.01, ***p<.001

Notes. Models were global regression models with linear functional form and also included a set of dummy codes for classroom. Statistics are rounded to two decimal places.

Details about the RD Design

Children were eligible to enroll in PEG in a given year if they turned four years old by September 1 of that year. The RDD takes advantage of this age cut-off to compare outcomes from children at the end of one year of PEG to children who have just begun participating in PEG preschool in the next year. Any observed differences between children who fall on opposite sides of the age cut-off are interpreted as estimates of the causal impact of PEG participation.

The fact that four of the five PEG communities primarily targeted children who have never before been enrolled in formal early education of any kind meant that the majority of students who enroll in PEG were not exposed to a formal program in the year prior to their preschool year. This requirement improved the precision of the treatment-control contrast in the RDD study. However, the fifth PEG community used different eligibility requirements for their PEG families, which meant that children could enroll in PEG whether or not they had previously been in other types of formal early childhood education. In the other four PEG communities, the eligibility requirements also relaxed in the 2016-17 school year when programs were not able to fully enroll by a certain date. Because the prior care experience of children is important in determining the impact of PEG, analyses were conducted that interacted previous care experience with treatment to determine if the PEG impact varied as a function of care experiences prior to PEG. Those results are described later in the Appendix.

Sample Eligibility Rules

The necessity of the assessment window in typical age-cutoff RDD studies, where children in both groups are assessed at the beginning of the prekindergarten year for the control group, poses certain difficulties in defining the sample. It is imperative that identical sample eligibility rules are used for both groups to

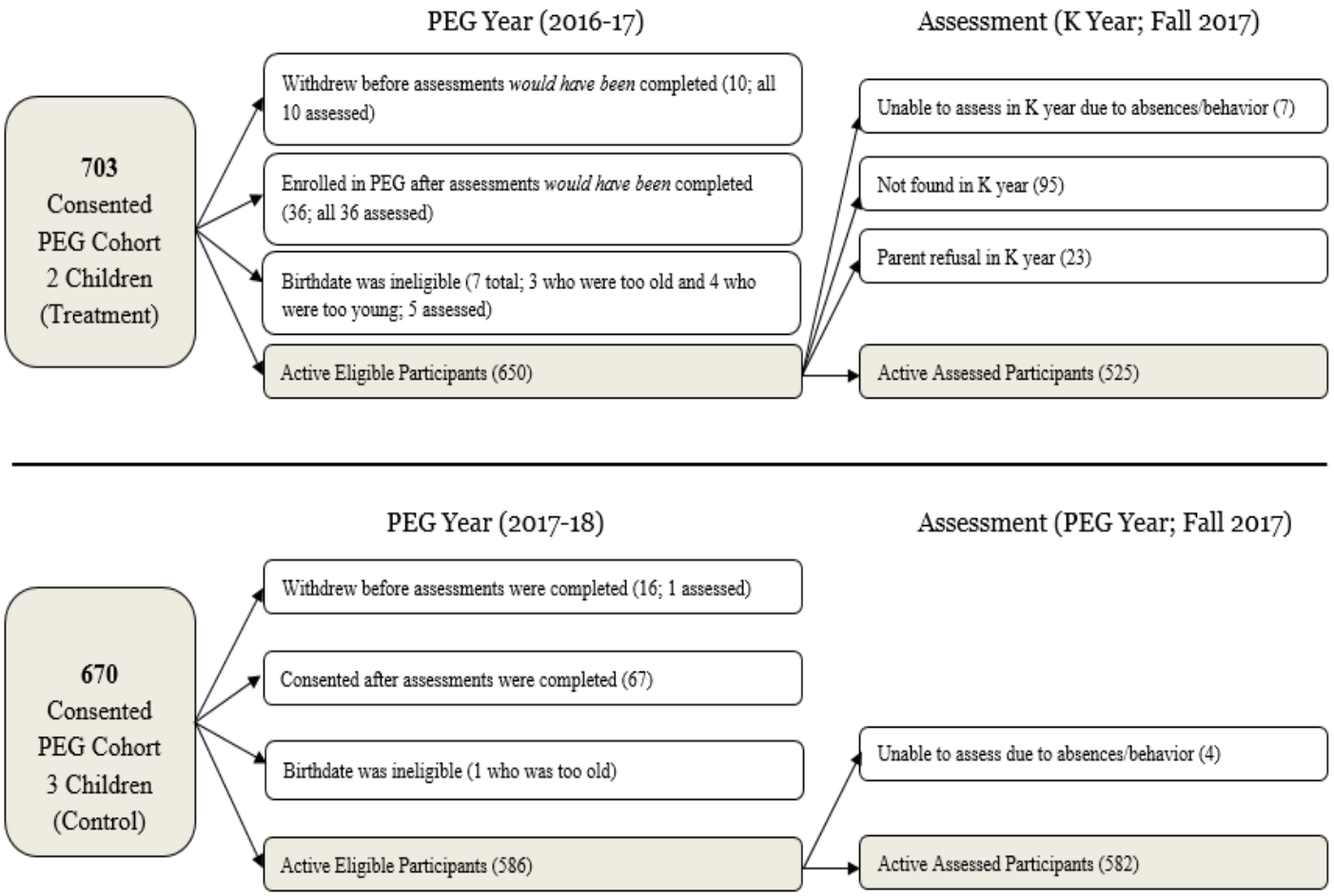
define participants eligible for the analytic sample. Thus, a series of eligibility rules in the PEG evaluation were imposed in order to meet this imperative. Eligibility requirements for inclusion in the analysis sample were:

- PEG Enrollment Date Before November of the PEG Year
  - In the PEG programs, while most children are enrolled within the first weeks of the school year, if classrooms are not filled early in the year or children leave and there are open slots, some children could enroll at another time during the year. Because parental consent for the treatment group was collected at the beginning of the 2016-17 PEG year, enrollment eligibility requirements were applied to both groups in order to include children who enrolled in their PEG year during the same window. To be eligible for the sample for the RD, a child must have been enrolled in the PEG classroom early in the school year, which, for the purposes of the study, was defined as prior to or during the PEG fall assessment window (August 18 – November 10)[15].
- PEG Withdrawal Date Later than November of the PEG Year
  - Children must not have withdrawn from the PEG program prior to the end of the fall assessment window of their PEG year. Kindergarteners who had withdrawn from their PEG program very early in the year would potentially not have been present for assessments had the team conducted assessments in the PEG year. Consequently, the same PEG enrollment period end date criteria was applied to both the treatment and control groups.
- Age Eligible for PEG Program (Turned 4 years of age by September 1 of the PEG year)
  - Children must have birthdates within the range that defines their cohort. For the treatment group, all birthdates were between (and including) September 2, 2011 and September 1, 2012. For the control group, all birthdates were between (and including) September 2, 2012 and September 1, 2013.
- Located in Any Setting in the Kindergarten Year
  - All efforts were made to locate and assess children in the treatment group who did not enroll in the local school district in the year following their PEG exposure. These children were *not* excluded from the sample, provided they could be located and assessed.

The flow of sample participants through the stages from consented to analysis sample is illustrated in the CONSORT chart in Exhibit A.2. There were only a small number of children who were assessed but were not ineligible for PEG based on age, and only 4 out of 703 were too young (the relevant margin for an RDD study). Furthermore, as reflected in the CONSORT chart, the large majority of sample losses were because individuals could not be located for assessment, not for technical reasons or refusal of consent.

[15] Occasionally, a student was assessed after November 10, which was typically due to an earlier partial assessment or multiple absences. The eligibility period was not extended because of these additional assessments. Thirty-five children in the treatment group were assessed by team members from the Expanding Children's Early Learning Network (ExCEL) project, a separate study conducted by MDRC and partners (University of Michigan, Harvard, Boston Public Schools, and Stanford) that overlaps with some of the PEG classrooms, and occasionally those assessments extended beyond the PEG fall assessment window, as well.

Exhibit A.2. Consort Chart



Analysis Sample Numbers by Classroom

This report includes assessments from 1,107 children total (582 children in the control group and 525 in the treatment group). Exhibit A.3 shows this total by community, classroom, and condition.

Exhibit A.3. Analysis Sample Numbers by Classroom and Condition

Community/ Classroom	Control	Treatment
Boston	***186***	***138***
116	5	6
117	14	9
118	18	8
119	10	9
120	16	12
121	16	10
122	10	10
123	12	15
124	11	8
125	12	10
126	16	8
127	9	6
128	12	9
129	14	6
130	11	12
Holyoke	***65***	***53***
305	20	9
306	17	10
307	12	17
308	16	17
Lawrence	***127***	***120***
211	14	15
212	17	12
213	8	8
214	7	9
215	8	9
216	8	9
217	19	18
218	10	9
219	18	18
220	18	13

Community/ Classroom	Control	Treatment
Lowell	***99***	***111***
409	13	16
410	14	13
411	12	11
412	10	12
413	15	16
414	11	11
415	14	15
416	10	17
Springfield	***105***	***103***
512	3	5
513	13	11
514	7	7
515	12	13
516	14	11
517	8	6
518	16	10
519	9	12
520	4	7
521	9	8
522	10	13

Unadjusted Outcome Scores

Exhibit A.4 shows the average unadjusted standard scores or percent correct (for executive function) for the treatment and control groups for the full sample and for a limited-bandwidth sample.

Exhibit A.4. Unadjusted Average Outcome Scores by Condition and Bandwidth Selection

Bandwidth Selection/ Outcome	Control Group	Treatment Group
Full Sample	(n=582)	(n=524-525)
Early Literacy	92.85	97.59
Early Math	93.92	97.57
Vocabulary	85.64	94.09
Executive Function (Mixed Trials)	49.97%	63.45%
Limited Bandwidth Sample (30 days)	(n=42)	(n=52)
Early Literacy	93.93	104.44
Early Math	95.00	100.65
Vocabulary	89.81	91.87
Executive Function (Mixed Trials)	59.52%	59.50%

Notes. Scores are not adjusted for anything other than age at time of test.

Data Examination Prior to Impact Analysis

Graphical Analysis of Discontinuity and Functional Form

The analyses looked at two questions related to discontinuity: (1) Is there evidence of discontinuity in the plotted relationships of age and outcomes at the cutoff (no visible discontinuity would not likely lead to significant impact estimates); and (2) Is there evidence of discontinuity in the plotted relationships of age and outcomes at ages other than the cutoff (which might suggest a threat to the internal validity of the study). Additionally, the analyses addressed a third question about functional form: What is the appropriate form of the analysis model based on the shape of the relationships between outcome and age?

Local linear regressions[16] were plotted for each of the four key outcome measures separately with child age in months and examined graphs (shown below in Exhibits A.5-A.8). Regarding question (1) above, some outcomes exhibited clear discontinuities at the age cutoff and others did not, but it did appear that there was a treatment effect for at least some tested outcomes. Regarding question (2) above, for each outcome, scores vary smoothly and continuously across age and do not exhibit any visual discontinuities at points other than the cutoff, suggesting that the RDD approach is appropriate. Regarding question (3) above, most outcomes appear to be linearly related to age, but there was modest evidence of quadratic curvature in some cases. Cattaneo, and Titiunik (2014) and Kamat (2018) give a variety of reasons to estimate both linear and quadratic models and indicate that estimating both forms helps improve the ultimate precision of the

[16] Local linear regressions in this step used a triangular kernel with a 300-day bandwidth and included child covariates and classroom fixed effects.

treatment effect estimates, and so the study analyzed models with both functional forms (explained later in this Appendix) and examined the robustness of effects across model variants.

Exhibit A.5. Relationship of Age and Outcome: Early Literacy (WJ-III Letter-Word Identification W-Score)

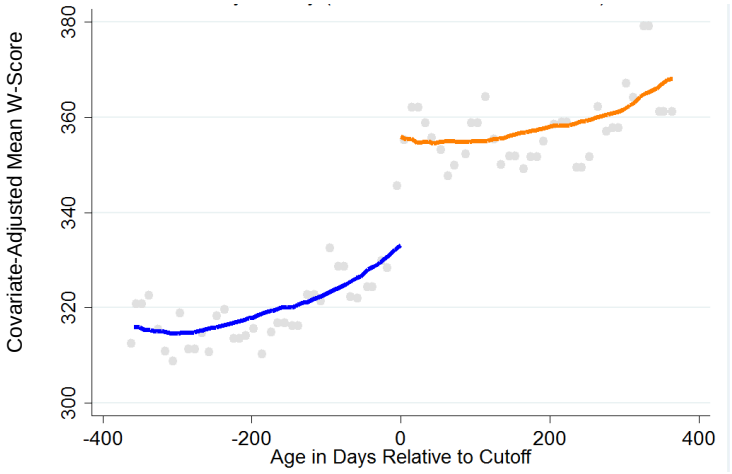


Exhibit A.6. Relationship of Age and Outcome: Early Math (WJ-III Applied Problems W-Score)

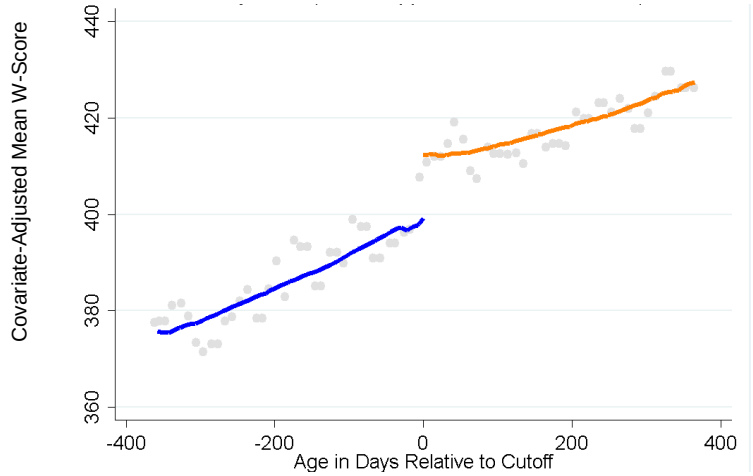


Exhibit A.7. Relationship of Age and Outcome: Vocabulary (PPVT Raw Score)

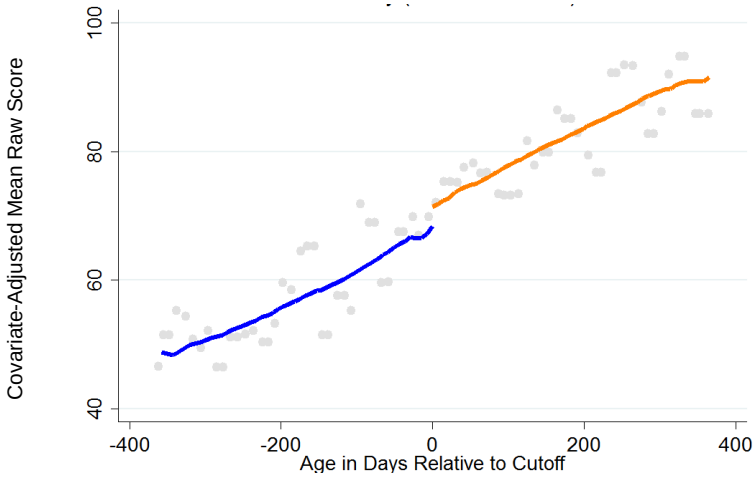
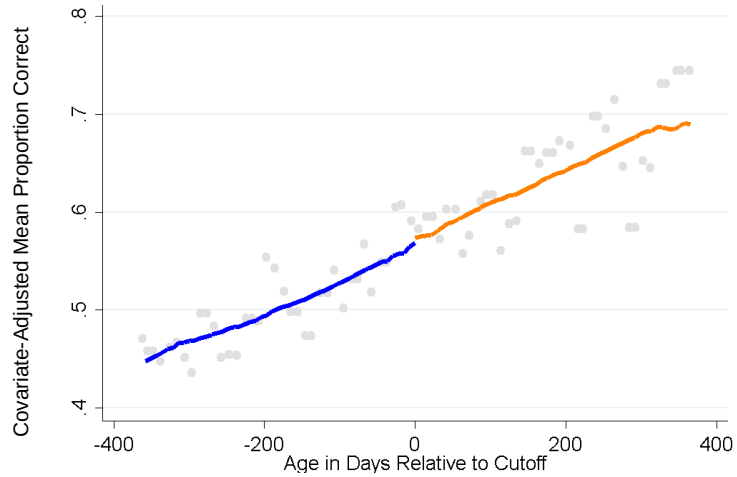


Exhibit A.8. Relationship of Age and Outcome: Executive Function (Hearts and Flowers Mixed Trials Raw Score)



Analysis of Participant Characteristics

The RD design requires exchangeability of students across the cutoff, or "as if random assignment" in the area around the cutoff. One way to examine this assumption is to look for evidence of breaks in the mean level of baseline covariates at the cutoff. Analyses, via local linear models, were conducted to examine gender (Exhibit A.9), language spoken at home (Exhibit A.10), and prior care in a formal setting (Exhibit A.11). The only characteristic that appears to show a large break at the cutoff is prior care, indicating that more of the control group right around the cutoff (i.e., the older children in that group) experienced prior care than the treatment group right around the cutoff (i.e., the younger children in the treatment group). This finding is not unexpected, since the otherwise identical cases on either side of the cutoff differ primarily in having an extra year of exposure to the risk of some formal care other than PEG prior to entering PEG. Testing for a statistically significant break in gender across bandwidths (Exhibit A.12) via local linear regressions with triangular kernels shows a precisely estimated zero difference in percent female at most bandwidths, and in home language (Exhibit A.13), a less precisely estimated difference that does not differ statistically from zero at any bandwidth. Testing for a statistically significant break in prior care across bandwidths (Exhibit A.14) shows positive differences at narrow bandwidths that do not differ from zero statistically, and negative differences at wider bandwidths that do differ statistically from zero at the largest bandwidths.

In summary, there was no systematic evidence of a jump in gender or home language at the cutoff. Further, there was very minimal evidence of a jump in prior care at the cutoff (only in some models but not in others). The majority of the time that prior care seemed somewhat differential by condition was in bandwidth-limited models where the sample size is smaller and the standard error is larger; therefore, it is impossible to parse out the effect of the covariate from the effect of the reduced sample.

Exhibit A.9. Probability of Being Female by Age

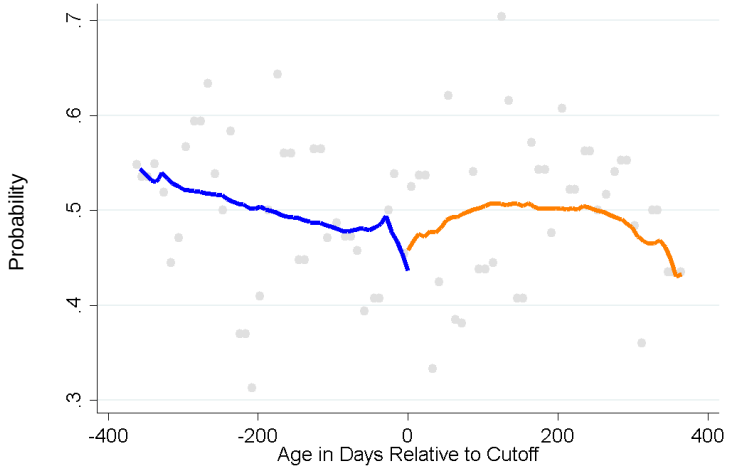


Exhibit A.10. Probability of Being from an English-Speaking Home by Age

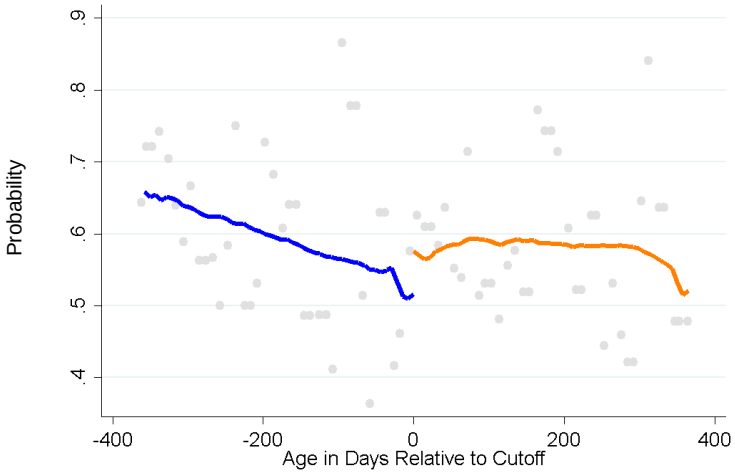


Exhibit A.11. Probability of Having Prior Childcare by Age

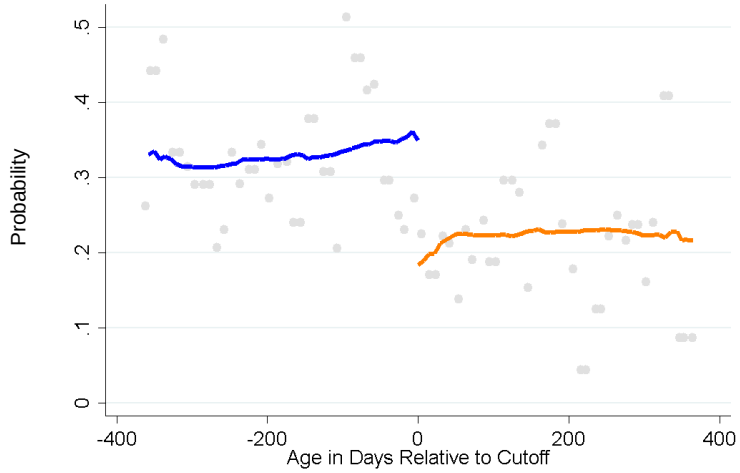
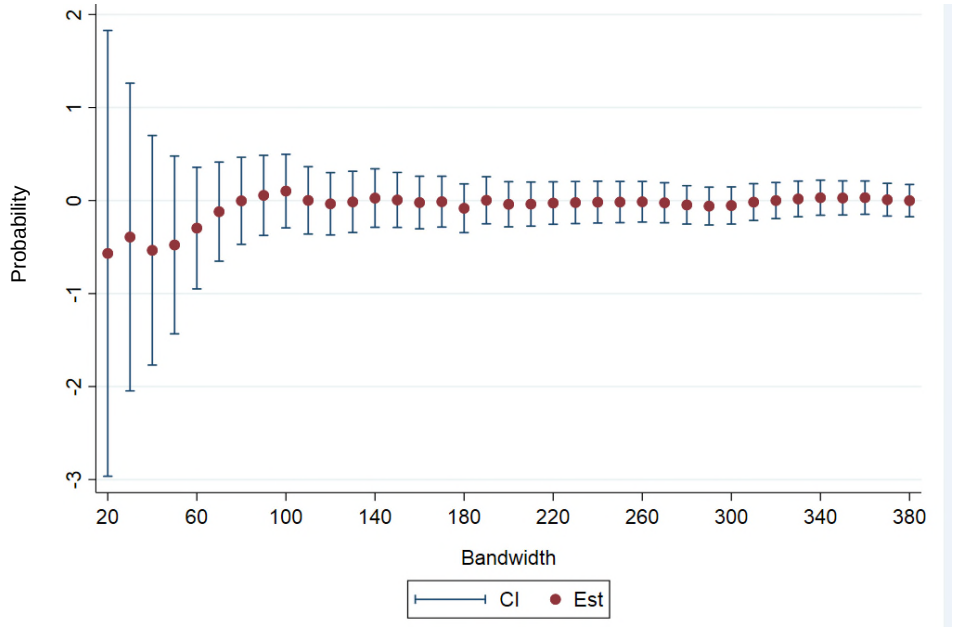
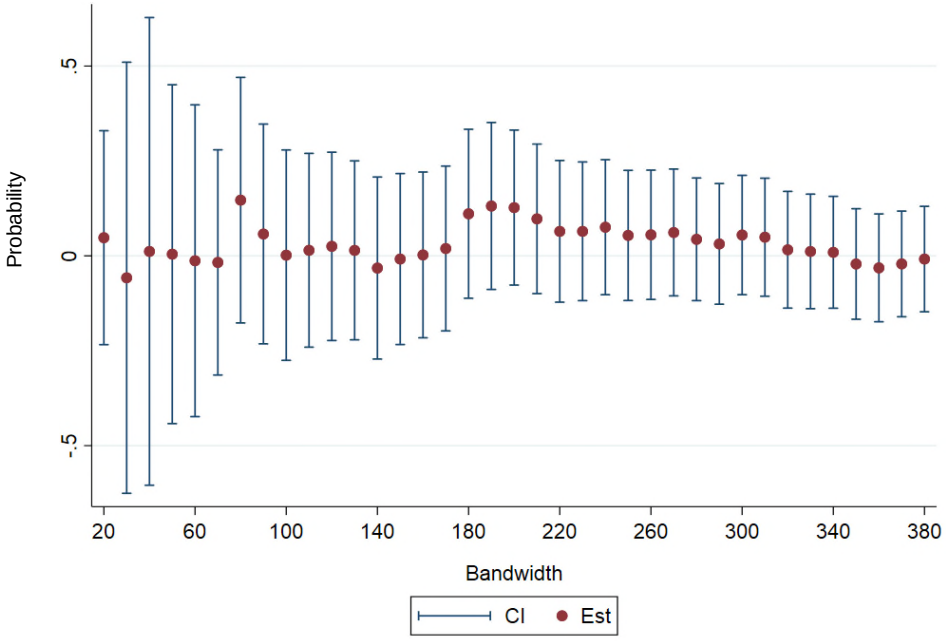


Exhibit A.12. Dependence on Bandwidth of the Differential Probability of Being Female at the Cutoff



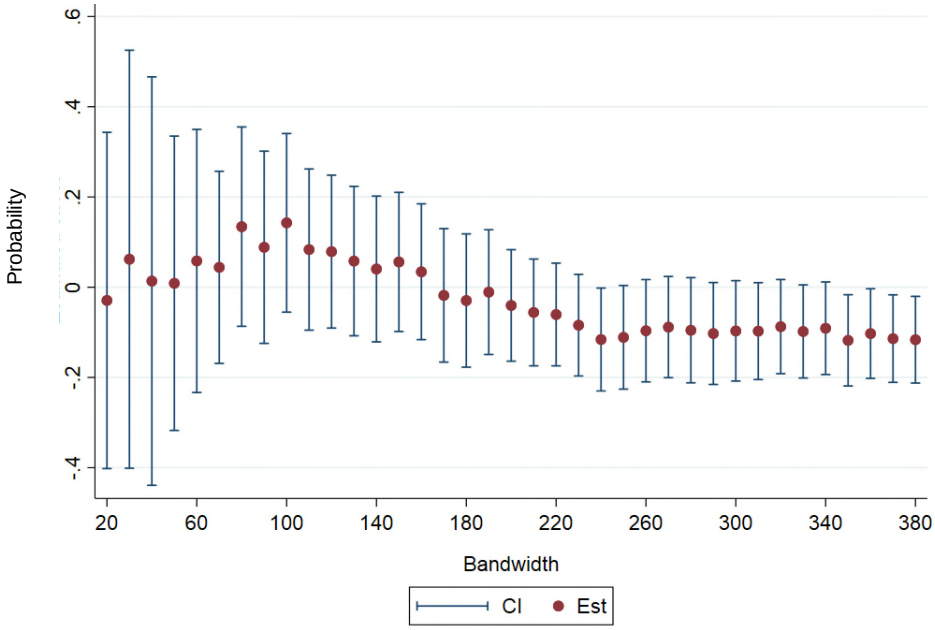
Note. 95% confidence interval

Exhibit A.13. Dependence on Bandwidth of the Differential Probability of Being from an English-Speaking Home at the Cutoff



Note. 95% confidence interval

Exhibit A.14. Dependence on Bandwidth of the Differential Probability of Having Prior Childcare at the Cutoff



Note. 95% confidence interval

Sensitivity Analyses

Analysis of Impact Variation Due to Sample Eligibility Requirements

Because compliance was imperfect (i.e., sometimes children who were too young to enroll in PEG ended up getting into the program and sometimes children who were too old to be in PEG ended up getting into the program, as well), it is important to know if impact estimates would be different when noncompliant cases are included (using a fuzzy RD design). The results from the main estimates using the analysis sample (i.e., those who meet both PEG eligibility requirements and analysis sample eligibility requirements) were compared to the equivalent regression model including all children who were assessed regardless of compliance with the age cutoff (but otherwise eligible). The second (fuzzy RD) design involves instrumenting for participation with eligibility based on age. Because only 5 of the 576 assessed treatment cases were the wrong age to be included in the analysis sample (0 in the control group), the differences between this instrumental variables (IV) model and the main results are negligible (see a comparison of the impact parameter estimates in Exhibit A.15). Taking into account the imperfect compliance using IV is to multiply the impact estimates in that slightly larger sample by 1.02 to 1.05 (dividing by first-stage compliance rates of .98 to .95) depending on bandwidth. But impact estimates are largely unaffected. By dropping the noncompliant cases, we improve precision (IV has higher asymptotic variance in every case, but in this type of exactly identified model, has a nonfinite mean and variance). The IV results in each case are qualitatively identical to the main analysis results. Thus, imperfect compliance is not substantively important.

Exhibit A.15. IV RDD Results: Includes Age-Ineligible Children and Instruments for Participation with Eligibility

Outcome	Global IV (Fuzzy) Parameter Estimate	Global RDD (Sharp) Parameter Estimate
Early Literacy	25.68***	24.54***
Early Math	12.51***	11.33***
Vocabulary	5.68*	4.93*
Executive Function (Mixed)	.02	.01

*p<.05, **p<.01, ***p<.001

Notes. Models are global regression models with linear functional form and included key child covariates as well as a set of dummy codes for classroom. Statistics are rounded to two decimal places.

Analysis of Impact Variation Due to Inclusion of Child Covariates

The effect on the impact estimate was analyzed when the models included child covariates as controls in the model. Exhibit A.16 shows that the parameter estimates did not change substantially with the inclusion of covariates, for each of the primary four outcomes.

Exhibit A.16. Variation in Treatment Parameter Estimate across Models with and without Child Covariate Controls

Outcome	Parameter Estimate without Child Covariates	Parameter Estimate with Child Covariates
Early Literacy	24.69***	24.54***
Early Math	11.64***	11.33***
Vocabulary	5.32*	4.93*
Executive Function (Mixed)	.02	.01

Notes. Models are global regression models with linear functional form and included key child covariates where indicated as well as a set of dummy codes for classroom. Statistics are rounded to two decimal places.

Analysis of Impact Variation Due to Functional Form

A critical piece in any RDD analysis is to correctly model the functional form of the relationship between child age (distance from the cut-off) and outcomes. The analyses examined whether impact estimates change substantially when the models are run using a quadratic rather than a linear functional form. Regression models were run using both quadratic and linear functional forms to facilitate comparisons across these specifications.

Exhibit A.17 below shows the parameter estimates and statistical significance of each of these model variations. Across all models, there is little difference in estimates regardless of whether a linear or quadratic functional form is used, with the exception of vocabulary. The quadratic models allow curvature but will give similar results as the linear model when the shape of the curve on each side of the cutoff is

the same. The linear and quadratic models for vocabulary differ because the shape differs: on one side of the cutoff, the slope is increasing everywhere, whereas on the other side, the slope is decreasing everywhere. Because the results from the two models differ for vocabulary, we have to exercise caution about either set of results.

When results from both functional form models are so similar, as is the case with three of the four key outcomes in this study, the linear model is the more parsimonious and therefore preferable in this design.

Analysis of Impact Variation Due to Bandwidth Size[17]

Analyses were conducted to examine whether impact estimates change substantially when the analysis sample is limited to those with birthdates falling in certain bandwidths rather than using the entire sample. To do so, local linear regression models were run with rectangular kernel shapes at different bandwidth sizes from 20 days to 380 days. These models included child-level covariates and fixed effects for classroom.

A local linear regression with a rectangular kernel simply restricts the regression to a range of ages, e.g., a 190 day bandwidth restricts the sample to those children who are 0 to 190 days older than the minimum age or 1 to 190 days younger than the minimum age (190 days on either side of the age cutoff). Exhibits A.18-A.21 summarize these estimates for the rectangular bandwidths for each outcome across a wide range of bandwidths. At larger bandwidths, impacts on early math, early literacy, and vocabulary (though impacts for the latter are not as robust across bandwidths as they decrease in size) are positive and significant (as confidence intervals do not overlap the axis), but impacts on executive function are rarely statistically distinguishable from zero across multiple bandwidths (confidence intervals overlap zero), and the very narrow confidence intervals around impact estimates for executive function rule out even modest impacts.

The following graphs indicate that both point estimates and confidence intervals are stable with bandwidths of 190 days or greater. At smaller bandwidths, confidence intervals are very large and point estimates are highly variable, where the bias is lower but variance dominates, so that a wide range of implausible true effects cannot be confidently rejected. At wider bandwidths, the models gain substantial reductions in variance at the cost of introducing more potential bias by including observations farther from the cutoff, and in each case, the models project to the cutoff to obtain inferences of treatment effects for a hypothetical child born at midnight on September 1, 2012.

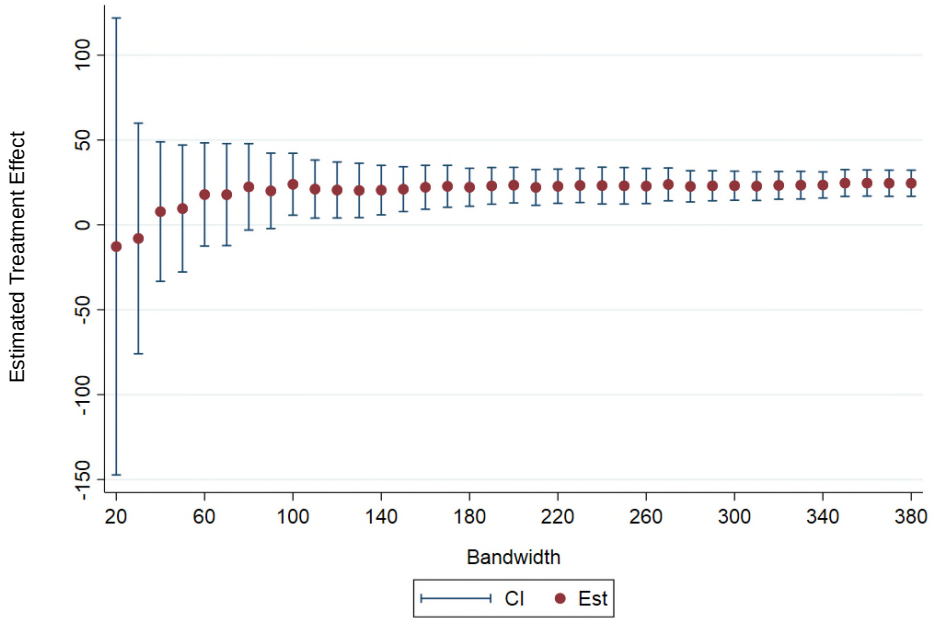
[17] Analyses that examined the robustness of model effects to variations in kernel size in the limited bandwidth models were also performed.

Exhibit A.17. Variation in Impacts Due to Functional Form

Parameter	Early Literacy		Early Math		Vocabulary		Executive Function (Mixed)	
	Linear Global Model	Quadratic Global Model	Linear Global Model	Quadratic Global Model	Linear Global Model	Quadratic Global Model	Linear Global Model	Quadratic Global Model
Treatment	**24.54*** (3.83)**	**22.46*** (6.11)**	**11.33*** (2.47)**	**9.96* (4.11)**	**4.93* (2.21)**	**.21 (3.24)**	**.01 (.03)**	**-.02 (.03)**
Age (Distance from Cut-off, Linear)	.05*** (.01)	.17** (.05)	.07*** (.01)	.14* (.05)	.07*** (.01)	.13*** (.04)	.00*** (.00)	.00*** (.00)
Treatment by Age Interaction	-.02 (.02)	-.24** (.08)	-.03* (.01)	-.14* (.07)	-.01 (.01)	-.08 (.05)	.00 (.00)	-.00 (.00)
Age (Distance from Cut-off, Quadratic)		.00* (.00)		.00 (.00)		.00 (.00)		.00* (.00)
Treatment x Age (Squared)		-.00 (.00)		-.00 (.00)		-.00 (.00)		-.00 (.00)
Female	1.87 (1.39)	2.02 (1.38)	4.65** (1.53)	4.72** (1.51)	4.86** (1.40)	4.84*** (1.37)	-.00 (.01)	-.00 (.01)
English as Home Language	3.73* (1.51)	3.53* (1.48)	9.98*** (2.03)	9.87*** (2.05)	16.52*** (1.66)	16.43*** (1.69)	.01 (.01)	.01 (.01)
Prior Childcare Exposure	7.18** (2.63)	7.39** (2.55)	5.64* (2.51)	5.75* (2.58)	5.14* (2.50)	5.16* (2.56)	-.01 (.02)	-.01 (.02)
Constant	315.00*** (2.06)	322.40*** (3.83)	386.00*** (2.30)	389.90*** (3.68)	48.21*** (1.76)	52.77*** (2.82)	.60*** (.02)	.63*** (.02)

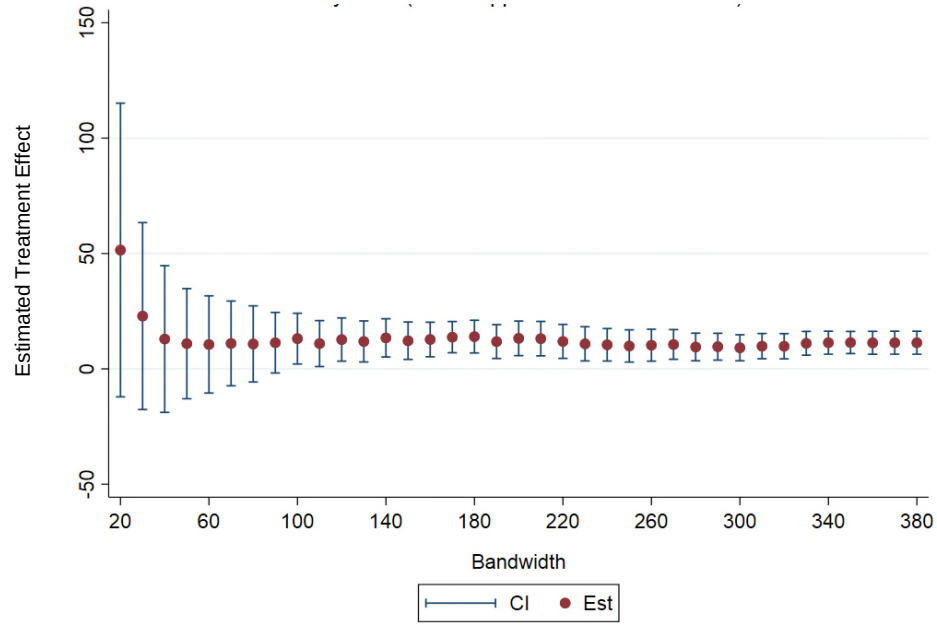
Notes. Models were global regression models with functional form as indicated and also included a set of dummy codes for classroom. Statistics are rounded to two decimal places.

Exhibit A.18. Variation in PEG Impact on Early Literacy by Bandwidth



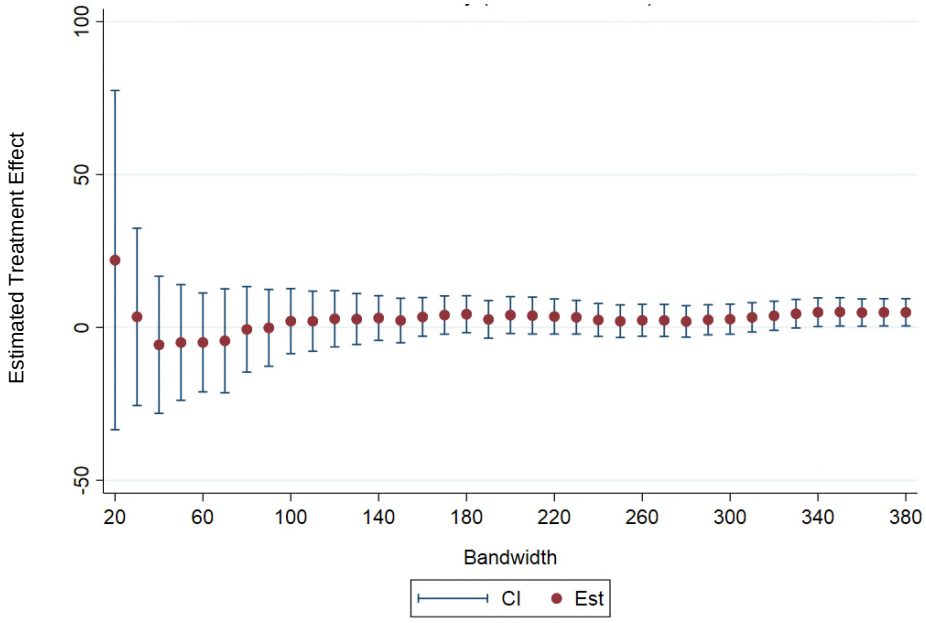
Note. 95% confidence interval

Exhibit A.19. Variation in PEG Impact on Early Math by Bandwidth



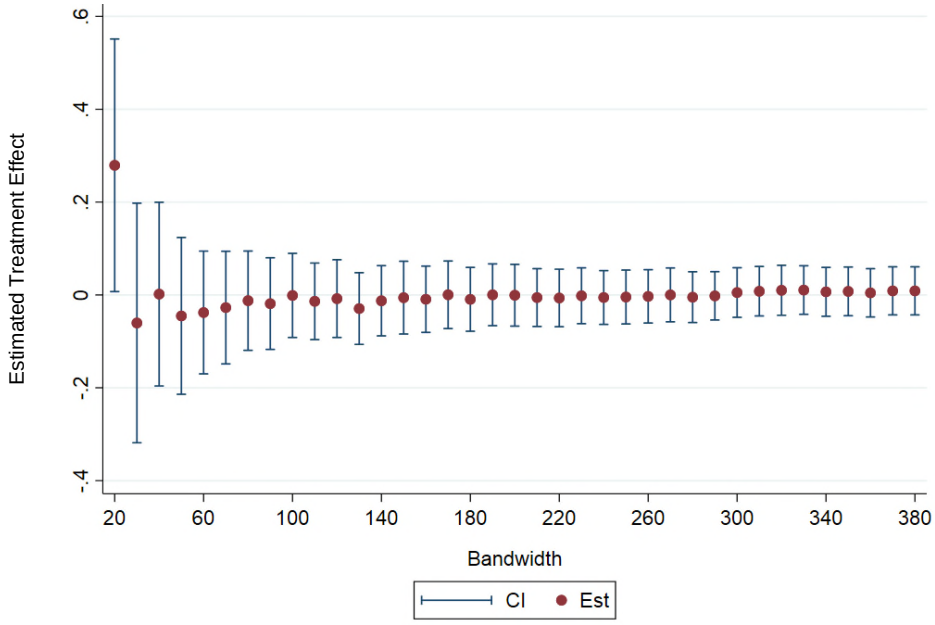
Note. 95% confidence interval

Exhibit A.20. Variation in PEG Impact on Vocabulary by Bandwidth



Note. 95% confidence interval

Exhibit A.21. Variation in PEG Impact on Executive Function (Mixed Trials) by Bandwidth



Note. 95% confidence interval

Regression Results for Child Subgroups

Exhibits in this section show the results of global linear and quadratic regressions estimated for various subsets of the sample, or where models include interactions of treatment with subgroup indicators (which is equivalent to estimating models in each subgroup and then combining the results to test for differences in treatment impact across subgroups). Because there are many coefficients being tested in these models, and no correction is made for multiple hypothesis testing, the reader is cautioned to interpret results with care.

Analyses were conducted to examine the interaction of treatment and child covariate (gender, home language, and prior care) in separate models, providing global and limited-bandwidth model (rectangular kernel with 190-day bandwidth) results (see Exhibits A.22-A.24).

Though there is not sufficient power to detect whether these patterns are due to chance or systematic variation, the most robust suggestive pattern is that treatment effects tend to be smaller for children with prior care than for children who have not had prior exposure to formal childcare.

Exhibit A.22. Impacts Results for Child Subgroup: Females

Parameter	Early Literacy		Early Math		Vocabulary	
	Linear Global Model	Limited Bandwidth Model (190 days)	Linear Global Model	Limited Bandwidth Model (190 days)	Linear Global Model	Limited Bandwidth Model (190 days)
Treatment	27.57***	27.60**	14.80***	14.31*	5.92	4.85
Age (Distance from Cut-off, Linear)	.04**	.07	.07***	.10	.06***	.12*
Treatment by Age Interaction	-.01	-.08	-.04	-.10	-.02	-.11*
Gender	5.97	12.06	7.15	5.37	4.35	1.08
Female by Treatment Interaction	**-6.31**	**-9.75**	**-7.30**	**-5.37**	**-2.12**	**6.63**
Female by Age Interaction	.01	.11	.00	.00	-.00	-.03
Female by Treatment by Age Interaction	-.01	-.12	.01	.02	.02	.12
English as Home Language	3.72*	2.00	10.03***	7.72**	16.58***	16.24***
Prior Childcare Exposure	6.97*	14.08***	5.32*	12.27***	5.04	12.11**
Constant	312.80***	322.10***	384.6***	388.80***	48.31***	51.40***

*p<.05, **p<.01, ***p<.001

Notes. Models also included a set of dummy codes for classroom. Limited bandwidth models used rectangular kernels and linear functional form. Statistics are rounded to two decimal places.

Exhibit A.23. Impacts Results for Child Subgroup: Children from English-Speaking Homes

Parameter	Early Literacy		Early Math		Vocabulary	
	Linear Global Model	Limited Bandwidth Model (190 days)	Linear Global Model	Limited Bandwidth Model (190 days)	Linear Global Model	Limited Bandwidth Model (190 days)
Treatment	31.73***	28.71**	19.26***	18.37*	9.86*	5.43
Age (Distance from Cut-off, Linear)	.04*	.12	.08***	.06	.06**	.10
Treatment by Age Interaction	-.03	-.13	-.04	.00	-.01	-.05
English as Home language	9.02*	9.50	15.43**	21.83*	19.84***	18.82*
English as Home language by Treatment Interaction	**-12.76**	**-11.71**	**-14.19****	**-13.95**	**-8.65**	**-5.49**
English as Home language by Age Interaction	0.01	.01	-.00	.09	.02	-.00
English as Home language by Treatment by Age Interaction	0.01	-.03	.02	-.16	.01	-.00
Female	1.93	1.47	4.76**	3.65*	4.87**	4.37*
Prior Childcare Exposure	6.85*	13.11**	5.28*	11.73***	4.91	11.86**
Constant	312.30***	324.20***	383.80***	383.10***	46.10***	49.03***

*p<.05, **p<.01, ***p<.001

Notes. Models also included a set of dummy codes for classroom. Limited bandwidth models used rectangular kernels and linear functional form. Statistics are rounded to two decimal places.

Exhibit A.24. Impacts Results for Child Subgroup: Children with Prior Childcare

Parameter	Early Literacy		Early Math		Vocabulary	
	Linear Global Model	Limited Bandwidth Model (190 days)	Linear Global Model	Limited Bandwidth Model (190 days)	Linear Global Model	Limited Bandwidth Model (190 days)
Treatment	29.34***	25.36***	17.19***	17.62***	9.49**	4.77
Age (Distance from Cut-off, Linear)	.04***	.12**	.05***	.07	.05***	.09*
Treatment by Age Interaction	-.02	-.14*	-.01	-.05	.01	-.03
Prior Childcare Exposure	13.30**	17.88*	20.01***	27.44***	16.20***	17.71*
Prior Childcare Exposure by Treatment Interaction	**-18.40***	**-9.13**	**-18.88*****	**-20.13****	**-14.73***	**-7.69**
Prior Childcare Exposure by Age Interaction	.01	-.01	.06*	.11	.05**	.02
Prior Childcare Exposure by Treatment by Age Interaction	.02	-.01	-.07*	-.12	-.05	-.05
English as Home Language	3.59*	7.05	9.76***	7.14**	16.34***	15.69***
Female	1.66	1.09	4.58***	3.47*	4.80***	4.16*
Constant	312.90***	326.30***	381.00***	385.3***	44.34***	48.65***

*p<.05, **p<.01, ***p<.001

Notes. Models also included a set of dummy codes for classroom. Limited bandwidth models used rectangular kernels and linear functional form. Statistics are rounded to two decimal places.

Regression Results for Community Subgroups

To investigate the extent to which the impact of PEG was consistent across the five communities, analyses were conducted to examine the interaction of community and treatment for each key outcome. The results of these models are shown in Exhibit A.23.

To examine how the impact of PEG differed across the five PEG communities, the following analyses were conducted: (a) separate regression models for each of the five PEG communities, and (b) regression models that included terms for the interactions of treatment and community that alternated the community reference group. Sensitivity analysis of the community differences were also performed.

The first row of Exhibit A.25 shows the statistical significance of the overall F-test, which tested for overall differences by community. The following rows of Exhibit A.25 show which community impacts differed significantly from which other community impacts. By and large, the impact of PEG across the five communities was similar, despite the freedom afforded them to develop their own PEG implementation model. There were significant overall differences by community for each early academic outcome in the global models, meaning that on each outcome, at least one of the five communities had a significantly different impact than one or more of the other communities. However, the only difference that was large enough to hold up in the limited bandwidth model was the community difference related to the impact on

early math skills. In particular, the average math score in one community ('Community 1' in the Exhibit) was significantly different that the average impact in two other communities. Three of the communities showed no significant differences from each other in all models.

Exhibit A.25. Global and Local 190-Day Bandwidth Regression Results (significance) for Community Comparisons

	Global Model			Limited Bandwidth Model		
Comparison	Early Literacy	Early math	Vocabulary	Early Literacy	Early math	Vocabulary
Overall Test of Community Differences	**	***	**	n.s.	*	n.s.
Individual Community Comparisons						
Community 1 v Community 2	n.s.	n.s.	n.s.	n.s.	**	n.s.
Community 1 v Community 3	*	***	***	n.s.	n.s.	n.s.
Community 1 v Community 4	n.s.	n.s.	*	n.s.	n.s.	n.s.
Community 1 v Community 5	n.s.	**	**	n.s.	**	n.s.
Community 2 v Community 3	***	*	n.s.	n.s.	n.s.	n.s.
Community 2 v Community 4	*	n.s.	n.s.	n.s.	n.s.	n.s.
Community 2 v Community 5	**	n.s.	n.s.	n.s.	n.s.	n.s.
Community 3 v Community 4	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Community 3 v Community 5	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Community 4 v Community 5	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

*p<.05, **p<.01, ***p<.001

Notes. Models included child covariates and a set of dummy codes for classroom. Global regressions had linear functional forms. Limited bandwidth models used rectangular kernels and linear functional form. Statistics are rounded to two decimal places.