Impacts of Participation in Preschool Expansion in the Commonwealth of Virginia (VPI+) on Children’s Early Academic Skills

Association for Public Policy and Management Annual Conference Fall 2018

Todd Grindal, Erika Gaylor, Kirby A. Chow, and Shari Golan, SRI International

Abstract

The Commonwealth of Virginia received a federal Preschool Development Grant to increase access and improve quality of a state-funded preschool program for four-year olds and their families in high need communities, launching Virginia Preschool Initiative Plus (VPI+) in 11 of Virginia’s 132 school divisions. All VPI+ classrooms are expected to include components consistent with a high-quality preschool program (e.g., highly educated teachers with child development expertise, use of an evidence-based curriculum, ongoing training and technical assistance). Evidence from Boston (Weiland & Yoshikawa, 2013) and Tulsa (Gormley, 2005) indicates that the provision of high-quality preschool across a large school district can yield positive impacts for children, but it remains unclear whether statewide efforts can replicate these results (Farran, 2016). This paper examines the impact of participation in VPI+ on children’s early academic skills at kindergarten entry by using a regression discontinuity design (RD).

Methods. To be eligible to participate in VPI+, a child’s family income had to be 200% or less of the federal poverty level (FPL), and the child had to be between the ages of four and five-years-old by September 30 of the preschool year. We leverage this strict age cutoff and compare children at kindergarten entry who just completed the VPI+ preschool intervention during the 2016–2017 school year (VPI+ treatment, n = 969) to children who just began VPI+ preschool during the 2017–2018 school year (control, n = 1,140). The impact of VPI+ is then operationalized as the difference in the expected achievement levels for a child that was just old enough to be eligible for VPI+ in fall of 2016 and a child who was just too young to be eligible for VPI+ in fall of 2016 and subsequently enrolled in the VPI+ program in fall 2017. Child outcome measures included the Phonological Awareness Literacy Screening (PALS) lowercase letter recognition and letter sounds tasks (literacy skills); Woodcock Johnson® III Applied Problems subtest (mathematics); Peabody Picture Vocabulary Test (receptive vocabulary); and the Head Toes Knees Shoulders task (self-regulation). Covariates included child race/ethnicity, child sex, child disability status, child English language learner status, maternal education, and family income at program entry.

Findings. Preliminary analyses indicate that RD assumptions were satisfied (compliance with the age cutoff, no evidence of “pile up,” and no discontinuities of baseline covariates at the age cutoff) (Lipsey et al., 2015). All analyses were conducted using a global regression as well as empirically derived optimal bandwidth approaches and included covariates and propensity score weights to account for baseline differences. Preliminary results show large impacts on literacy skills [effect size (ES) > 1.0] and moderate effects on self-regulation (ES = 0.4) across all analytic models. We observe small impacts (ES = 0.2) on mathematics skills in some models and do not observe statistically significant impacts on vocabulary.

Implications. For policy makers, the findings suggest the importance of strengthening some aspects of instructional quality, including mathematics and language supports. The presentation will discuss how these data are being used to inform program improvement and progress monitoring.
Impacts of Participation in Preschool Expansion in the Commonwealth of Virginia (VPI+) on Children’s Early Academic Skills

Todd Grindal, Erika Gaylor, Kirby A. Chow, and Shari Golan, SRI International

Paper presented at Association for Public Policy and Management Annual Conference Fall 2018, Washington DC.

Note that findings presented in this paper are preliminary and may change following the integration of additional data. Final findings expected in Spring 2019. Please consult the first author before citing. Todd.Grindal@SRI.com
Enrollment in state-funded preschool programs has become common among four-year-old children in the United States. In 2017, more than 1.3 million or approximately one third of four-year-old children attended a state-funded preschool program (Friedman-Krauss et. al., 2017). The evidence on the impacts of participation in these programs on children’s development is mixed. Local programs in Boston, MA (Weiland & Yoshikawa, 2013) and Tulsa, OK (Gormley, 2005) indicate that the provision of high-quality preschool across a large school district can yield large short-term positive impacts on the development of four-year old children. A recent quasi-experimental examination of eight state-funded preschool programs indicated that these programs, on average, yield positive results on a range of early academic skills (Barnett et al., 2018). Yet, experimental evidence from Tennessee calls into question whether these positive early results persist as children progress into elementary school (Lipsey, Farran, & Durkin, 2018).

In 2014, the US federal government awarded $226 million of grants to eighteen states to develop and/or expand high-quality early childhood education for four-year-old children from low-income families within a mixed delivery system of care. To receive one of these Preschool Development Grants (PDG) or Preschool Expansion Grants (PEG), states were expected to implement components regarded as key ingredients of quality programs. These include but are not limited to: high staff qualifications, low child-staff ratios and small class sizes, a full-day program, salary parity for lead teachers, and comprehensive services for children.

The Virginia Department of Education (VDOE) received a Preschool Expansion Grant in January 2015. With this 4-year, $17.5 million per year grant, VDOE launched an initiative to provide access to high-quality preschool programs for four-year-old children from low-income families in selected high need communities. Since the PDG grant augmented Virginia’s existing state-funded Virginia Preschool Initiative (VPI), Virginia named the work being carried out through its PDG grant the Virginia Preschool Initiative Plus (VPI+). VDOE has used these PDG funds to support two types of preschool classrooms in high-need communities in participating school divisions across the state: (1) VPI+ classrooms: newly-opened VPI+ classrooms that implement all of the VPI+ grant requirements; and (2) VPI Improved classrooms: existing state-funded VPI classrooms that enhance their quality by implementing at least one of five program quality enhancements. Due to evaluation budgetary constraints, VDOE decided to focus the external evaluation on only VPI+ classrooms in the 11 participating school divisions¹, given these VPI+ classrooms receive the full treatment of initiative supports.

VPI+ went beyond the requirements of the PDG grant and included the following features: teachers with an active Virginia teaching license with an elementary endorsement including Pre-K; child-to-instructional staff ratios of no more than 9 to 1 and class sizes of no more than 18 children; a formative assessment system and training; evaluation and monitoring from the Virginia Quality Rating and Improvement System (QRIS); family engagement coordinators to help with outreach to hard-to-reach families and to connect families to services; and significant additional resources (e.g., instructional technology for classrooms, classroom libraries). Each school division hired at least one instructional coach to provide VPI+ teachers with job-embedded professional development and personalized support on improving their instructional practice. Coaches were tasked to help teachers implement evidence-based curricula to target learning in the five Essential Domains of School Readiness (language and literacy, early mathematics and early scientific development, approaches to learning, physical well-being

¹ Virginia uses the term “division” rather than “district” to refer to public local education agencies.
and motor development, and social and emotional development); engage in effective teacher-child interactions, and individualize instruction based on formative assessments.

During its first three years of implementation, VPI+ has provided center-based preschool services to 4,022 four-year-old children across 11 school divisions. To be eligible to participate in VPI+, a child’s family income had to be less than 200 percent of the federal poverty level (FPL) and the child had to be between the ages of four and five-years-old by September 30th of the preschool year. Unlike some other state initiatives funded by the PDG/PEG (i.e., Massachusetts), prior participation in center-based preschool was not considered in determining children’s eligibility to participate in VPI+.

VPI+ programs are locally administered but governed by a set of core requirements set by VDOE and the federal grant. All VPI+ programs are full day, providing 5.5 hours or more of instructional time per day and operate on an elementary school calendar with approximately 180 days of instruction provided between September and June (but this number varied across school divisions due to differences in school calendars). The VPI+ school divisions used one of four different comprehensive curricula. VPI+ teachers in 8 of the 11 divisions used The Creative Curriculum®. Among the other three divisions, one used Houghton Mifflin Harcourt, one used High Scope curriculum, and one used a locally developed curriculum that was vetted by the Center for Advanced Study of Teaching and Learning (CASTL) at the University of Virginia (UVA), which supported VPI+ implementation and continuous quality improvement efforts.

Divisions also varied in which coaching model they used. Coaches from all but one division reported using the practice-based coaching framework. Most divisions that used practice-based coaching also used a second coaching model or program such as MyTeaching Partner ($n = 3$ divisions), the Pyramid Model ($n = 2$ divisions), PATHS ($n = 1$ division), Erikson math ($n = 2$ divisions), or Teachstone ($n = 1$ division). VPI+ teachers, on average, met with coaches for three hours two times per month.

In this draft conference paper, we examine the short-term impacts of participating in VPI+ on the vocabulary, mathematics, self-regulation, and literacy skills of children who enrolled in VPI+ in the fall of 2016 and subsequently entered kindergarten in public school in the same school division in the fall of 2017.

Research Design

We use a regression discontinuity design (RD) to assess the impact of VPI+ participation on children’s skills at kindergarten entry. Following the approach used for estimating the impact of preschool programs in numerous previous studies of preschool programs (see for example, Gormley et al., 2005; Hofer et al., 2018; Weiland and Yoshikawa, 2013), we also use age-related program requirements as a forcing variable for conducting the RD analysis. We considered this birthdate age cutoff as an exogenous factor that determines whether otherwise interested and eligible children enroll in the program in a given year. We leverage this strict age cutoff and compare children at kindergarten entry who just completed the VPI+ preschool intervention during the 2016–2017 school year (VPI+ treatment) to children who just began VPI+ preschool during the 2017–2018 school year (control). The impact of VPI+ is then operationalized as the difference in the expected achievement levels for a child that was just old enough to be eligible for VPI+ in fall of 2016 and a child who was just too young to be eligible for VPI+ in fall of 2016 and subsequently enrolled in the VPI+ program in fall 2017.
Data Sources

Data for this study were collected as part of the SRI Education evaluation of the VPI+ program implementation and impacts. Staff from School Readiness Consulting (SRC), who served as a subcontractor to SRI on the evaluation, helped collect these data. These independent assessors from SRC administered three direct assessments (vocabulary, mathematics, and self-regulation) and children’s teachers administered a literacy assessment. If children spoke a language other than English at home (based on parent report), they were administered a language screener at the beginning of the assessment to see whether they could complete the direct assessments for mathematics and self-regulation in English. For children who did not pass the screener and whose home language was Spanish, parallel Spanish measures were administered. Children who did not pass the language screener and spoke a language other than Spanish were not administered these two assessments. All children, regardless of home language, completed the vocabulary assessment in English. The assessors received training on conducting the vocabulary, mathematics, and self-regulation assessments prior to each assessment cycle and were required to achieve 90% reliability prior to data collection. Assessment supervisors made unannounced visits to study sites to observe assessors, ensuring that the SRC assessors maintained high reliability with a master assessor and followed all protocols. These assessors collected data on the treatment and control children between September 18, 2017 and November 27, 2017, as treatment students began Kindergarten and control students began VPI+ Pre-K.

Population and Analytic Sample

VPI+ provided funds for the four consecutive school years from 2015-2016 through 2018-2019. We refer to each group of children who enrolled in VPI+ as a cohort. Children who participated in VPI+ during the 2016–2017 school year were part of VPI+ Cohort 2 and represent the treatment group for this study. Children who participated in VPI+ during the 2017–2018 school year were part of VPI+ Cohort 3 and represent the control group for this study. SRI researchers, in collaboration with VDOE program leaders, decided not to assess the impact of VPI+ on the first cohort of VPI+ participants because some divisions were not able to fully implement their programs by the fall of Year 1 due to delays from initial program set up, recruitment, and training. This decision to not evaluate program impacts on Cohort 1 children was made before the collection of any child data.

VPI+ program features (i.e., hours of operation, teacher qualifications, curriculum) as well as eligibility requirements for children, were consistent across Cohorts 2 and 3. The procedures for recruiting families to participate in VPI+ were all consistent across the two years, but VPI+ staff report there were stronger efforts to enroll children to participate in Cohort 3 who were dual language learners (DLL), children with disabilities, and children with incomes at or below the federal poverty level.

We limited the sample of Cohort 2 and Cohort 3 children included in these analyses to create groups of children who are, to the greatest extent possible, equal on the observed and unobserved factors that

---

2 By parent report, approximately 24–30% of participating VPI+ children were identified as speaking a language other than English at home. Those children who spoke a language other than English at home took an English language screener to see if they could understand and express themselves in English well enough to complete assessments in English. Of those children, 55% passed the screener in fall 2017. See Appendix G for copy of the screening protocol.
might be associated with their scores on the outcome measures. The procedures for our sample reduction are displayed graphically in Appendix A and described below.

**Cohort 2 (treatment group) analytic sample**

VDOE administrative records indicate that 1,295 children participated in Cohort 2 of VPI+ at some point during the 2016–2017 school year. We distinguish between two types of children that we retain for the analytic sample. The first category of children, who we classify as Consistently enrolled and assessed,³ (n = 926) are those children that were enrolled in VPI+ Pre-K in fall 2016 and spring 2017, enrolled in Kindergarten in the same division in fall 2017, and had valid assessment data from the fall of 2017 on one or more outcome measures. We also include in the analysis a second group of children who we classify as Pre-K non-completers with outcome data (n = 44). These children were enrolled in VPI+ Pre-K in fall 2016, were not enrolled in VPI+ Pre-K in spring 2017, yet were enrolled in Kindergarten in the same school division where they had originally been enrolled in VPI+ and had valid assessment data in the fall of 2017 on one or more outcome measures.⁴

We distinguish between four types of Cohort 2 children who are not included in the analytic sample, Pre-K attritors (n = 68) are children who were enrolled in VPI+ Pre-K in fall 2016, who departed before spring 2017, and not enrolled in Kindergarten within the same division in fall 2017. Kindergarten no shows (n = 139) are children who were consistently enrolled in VPI+ Pre-K in fall 2016 and spring 2017 but not enrolled in Kindergarten within the same division in fall 2017. Kindergarten not tested (n = 26) are children who were enrolled in VPI+ Pre-K in fall 2016 and enrolled in Kindergarten within the same division in fall 2017, but did not have data on at least one Kindergarten outcome measure at fall 2017. Late joiners (n = 93) are children who were enrolled in VPI+ Pre-K after fall 2016.

Children classified as Pre-K attritors, Kindergarten no-shows, and Kindergarten not tested are not included in the sample because they are missing outcome data. We chose to also remove the late joiners to reduce potential unobserved differences between the Cohort 2 and Cohort 3 samples. Late joiners represent a concern for cross-cohort comparability because the collection of outcome data for this study occurred as the Cohort 3 children entered Pre-K. This means that the Cohort 3 sample will not include any of the children who might later be Cohort 3 Late joiners.

**Cohort 3 (control group) analytic sample**

As of December 2017, 1,163 children had participated in Cohort 3 of VPI+ Pre-K during the 2017–2018 school year. The Cohort 3 analytic sample includes children consistently enrolled and assessed, so children who were enrolled in VPI+ Pre-K in the fall 2017 and have valid data on at least one of the measures of student skills in fall 2017 (n = 1,121) as well as Pre-K non-completers (n = 34) who were children enrolled in VPI+ Pre-K in fall 2017, who had valid data on at least one focal outcome, but who left the VPI+ Pre-K program by December 2017. The only Cohort 3 children who are not included in the analyses are the Pre-K not tested, who were missing fall 2017 outcome data (n = 19).

³ It is possible that a child could be enrolled in fall 2016 and spring 2017 but not enrolled at some point in the middle. The evaluation team obtained rosters at three points across the year: September, December, and March to determine these enrollment patterns.

⁴ One of these 44 children who are Pre-K non-completers with outcome data does not have data on at least one outcome measure at fall 2017 and is thus also included in the Kindergarten not tested category.
Measures

The analyses use enrollment data on children’s demographic characteristics, as well as data from measures of children’s academic skills administered by teachers and independent assessors.

Student age. We created an indicator of each child’s age in days for use as the independent variable for these RD analyses. This variable was created with information using the children’s birthdate to determine the number of days the child had been alive as of September 30, 2016, the cut-off for eligibility to participate in Cohort 2 of VPI+. We then center each child’s age at the cutoff so that a child born on September 30, 2016 is given a value of 0. Children born before September 30, 2016, and thus eligible for VPI+ Cohort 2, have a positive value for the number of days from the cutoff on the age variable. Conversely, children born after September 30, 2016, and thus not eligible for VPI+ Cohort 2, have a negative value for the number of days from the cutoff.

Student demographics. Information on student demographics (e.g., gender, race/ethnicity, home language, family poverty level, mother’s education), was derived from administrative records. In each year, participating VPI+ school divisions provided data on VPI+ enrollment and student demographics and teacher and program characteristics in the fall, and updates to student enrollment in the spring.

Student academic skills. Information on children’s early academic skills were measured across four domains of school readiness with tools that have been extensively validated and used with children 4 to 6 years of age, living in different geographic regions, from different socioeconomic backgrounds, and from different racial/ethnic groups.

Literacy skills: Information on children’s early literacy skills was collected using the Phonological Awareness Literacy Screening (PALS)-PreK. Pre-K teachers administered eight PALS assessments (e.g., upper-case alphabet recognition, lower-case alphabet recognition, letter sounds). Then, in kindergarten, children’s teachers administered the PALS-K, which screens for more advanced reading and writing skills across seven tasks (e.g., spelling, concept of word). A prior validation study showed that interrater reliability estimates for PALS-PreK and PALS-K, expressed as Pearson coefficients, were .96 to .99 across tasks (Invernizzi et al., 2004), and internal consistency assessed using Cronbach’s alpha ranged from .75 to .93 for both versions of the PALS. In addition, PALS-PreK scores are predictive of first and second grade reading levels and fall PALS-K scores are predictive of PALS scores through third grade (Invernizzi et al., 2004). For these analyses we use only the scores from the PALS letter sounds and lower-case alphabet recognition tasks because these are the only two tasks that are identical across the PALS Pre-K and PALS-K assessments.

Mathematics skills: The Woodcock-Johnson Applied Problems subtest (Woodcock et al., 2001) was used to measure the mathematics skills of students who passed the English language screener, and the Batería III Woodcock-Muñoz Problemas Aplicados (Muñoz-Sandoval et al., 2009) was used for Spanish-speakers who did not pass the language screener. The subtest is a widely used norm-referenced measure of early mathematics skills (e.g., counting, number sense). When reporting descriptive information, we use the standard scores of the Woodcock-Johnson, which has a mean of 100 and standard deviation of 15. We use WJ W scores for the regression discontinuity analyses. Reliability coefficients for children ages 2 to 6 range from .88 to .94 for this subtest (Woodcock et al., 2001), and prior studies have shown that internal consistency coefficients for this tool range from .92 to .94 (Burchinal et al., 2016). For preschool-
age children, this measure shows evidence of concurrent validity, with Applied Problems scores moderately to highly correlated with 12 early numeracy assessments that measure skills and concepts identified as fundamental to mathematics development (rs = .35-.70; Purpura & Lonigan, 2015).

**Self-regulation:** The Head Toes Knees Shoulders task (HTKS; Ponitz et al., 2008, 2009) was used to measure children’s self-regulation. The task was administered by trained assessors and required children to play a game in which they had to do the opposite of what the assessor asked; it took approximately 10 minutes to administer. Possible scores range from 0 to 60. This measure does not have norm references, but research on the HTKS provides strong evidence for its validity, including substantial correlation with children’s achievement level and gains during preschool and kindergarten, with rs in the .18 to .49 range (McClelland et al., 2007; Ponitz et al., 2009). Moreover, the HTKS is moderately to strongly correlated with other direct assessments of executive function, with rs ranging from .32 to .77 (Allan & Lonigan, 2011; Carlson & Harrod, 2013). Finally, the HTKS task is moderately correlated with teacher ratings of children’s self-regulation in preschool and kindergarten (rs ranging from .15 to .35; Ponitz et al., 2008, 2009; Schmitt, Pratt, & McClelland, 2014).

**Vocabulary:** The Peabody Picture Vocabulary Test (PPVT-4) was used to measure children’s receptive vocabulary. An assessor shows the child an easel with four black-and-white drawings on each page, and the child is asked to indicate the picture that corresponds to the word spoken by the assessor. The assessment ends when a child answers eight or more questions incorrectly in a set. There are 12 questions in a set, and it takes about 20-25 minutes to administer. The PPVT-4 is a norm-referenced assessment with high published reliabilities (α = .89 to .97; West et al., 2010). We use standard scores which has a mean of 100 and standard deviation of 15 when reporting descriptive statistics and raw scores for the regression discontinuity analyses.

**Analyses**

The first step in our analyses was to determine whether the hypothesized forcing variable (age in days) distinguished between the two cohorts of children as expected. Program eligibility guidelines stipulate that children must be no younger than four years old and no older than five years old by September 30 of the preschool year. All of the children in Cohort 3 (control) analysis sample were age-eligible to participate in VPI+ Pre-K in fall 2017. Among children in the Cohort 2 (treatment) analysis sample, 1 out of 1,140 children did not appear to have been age eligible to participate in VPI+ Pre-K in fall 2016 (Exhibit 1). This child was too old to participate in VPI+ Pre-K in 2016 by 80 days. Program officials were unable to explain this child’s enrollment in VPI+ Cohort 2. We dropped this child from subsequent analyses.

---

5 There is a Spanish version of the PPVT called the Test de Vocabulario en Imágenes Peabody (TVIP) (Dunn, Lugo, Padilla, & Dunn, 2007). However, this assessment was not used because it was normed on a monolingual sample in Puerto Rico and Mexico. All children were administered the English version of the measure unless they did not answer the assessors’ initial rapport-building questions.
**Exhibit 1.** Histogram of ages in days (centered at the September 30, 2016 cutoff) for children in the Cohort 3 (control) and Cohort 2 (treatment) analytic samples.

We next examined the distributions of the ages of Cohort 2 and Cohort 3 children. A participant’s ability to manipulate their value on the forcing variable, and thus their eligibility for the treatment, represents a common concern for regression discontinuity studies (Murnane and Willett, 2010). This manipulation of the forcing variable is often observed through a “pileup” of observations on one side of the cutoff. Although it is unlikely that any Virginia parents planned their child’s birthdate to make him or her eligible for VPI+ in September 2016 rather than September 2017, it is possible that parents’ choice regarding the decision to enroll in VPI+ might be influenced by the child’s age relative to the birthdate cut-off. For example, if parents of older four-year old children were more likely to seek enrollment in VPI+, pileup might indicate that parents somehow manipulated their position on either side of the cutoff to access VPI+ each year. We examined whether there was pile up at the cut off using the procedures described by McCrary (2008) as well as those outlined by Cattaneo, Jansson, and Ma (2017). Neither test indicated any evidence of pileup.

We also examined the demographic and test score characteristics of the Cohort 2 and Cohort 3 analysis samples (see Exhibit 2) at Pre-K entry to determine whether the types of children who enrolled in VPI+ in the fall of 2016 differed from children who enrolled in VPI+ in fall 2017. Cohort 2 and Cohort 3 are similar on most demographic characteristics. Cohort 3 (control) analysis sample has a smaller proportion of Asian children and children whose annual family incomes are at or below 100% FPL, and proportionately more children who speak a language other than English at home. Children in the two
cohorts were similar at Pre-K entry, on average, on the two child skills indicators for which we have consistent measurement.6

### Exhibit 2. Characteristics of treatment and control analysis samples at Pre-K Entry

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cohort 2 (Treatment) analysis sample (n = 969)</th>
<th>Cohort 3 (Control) analysis sample (n = 1140)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child age in days as of Sept. 30 of Pre-K year: mean</td>
<td>1,646.2 (108.9)</td>
<td>1,643.2 (106.5)</td>
</tr>
<tr>
<td>standard deviation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>49.0%</td>
<td>48.3%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>30.7%</td>
<td>30.3%</td>
</tr>
<tr>
<td>African American</td>
<td>45.1%</td>
<td>44.9%</td>
</tr>
<tr>
<td>White</td>
<td>17.1%</td>
<td>17.9%</td>
</tr>
<tr>
<td>Asian</td>
<td>3.7%</td>
<td>2.3%*</td>
</tr>
<tr>
<td>At or below 100% of federal poverty level (FPL)</td>
<td>58.1%</td>
<td>53.0%**</td>
</tr>
<tr>
<td>Mother’s highest education is high school</td>
<td>64.0%</td>
<td>62.9%</td>
</tr>
<tr>
<td>Non-English</td>
<td>26.9%</td>
<td>30.4%*</td>
</tr>
<tr>
<td>Child with a disability</td>
<td>4.7%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Test score data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WJ applied problems standard score (SD)</td>
<td>97.9 (12.0)</td>
<td>97.8 (12.5)</td>
</tr>
<tr>
<td>HTKS raw score (SD)</td>
<td>6.0 (12.8)</td>
<td>6.0 (12.3)</td>
</tr>
</tbody>
</table>

T-tests are used for continuous measures, chi-squared tests for binary measures. *p < .10, **p < .05, ***p < .01

Following the guidance of Jacob and colleagues (2012), we first examined the data graphically via scatter plots which display the focal outcome measures on the vertical axis and the running variable (in this case child age in days centered at the age cutoff) on the horizontal axis. We fit a lowess (local polynomial regression) line to the data on each side of the age cutoff to aid in interpretation. These scatter plots are displayed in Appendix Exhibits B1, C1, D1, E1, and F2. First, we examine whether there appears to be a discontinuity in the relationship between child outcomes and age at the eligibility cut-point. We find strong visible evidence of a discontinuity in this relationship at the cut-point for the letter sounds and lower-case alphabet recognition measures, possible evidence of a potential discontinuity for the HTKS and WJ applied problems, and no evidence of a discontinuity for PPVT.

---

6 The study team did not measure the vocabulary skills of Cohort 2 students at PreK entry. The team did measure the literacy skills of both Cohort 2 and Cohort 3 students at PreK entry but inconsistencies in the administration procedures across the years make these scores noncomparable. In fall 2016, Cohort 2 followed the default administration rules for PALS for Pre-K-age children which directs teachers to only administer lower-case alphabet recognition and letter sounds awareness if children demonstrate proficiency in other tasks. In fall 2017, Cohort 3 Pre-K teachers were asked to administer those two tasks (lower-case alphabet recognition and letter sounds awareness) to as many children as appropriate and ignore the default administration guidelines. These differences in the testing rules bias the average baseline (Pre-K entry) literacy skills upward for Cohort 2. In fall 2017, the data collection period that is of interest for the RD impact analyses, the procedures for determining which children were administered the PALS were consistent for the Cohort 2 and Cohort 3 children. As a result, the percentage of Cohort 2 and Cohort 3 children who were assessed on PALS in fall 2017 was nearly identical (98 percent for Cohort 2 and 96 percent for Cohort 3). We are therefore not concerned that the observed differences in average levels of literacy skills at Pre-K entry reflect differences in the baseline skills of the Cohort 2 and Cohort 3 children.
Again, following the Jacob et al. (2012) guidance, we then examined the scatter plots for evidence of discontinuities in the relationship between age and outcomes at ages other than the age eligibility cut-point. Multiple discontinuities would indicate that some factor other than or in addition to participation in VPI+ Pre-K may have influenced the relationship between outcomes and age, and thus threaten the internal validity of the study. Examinations of the scatter plots do not provide evidence of multiple discontinuities for any of the outcome measures.

To determine whether participation in VPI+ led to impacts on children’s developmental skills, we conducted a set of regression analyses for each outcome. We conduct analyses of the effect of enrollment in VPI+ Pre-K on child outcomes using two samples:

1. Complete cases only (excluding the Late joiners, Attritors, Kindergarten no-shows, Kindergarten not tested, and Pre-K not tested)
2. All children except for the Late joiners, in which missing data are imputed.7

For each sample, we then use two analytic approaches for conducting the regression discontinuity analyses:

1. We use a global regression approach in which we use all observations to fit regression lines on either side of the cut-point.8
2. We also fit models in which the bandwidth is empirically derived to minimize mean squared error (following the procedure outlined in Calonico, Cattaneo, Farrell, & Titiunik, 2017). All regression analyses were conducted in STATA using the rdrobust command (Calonico et al., 2017). Regression models account for clustering within school division and include the following covariates (gender, race/ethnicity, income less than 100 percent FPL, mother’s highest level of education, DLL status, and disability status).9

Results

In Exhibit 3 we provide the unadjusted sample means for our Cohort 2 (treatment) and Cohort 3 (control) children in the fall of 2017.

<table>
<thead>
<tr>
<th></th>
<th>Treatment (Cohort 2 at Kindergarten entry)</th>
<th>Control (Cohort 2 at PreK entry)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literacy skills (PALS letter sounds)</td>
<td>15.2 (7.1)</td>
<td>2.6 (5.1)</td>
</tr>
<tr>
<td>Literacy skills (PALS lower-case alphabet recognition)</td>
<td>20.5 6.6</td>
<td>6.4 (12.2)</td>
</tr>
<tr>
<td>Self-regulation (HTKS)</td>
<td>24.4 (21.4)</td>
<td>6.0 (12.3)</td>
</tr>
<tr>
<td>Mathematics (WJ)</td>
<td>97.5 (10.8)</td>
<td>97.7 (12.5)</td>
</tr>
<tr>
<td>Vocabulary (PPVT)</td>
<td>94.4 (15.3)</td>
<td>87.8 (20.5)</td>
</tr>
</tbody>
</table>

7 We created five fully imputed datasets and calculated the average mean effect size across the five.
8 Appendix Exhibits B2, C2, D2, E2, and F2 provide a graphical representation of the results for each of the five outcomes.
9 See Appendices G and H for regression tables
In Exhibit 4 we report the effect sizes differences between treatment and control. We find evidence of positive impacts of enrollment in VPI+ Pre-K on children’s literacy skills and self-regulation across all four analytic models: global regression with complete cases, global regression with missing data imputed, optimal bandwidth with complete cases, and optimal bandwidth with missing data imputed. For measures of literacy skills (lower-case alphabet recognition and letter sounds), we find large (effect sizes between 0.85 to 1.12) and statistically significant ($p < .01$) impacts using both analytic approaches with complete cases and multiply imputed data. Impacts on self-regulation were moderate (effect sizes between 0.26 to 0.38) and statistically significant ($p < .05$ in three models, $p < .10$ in the imputed data/optimal bandwidth analyses) in all analyses.

Estimates of the impacts of enrollment in VPI+ Pre-K on mathematics and vocabulary were not consistent across models. We observe positive and moderately sized effects (effect sizes between 0.28 to 0.33) and statistically significant ($p < .05$) impacts of enrollment in VPI+ Pre-K on mathematics skills when using a global regression approach. We do not observe statistically significant effects on mathematics skills when using the optimal bandwidth approach. We observe small positive impacts of VPI+ Pre-K enrollment on vocabulary skills in the analyses using global regression with complete cases. All other analyses were not statistically significant.$^{10}$

**Exhibit 4.** Effect sizes from RD analyses using complete cases and multiply imputed data paired with global regression and optimal bandwidth approaches. Significance tests use robust, bias corrected standard errors

<table>
<thead>
<tr>
<th></th>
<th>Global regression</th>
<th>Optimal bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete cases</td>
<td>With missing data</td>
</tr>
<tr>
<td></td>
<td>(No imputation)</td>
<td>imputed</td>
</tr>
<tr>
<td>Literacy skills (PALS letter sounds)</td>
<td>1.12***</td>
<td>0.96***</td>
</tr>
<tr>
<td>Literacy skills (PALS lower-case alphabet recognition)</td>
<td>1.10***</td>
<td>0.97***</td>
</tr>
<tr>
<td>Self-regulation (HTKS)</td>
<td>0.38**</td>
<td>0.32**</td>
</tr>
<tr>
<td>Mathematics (WJ)</td>
<td>0.33**</td>
<td>0.28**</td>
</tr>
<tr>
<td>Vocabulary (PPVT)</td>
<td>0.15**</td>
<td>0.14</td>
</tr>
</tbody>
</table>

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

**Sensitivity Analysis**

**Testing for Discontinuity of Baseline Covariates at the Cutoff**

We also tested whether key baseline covariates that are correlated with the outcome variable are continuous at the cutoff. We wanted to know whether there are discontinuities in these covariates at the eligibility cutoff. These can be thought of as falsification tests in which we consider each of the demographic variables to be the outcome in its own RD analysis. As in the outcome analyses, we fit global regression and optimal bandwidth models for each of the covariates. If the birthdate cutoff is

$^{10}$ The addition of the propensity score weights produced meaningfully different results only for the complete case/global regression analyses of vocabulary skills (differences no longer statistically significant).
exogenous we would not observe an impact of the intervention on baseline covariates at the cutoff. When using the optimal bandwidth approach, we do not observe statistically significant differences for any of the covariates at the cutoff. We do observe a marginally significant difference ($p < .10$) in the probability that a child will have a disability.

**Comparison of Children Included Versus Excluded from the Analysis Sample**

For each cohort, we compared the demographic and test score data of the *consistently enrolled and assessed* children to each of the other groups (see Exhibits 5a and 5b below).

In Cohort 2 (treatment condition), we find three statistically significant differences in demographic characteristics between the children in our analytic sample and those children who were not in the analysis sample (*Attritors, Kindergarten no shows, Kindergarten not tested*). There are proportionately 1) more Hispanic children, 2) fewer African American children, and 3) more Non-English speakers in the analysis sample than in the group of children excluded from the analysis sample (Exhibit 5a). The two samples are comparable regarding proportion of White and Asian children, percentage of children at or below 100% FPL, percentage of children whose mother's highest education is high school, and percentage of children with a disability. The groups also perform similarly on all pretest measures.

---

11 We do not include late joiners in this list or subsequent tables. Our reason is that since we will not collect information for this group of children in the Cohort 3 sample they will never be part of the analyses.
Exhibit 5a. Characteristics of analysis sample and subsamples of students in Cohort 2 (treatment) ‡

<table>
<thead>
<tr>
<th>Analysis sample† (n = 969)</th>
<th>Not included in the analysis sample (n = 233)</th>
<th>Attriters (n = 68)</th>
<th>Kindergarten no show (n = 139)</th>
<th>Kindergarten not tested (n = 26)</th>
<th>Pre-K Non-completers (n = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic characteristics (fall 2016)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>48.6%</td>
<td>48.7%</td>
<td>54.1%</td>
<td>47.8%</td>
<td>38.5%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>30.7%</td>
<td>15.6%***</td>
<td>16.2%</td>
<td>16.2%</td>
<td>11.5%</td>
</tr>
<tr>
<td>African American</td>
<td>45.1%</td>
<td>61.7%***</td>
<td>69.1%</td>
<td>61.8%</td>
<td>42.3%</td>
</tr>
<tr>
<td>White</td>
<td>17.1%</td>
<td>13.5%</td>
<td>7.4%</td>
<td>14.0%</td>
<td>26.9%</td>
</tr>
<tr>
<td>Asian</td>
<td>3.7%</td>
<td>5.2%</td>
<td>5.9%</td>
<td>2.2%</td>
<td>19.2%</td>
</tr>
<tr>
<td>At or below 100% FPL</td>
<td>58.1%</td>
<td>62.4%</td>
<td>70.6%</td>
<td>59.9%</td>
<td>53.9%</td>
</tr>
<tr>
<td>Mother’s highest education is HS</td>
<td>64.0%</td>
<td>58.9%</td>
<td>65.1%</td>
<td>57.8%</td>
<td>47.8%</td>
</tr>
<tr>
<td>Non-English</td>
<td>26.9%</td>
<td>15.2%***</td>
<td>14.7%</td>
<td>13.7%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Child with a disability</td>
<td>4.7%</td>
<td>2.6%</td>
<td>0.0</td>
<td>2.9%</td>
<td>7.7%</td>
</tr>
<tr>
<td><strong>Test score data (fall 2016)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-case alphabet recognition raw score (SD)</td>
<td>8.8 (9.2)</td>
<td>10.3 (9.1)</td>
<td>7.7 (8.4)</td>
<td>10.4 (8.9)</td>
<td>14.9 (10.0)</td>
</tr>
<tr>
<td>Letter sounds raw score (SD)</td>
<td>4.2 (6.2)</td>
<td>4.6 (6.1)</td>
<td>3.7 (5.2)</td>
<td>3.9 (5.6)</td>
<td>9.0 (7.8)</td>
</tr>
<tr>
<td>WJ applied problems standard score (SD)</td>
<td>97.9 (12.0)</td>
<td>98.6 (11.5)</td>
<td>96.7 (9.09)</td>
<td>98.7 (12.1)</td>
<td>102.2 (12.2)</td>
</tr>
<tr>
<td>HTKS raw score (SD)</td>
<td>6.0 (12.8)</td>
<td>7.4 (14.1)</td>
<td>4.6 (10.4)</td>
<td>7.7 (14.4)</td>
<td>12.2 (18.3)</td>
</tr>
</tbody>
</table>

† The analysis sample consists of those consistently enrolled and assessed as well as Pre-K Non-completers with outcomes data.
‡ Significance tests report on the difference between sub-sample and analysis samples. T-tests are used for continuous measures, chisq tests for binary measures. *p < .10, **p < .05, ***p < .01

In Cohort 3 (control condition), the analysis sample and the group of children not included in the analysis sample differ significantly on one demographic characteristic: there are proportionately more White children in the non-analysis group than there are in the analysis group (Exhibit 5b).

The two groups are similar on all other demographic characteristics and pretest measures. However, the sample size of the group not included in the analysis is small, and statistical tests should be interpreted with caution. For example, none of the children in the Pre-K not tested or Pre-K non-completer groups had a disability compared to nearly 6 percent of children in the analysis sample.
**Exhibit 5b.** Characteristics of analysis sample and subsamples of students in Cohort 3 (control) ‡

<table>
<thead>
<tr>
<th>Demographic characteristics (fall 2017)</th>
<th>Analysis sample (n = 1140) †</th>
<th>Not included in the analysis sample Pre-K not tested (n = 19)</th>
<th>Pre-K non-completers (n =34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>48.3%</td>
<td>38.9%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>30.3%</td>
<td>11.1%</td>
<td>17.9%</td>
</tr>
<tr>
<td>African American</td>
<td>44.9%</td>
<td>38.9%</td>
<td>50.0%</td>
</tr>
<tr>
<td>White</td>
<td>17.9%</td>
<td>38.9%**</td>
<td>17.9%</td>
</tr>
<tr>
<td>Asian</td>
<td>2.3%</td>
<td>0.0%</td>
<td>3.6%</td>
</tr>
<tr>
<td>At or below 100 % FPL</td>
<td>53.0%</td>
<td>68.4%</td>
<td>72.4%</td>
</tr>
<tr>
<td>Mother’s highest education is high school</td>
<td>62.9%</td>
<td>50.0%</td>
<td>61.5%</td>
</tr>
<tr>
<td>Non-English</td>
<td>30.4%</td>
<td>21.1%</td>
<td>19.4%</td>
</tr>
<tr>
<td>Child with a disability</td>
<td>5.9%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Test score data (fall 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-case alphabet recognition raw score</td>
<td>6.4 (8.2)</td>
<td>missing</td>
<td>8.4 (9.9)</td>
</tr>
<tr>
<td>Letter sounds raw score</td>
<td>2.7 (5.1)</td>
<td>missing</td>
<td>5.7 (7.8)</td>
</tr>
<tr>
<td>WJ applied problems standard score</td>
<td>97.8 (12.5)</td>
<td>missing</td>
<td>97.1 (14.9)</td>
</tr>
<tr>
<td>HTKS raw score</td>
<td>6.0 (12.3)</td>
<td>missing</td>
<td>5.3 (10.2)</td>
</tr>
</tbody>
</table>

†The analysis sample includes children consistently enrolled and assessed and Pre-K non-completers with outcomes data. ‡T-tests are used for continuous measures, chi-square tests for binary measures. *p < .10, **p < .05, ***p < .01

We also conducted sensitivity checks in which we used propensity scores to account for cohort differences. Following the example in the Weiland and Yoshikawa (2013) RD study of the impacts of the Pre-Kindergarten program in the Boston Public Schools, we use propensity scores to account for differences between the cohorts on observable characteristics. We do not use the children’s test scores from the fall of the Pre-K year in the calculation of propensity scores, as these represent the outcome data for the Cohort 3 (control) children. The propensity scores calculated the probability of being in the treatment group using the family demographic characteristics listed in Exhibit 2. We then take the inverse of the resulting propensity score for the Cohort 3 children and use them as weights in our regression analyses. When we apply these weights to our Cohort 2 and Cohort 3 analytic samples, we no longer see significant differences on any of the observed measures of baseline child and family demographics. We do continue to see differences on children’s fall Pre-K scores for letter sounds, lower-case alphabet recognition, and self-regulation skills.

**Limitations**

Prior research has noted several important limitations to examining the impacts of four-year-old preschool programs using an age cut-off regression discontinuity design (Lipsey et al., 2015). Although we can address many of these issues, there are two that remain as limitations.
Differential Attrition

As noted above, children in the control group are measured at entry into prekindergarten, while the treatment group children are examined after the completion of the preschool and matriculating into kindergarten. This means that although treatment group children have 12 months in which to attrite from the sample, control group children do not have the same opportunity to attrite. We do observe a higher rate of attrition among treatment group children (22%) than from control group children (4%). If attritors are, on average, more disadvantaged than non-attritors, then this would bias our estimates of the treatment impact upward. Although we are not able to address this concern in this report of preliminary findings, we will be able to address this concern prior to completing this study. The study team is collecting information on the control group students as they enter kindergarten in the fall of 2018. The team is using the identical data collection procedures for these students as was used for the treatment group students who entered kindergarten in the fall of 2017. We will remove control group students for whom we cannot collect outcome data in the fall of 2018 and report findings on this reduced sample.

Differential Experience with Data Collection

For some children, preschool represents the first time that they have needed to interact with a set of unfamiliar adults such as data collectors. Treatment group children will have had the opportunity to grow accustomed to this through participation in preschool and interactions with study data collectors on two other occasions (fall of 2016 and spring of 2017). It is possible that this greater familiarity with the data collection process among treatment group students biases the results of impact estimates upwards. Data collectors followed protocols designed to reduce any anxieties that children might feel, but it is impossible for us to rule out that the observed treatment effect is not a result of control group children’s unfamiliarity with testing procedures.

Discussion

These analyses provide evidence that participation in VPI+ has a positive impact on key skills that predict children’s school success. Findings show strong impacts on literacy, moderate impacts on self-regulation, and small to moderate impacts on mathematics. This pattern of findings is generally similar to those seen in studies of nationally recognized, high-quality preschool programs, such as those in Tulsa, Oklahoma and Boston, Massachusetts (Gormley et al., 2011; Weiland & Yoshikawa, 2013). Taken together, these analyses provide evidence that participating in VPI+ accelerated children’s development of important school readiness skills. In the 12 months between enrolling in VPI+ and beginning kindergarten, VPI+ children developed more than 20 months of literacy skills and 15 months of mathematics skills. These impacts on literacy are consistent with what has been found in other analyses of high-quality public preschool programs. We also observed a moderately-sized impact on children’s self-regulation skills. These effects are of similar magnitude to the impacts observed in the Boston Public Schools Pre-K program. However, we did not observe any impacts on children’s vocabulary skills.

The findings about VPI+ impacts support and replicate other contemporary evaluations of state (Barnett, et al., 2018; Lipsey, Farran, & Durkin, 2018) and locally-funded preschool initiatives, including Boston, San Antonio, Seattle, and Denver. Mainly, it provides evidence that improving child outcomes in the areas of discrete skills like letter and sound recognition and counting is achievable when programs invest in a set of standards that are universally recognized as necessary, including teacher-child ratios, teacher qualifications, and resources and supports including an evidence-based curriculum and professional development and coaching. The particularly strong impacts on children’s literacy skills may reflect a long history of focusing on literacy in VPI and now VPI+ programs. Data from the VPI+ evaluation show that a clear majority of teachers reported spending time every day building and supporting literacy skills, and much of the VPI+ coaching focuses on this area (Gaylor et al., 2018). Data from the VPI+ evaluation also indicates that VPI+ teachers receive substantial training and coaching with supporting children’s social and emotional skills.

Improving children’s vocabulary skills appears to represent a greater challenge for the VPI+ program. This finding is consistent with what has been observed in evaluations of prekindergarten programs in other states (Barnett, 2018) where impacts on vocabulary are smaller and less robust. It also reflects the challenge of building an unconstrained skill (Snow & Matthews, 2016), like vocabulary, in a one-year program and being able to accurately assess growth in vocabulary skills given that assessments may not tap the specific new words learned. Yet, we know from prior research that children’s vocabulary skills at kindergarten entry are highly predictive of long-term literacy success in third grade and beyond. We also know that approximately 80% of the words children acquire in the early years of schooling are learned through direct instruction (Biemiller & Slonim, 2001). VDOE and the VPI+ school divisions might therefore consider additional resources for teachers and professional development efforts focused on helping to build children’s vocabulary skills.
References


Appendix A: Attrition in treatment and control groups.

Full Sample (n=2,458)

Cohort 2 (Treatment) (n=1,295)
Enrolled in VPI+ Pre-K in fall 2016

Attriters (n=68)
Enrolled in VPI+ Pre-K in fall 2016, departed before spring 2016 and did not enroll in K in fall 2017

Kindergarten no shows (n=139)
Enrolled in VPI+ Pre-K in fall 2016 and spring 2017 but did not enroll in Kinder in fall 2017

Late joiners (n=93)
Enrolled in VPI+ Pre-K after fall 2016

Consistently enrolled and assessed (n=926)
Enrolled in VPI+ Pre-K in fall 2016, and spring 2017, enrolled in Kinder in fall 2017, and has valid outcomes data

PreK non-completers (n=44)
Enrolled in VPI+ Pre-K in fall 2016, but not spring 2017 and enrolled in Kinder in fall 2017

RD Treatment Group n=969

Cohort 3 (Control) (n=1,163)
Enrolled in VPI+ Pre-K in fall 2017

Kindergarten no shows (n=26)
Enrolled in VPI+ Pre-K in fall 2016, and enrolled in Kinder in fall 2017 but did not have Outcomes data in fall 2017

Pre-K not tested (n=19)
Enrolled in fall 2017 and December 2017, but has no Outcomes measures

Consistently enrolled and assessed (n=1,121)
Enrolled in VPI+ Pre-K in fall 2017 through December 2017 and had outcomes data

PreK non-completers (n=34)
Enrolled in VPI+ Pre-K in fall 2017 but left before December 2017

RD Control Group (n=1,140)

Full RD Analysis Sample N=2,109

*Shaded in orange are children we chose to drop; shaded in yellow are children with missing data; shaded in blue are children included in the sample if they had valid outcomes data. Please note that the two blue boxes will not add up to the number in the light green box because the Treatment and Control groups exclude any Pre-K non-completers that did not have valid outcomes data. For Cohort 2, 1 Pre-K non-completer does not have outcome data and is excluded from the Treatment group. For Cohort 3, 15 children do not have outcomes data and are excluded from the Control group.
Appendix Exhibit B1: Scatterplot of vocabulary skills by child age in day (centered at the cutoff)

![Scatterplot of vocabulary skills by child age in day (centered at the cutoff)](image)

Appendix Exhibit B2: Regression discontinuity plot of vocabulary skills by child age in days (centered at the cutoff). Regression lines are smooth “global” polynomial regression curve estimates for Cohort 3 and Cohort 2 units separately. Error bars represent 95% CI around the average within each bin.

![Regression discontinuity plot of vocabulary skills by child age in days (centered at the cutoff)](image)
Appendix Exhibit C1: Scatterplot of mathematics skills by child age in day (centered at the cutoff)

![Scatterplot of mathematics skills by child age in day](image)

Appendix Exhibit C2: Regression discontinuity plot of mathematics skills by child age in days (centered at the cutoff). Regression lines are smooth “global” polynomial regression curve estimates for Cohort 3 and Cohort 2 units separately. Error bars represent 95% CI around the average within each bin.

![Regression discontinuity plot](image)
Appendix Exhibit D1 Scatterplot of HTKS by child age in day (centered at the cutoff)

[Scatterplot image]

Appendix Exhibit D2: Regression discontinuity plot of HKTS skills by child age in days (centered at the cutoff). Regression lines are smooth “global” polynomial regression curve estimates for Cohort 3 and Cohort 2 units separately. Error bars represent 95% CI around the average within each bin.

[RD Plot image]
Appendix Exhibit E1: Scatterplot of number of letter sounds identified by child age in day (centered at the cutoff)

Appendix Exhibits E2. Regression discontinuity ploy of letter sounds identified by child age in days (centered at the cutoff). Regression lines are smooth “global” polynomial regression curve estimates for Cohort 3 and Cohort 2 units separately. Error bars represent 95% CI around the average within each bin.
Appendix Exhibit F1: Scatterplot of lowercase alphabet recognition by child age in day (centered at the cutoff)

Appendix Exhibit F2: Regression discontinuity ploy of lowercase alphabet letters identified by child age in days (centered at the cutoff). Regression lines are smooth “global” polynomial regression curve estimates for Cohort 3 and Cohort 2 units separately. Error bars represent 95% CI around the average within each bin.
Appendix Exhibit G: Regression tables with no imputation

Optimal bandwidth with no imputation

<table>
<thead>
<tr>
<th></th>
<th>(1) Mathematics</th>
<th>(2) Vocabulary</th>
<th>(3) Behavior regulation</th>
<th>(4) Literacy skills: Letter sounds</th>
<th>(5) Literacy skills: Lowercase letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>4.226 (3.850)</td>
<td>2.737 (2.934)</td>
<td>7.524*** (2.645)</td>
<td>9.564*** (1.225)</td>
<td>10.45*** (1.590)</td>
</tr>
<tr>
<td>Bandwidth in days</td>
<td>117.5</td>
<td>158.2</td>
<td>142.9</td>
<td>114.2</td>
<td>101.2</td>
</tr>
<tr>
<td>from the cutoff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>367</td>
<td>502</td>
<td>452</td>
<td>346</td>
<td>315</td>
</tr>
<tr>
<td>Cohort 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>315</td>
<td>408</td>
<td>370</td>
<td>312</td>
<td>278</td>
</tr>
<tr>
<td>Cohort 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Global regression approach with no imputation

<table>
<thead>
<tr>
<th></th>
<th>(1) Mathematics</th>
<th>(2) Vocabulary</th>
<th>(3) Behavior regulation</th>
<th>(4) Literacy skills: Letter sounds</th>
<th>(5) Literacy skills: Lowercase letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>8.259*** (2.315)</td>
<td>4.079** (2.060)</td>
<td>7.410*** (1.201)</td>
<td>9.631*** (0.841)</td>
<td>11.36*** (1.000)</td>
</tr>
<tr>
<td>Bandwidth in days</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>from the cutoff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>1094</td>
<td>1086</td>
<td>1098</td>
<td>1076</td>
<td>1088</td>
</tr>
<tr>
<td>Cohort 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>945</td>
<td>946</td>
<td>946</td>
<td>940</td>
<td>940</td>
</tr>
<tr>
<td>Cohort 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix Exhibit H: Regression tables with imputation

### Optimal bandwidth

#### RD Estimates-Optimal bandwidth, **Vocabulary**

<table>
<thead>
<tr>
<th></th>
<th>(1) Imputation 1</th>
<th>(2) Imputation 2</th>
<th>(3) Imputation 3</th>
<th>(4) Imputation 4</th>
<th>(5) Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>3.496</td>
<td>4.948*</td>
<td>3.463</td>
<td>2.715</td>
<td>2.786</td>
</tr>
<tr>
<td></td>
<td>(3.181)</td>
<td>(2.834)</td>
<td>(3.149)</td>
<td>(3.200)</td>
<td>(3.000)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>168.4</td>
<td>164.4</td>
<td>115.3</td>
<td>157.7</td>
<td>148</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>553</td>
<td>544</td>
<td>374</td>
<td>523</td>
<td>488</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>544</td>
<td>527</td>
<td>384</td>
<td>508</td>
<td>477</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \)

#### RD Estimates-Optimal bandwidth, **Mathematics**

<table>
<thead>
<tr>
<th></th>
<th>(1) Imputation 1</th>
<th>(2) Imputation 2</th>
<th>(3) Imputation 3</th>
<th>(4) Imputation 4</th>
<th>(5) Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>4.487</td>
<td>2.244</td>
<td>3.608</td>
<td>1.292</td>
<td>2.098</td>
</tr>
<tr>
<td></td>
<td>(3.327)</td>
<td>(3.425)</td>
<td>(3.673)</td>
<td>(3.900)</td>
<td>(3.807)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>139.9</td>
<td>116.6</td>
<td>104.6</td>
<td>116.9</td>
<td>112.3</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>458</td>
<td>379</td>
<td>340</td>
<td>378</td>
<td>362</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>451</td>
<td>389</td>
<td>354</td>
<td>390</td>
<td>378</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \)
## RD Estimates - Optimal bandwidth, Behavior regulation

<table>
<thead>
<tr>
<th></th>
<th>(1) Imputation 1</th>
<th>(2) Imputation 2</th>
<th>(3) Imputation 3</th>
<th>(4) Imputation 4</th>
<th>(5) Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>3.689 (3.575)</td>
<td>5.749** (2.563)</td>
<td>6.117** (2.516)</td>
<td>4.812 (3.501)</td>
<td>6.098** (2.622)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>95.46</td>
<td>125.1</td>
<td>128.2</td>
<td>102.5</td>
<td>120.6</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>311</td>
<td>411</td>
<td>417</td>
<td>336</td>
<td>392</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>325</td>
<td>415</td>
<td>421</td>
<td>347</td>
<td>405</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1

## RD Estimates - Optimal bandwidth, Letter Sounds

<table>
<thead>
<tr>
<th></th>
<th>(1) Imputation 1</th>
<th>(2) Imputation 2</th>
<th>(3) Imputation 3</th>
<th>(4) Imputation 4</th>
<th>(5) Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>8.703*** (1.236)</td>
<td>7.674*** (1.380)</td>
<td>9.251*** (1.058)</td>
<td>9.224*** (1.211)</td>
<td>8.512*** (1.518)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>127</td>
<td>86</td>
<td>126.9</td>
<td>130</td>
<td>98.40</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>411</td>
<td>280</td>
<td>412</td>
<td>420</td>
<td>324</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>417</td>
<td>299</td>
<td>416</td>
<td>427</td>
<td>327</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1
### RD Estimates-Optimal bandwidth, Lowercase letters

<table>
<thead>
<tr>
<th></th>
<th>Imputation 1</th>
<th>Imputation 2</th>
<th>Imputation 3</th>
<th>Imputation 4</th>
<th>Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.761)</td>
<td>(1.577)</td>
<td>(1.584)</td>
<td>(1.644)</td>
<td>(1.835)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>99.27</td>
<td>93.60</td>
<td>103.6</td>
<td>91.67</td>
<td>90.07</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>325</td>
<td>299</td>
<td>339</td>
<td>296</td>
<td>295</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>332</td>
<td>317</td>
<td>352</td>
<td>310</td>
<td>308</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1

### Imputed data, Global regression

### RD Estimates-Global Regression, Vocabulary

<table>
<thead>
<tr>
<th></th>
<th>Imputation 1</th>
<th>Imputation 2</th>
<th>Imputation 3</th>
<th>Imputation 4</th>
<th>Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>4.162*</td>
<td>4.259*</td>
<td>3.424*</td>
<td>4.903**</td>
<td>3.518</td>
</tr>
<tr>
<td></td>
<td>(2.361)</td>
<td>(2.237)</td>
<td>(2.027)</td>
<td>(2.499)</td>
<td>(2.218)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>1152</td>
<td>1153</td>
<td>1153</td>
<td>1152</td>
<td>1153</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>1201</td>
<td>1200</td>
<td>1200</td>
<td>1201</td>
<td>1200</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1
## RD Estimates - Global Regression, Mathematics

<table>
<thead>
<tr>
<th></th>
<th>(1) Imputation 1</th>
<th>(2) Imputation 2</th>
<th>(3) Imputation 3</th>
<th>(4) Imputation 4</th>
<th>(5) Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>7.386***</td>
<td>6.586***</td>
<td>7.593***</td>
<td>6.955***</td>
<td>6.540***</td>
</tr>
<tr>
<td></td>
<td>(2.402)</td>
<td>(2.308)</td>
<td>(2.322)</td>
<td>(2.323)</td>
<td>(2.420)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>1152</td>
<td>1153</td>
<td>1153</td>
<td>1152</td>
<td>1153</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>1201</td>
<td>1200</td>
<td>1200</td>
<td>1201</td>
<td>1200</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

## RD Estimates - Global Regression, Behavior regulation

<table>
<thead>
<tr>
<th></th>
<th>(1) Imputation 1</th>
<th>(2) Imputation 2</th>
<th>(3) Imputation 3</th>
<th>(4) Imputation 4</th>
<th>(5) Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.241)</td>
<td>(1.228)</td>
<td>(1.019)</td>
<td>(1.154)</td>
<td>(1.209)</td>
</tr>
<tr>
<td>Bandwidth in days from the cutoff</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>Number of obs Cohort 2</td>
<td>1152</td>
<td>1153</td>
<td>1153</td>
<td>1152</td>
<td>1153</td>
</tr>
<tr>
<td>Number of obs Cohort 3</td>
<td>1201</td>
<td>1200</td>
<td>1200</td>
<td>1201</td>
<td>1200</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
### RD Estimates - Global Regression, Letter Sounds

<table>
<thead>
<tr>
<th></th>
<th>Imputation 1</th>
<th>Imputation 2</th>
<th>Imputation 3</th>
<th>Imputation 4</th>
<th>Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.798)</td>
<td>(0.671)</td>
<td>(0.767)</td>
<td>(0.927)</td>
<td>(0.894)</td>
</tr>
<tr>
<td>Bandwidth in days</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>from the cutoff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>1152</td>
<td>1153</td>
<td>1153</td>
<td>1152</td>
<td>1153</td>
</tr>
<tr>
<td>Cohort 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>1201</td>
<td>1200</td>
<td>1200</td>
<td>1201</td>
<td>1200</td>
</tr>
<tr>
<td>Cohort 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

### RD Estimates - Global Regression, Lowercase letters

<table>
<thead>
<tr>
<th></th>
<th>Imputation 1</th>
<th>Imputation 2</th>
<th>Imputation 3</th>
<th>Imputation 4</th>
<th>Imputation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient from RD</td>
<td>10.33***</td>
<td>10.28***</td>
<td>10.47***</td>
<td>10.94***</td>
<td>10.35***</td>
</tr>
<tr>
<td></td>
<td>(1.013)</td>
<td>(0.853)</td>
<td>(1.001)</td>
<td>(0.943)</td>
<td>(1.101)</td>
</tr>
<tr>
<td>Bandwidth in days</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>from the cutoff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>1152</td>
<td>1153</td>
<td>1153</td>
<td>1152</td>
<td>1153</td>
</tr>
<tr>
<td>Cohort 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of obs</td>
<td>1201</td>
<td>1200</td>
<td>1200</td>
<td>1201</td>
<td>1200</td>
</tr>
<tr>
<td>Cohort 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Appendix G: Dual Language Learner (DLL) Child Assessment Protocol

1. **Assignment of assessors.** Assessment coordinators will review all child demographic information, and assign the assessments of DLL children to bilingual assessors (in the case of Spanish-speaking children, a Spanish-English bilingual assessor will be assigned, and in the case of children with a first language other than English, efforts will be made to assign an assessor that is bilingual in English and the first language of the child, for the purposes of building rapport).

   If the child speaks a language other than Spanish or English at home, but there is not an assessor who speaks that language, the assessor should ask the child the English Screener Questions (see Box 2) to gain a sense of the child’s ability to understand and speak in English. After completing the questions in Box 2, the assessor should record how many questions in English the child responded to with acceptable and sufficiently complex answers.

2. **Rapport-building.** The assessor will begin by asking the child a short series of rapport-building questions in Spanish to ensure the child is comfortable interacting with the assessor and so the child knows it’s okay to speak in Spanish (or other first language if available).

   a. Because children often associate English with the language they are expected to use in preschool the assessor should speak in Spanish when in the classroom (if possible), and may use the following prompt to encourage the child to speak in Spanish:
      
      • “Do you speak Spanish? Oh, I like Spanish. Let’s talk in Spanish for a couple minutes.”

   b. The assessor will ask the child all five questions in Spanish from Box 1 to build rapport and facilitate having a conversation with the child.
Box 1: Rapport-Building Questions in Spanish

| Question #1: | ¿Qué comida te gusta comer para el desayuno?  
Acceptable answers would be the name of a specific food item |
| Question #2: | ¿Quién te dio esos zapatos que tienes puestos? [O “¿Quién es ese en tus zapatos?” Si es un personaje]  
Acceptable answers would be the name of a person who got them their shoes, or “I don’t know.” If asked the second question, the name of the character |
| Question #3: | ¿Cuáles son algunas de tus cosas favoritas que haces en la escuela (o el colegio)?  
Acceptable answers would be any activities children participate in at school – e.g., singing, reading books, playing outside |
| Question #4: | ¿Te gusta colorear o dibujar? [Si es así], ¿Qué te gusta colorear o dibujar?  
Acceptable answers are yes or no. If yes, the name of a person, animal, place, or thing the child likes to color or draw. The child could also mention coloring books.  
Si al niño no le gusta dibujar o colorear, pregúntale:  
¿Te gusta cantar canciones o bailar? [Si le dice que sí], ¿Qué tipos de canciones te gusta cantar o bailar?  
Acceptable answers would be yes or no. If yes, the name of a song or type of music |
| Question #5: | ¿Qué tipo de juegos te gusta jugar con tus amigos?  
Acceptable answers would be a type of game |

3. Determining whether the child should be assessed in English or Spanish.

a. To facilitate the process of shifting from speaking Spanish to English, the assessor will say, “Thank you for talking with me in Spanish. Some of the games today are in English so I just want to ask you some questions in English. Is that okay?”

b. The assessor will ask the child all five questions in Box 2.
Box 2: English Screener Questions

Question #1: What food do you like to eat at lunch?  
(Acceptable answers would be the name of a specific food item)

Question #2: Who got you that shirt (or dress) you are wearing? [Or ‘Who is that on your shirt (or dress)’ if it’s a character]  
(Acceptable answers would be the name of the person who got them their shirt or dress, or “I don’t know.” If asked the second question, the name of the character.)

Question #3: What games do you like to play when you go outside?  
(Acceptable answers would be any outdoor activities – e.g., going on the swings, running, playing in the sandbox)

Question #4: Do you like to watch TV or movies? [If yes], What TV shows or movies do you like to watch?  
(Acceptable answers would be yes or no. If yes, the name of a TV show or movie)  
If the child does not watch TV or movies, ask:  
Do you like to read books or listen to stories? [If yes], What types of books or stories do you like?  
(Acceptable answers would be yes or no. If yes, the name of a specific book or story, or a type of book or story [e.g., about animals, alphabet])

Question #5: What do you like to do at home with your family?  
(Acceptable answers would be an activity the child does at home with his/her family)

C. In addition to considering whether the child provided acceptable answers to the questions asked in Spanish (Box 1) and English (Box 2), the assessor will also take into consideration the complexity (i.e., the child answers with multiple words, without long pauses between words, and using vocabulary beyond simple terms such as “yes,” “no,” “things,” “it,” etc.) when determining the child’s dominant language to decide which language to use for assessing the child.
The assessor will answer the following two questions on the score sheet.

- On a scale of 0 to 5, how many questions did the child answer with sufficient complexity in Spanish?
- On a scale of 0 to 5, how many questions did the child answer with sufficient complexity in English?

Scoring instructions for both Spanish and English questions: If the child answered all questions with acceptable responses and with sufficient complexity across the questions, the child would be rated as 5. If the child did not answer any of the questions, the child would be rated as 0. If the child answered some questions appropriately and with sufficient complexity but not others, the child would be rated 1 to 4 depending on their responses.

### d. The assessor will use the following guidelines to determine the appropriate language battery:

- If language proficiency seems about even, (e.g., the child answers four or five questions with sufficient complexity in Spanish and answers 4 or 5 questions with sufficient complexity in English), the assessor will proceed with the English-language battery.

- If the child answers more questions with sufficient complexity in Spanish as compared to English [rating is higher for Spanish items than English], the assessor will proceed with the Spanish-language battery.

- If the child answers more questions with sufficient complexity in English as compared to Spanish [rating is higher for English items than Spanish], the assessor will proceed with the English-language battery.

- If the child receives a rating of 1 to 3 on Spanish and English questions, and the child is listed as having Spanish as their home language, the assessor will proceed with the Spanish-language battery.

- If the child receives a rating of 1 to 3 on English questions, and the child is listed as having another language at home (e.g., Arabic, Vietnamese), the assessor will administer the PPVT-4 in English (and for Fairfax kindergarteners ONLY, the two PALS-K tasks in English) and then terminate the assessment.

- If the child does not answer any questions in any language, the assessor will try to continue building rapport. If the child continues to be non-responsive, the assessor will bring the child back to the classroom and will try to assess the child at a later time.
### Assessment Battery Order and Instructions

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Instructions</th>
<th>Spanish-language battery</th>
<th>English-language battery</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Head Toes Knees Shoulders (HTKS)</strong></td>
<td>The assessor should ONLY use the language determined through the screening process.</td>
<td>Assessor will administer HTKS in Spanish.</td>
<td>Assessor will administer HTKS in English.</td>
</tr>
<tr>
<td></td>
<td>Only in extreme circumstances (e.g., the child requests to switch to the other language) should the assessor switch the language. We anticipate that in 99% of cases, the assessor will NOT switch languages.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessor will administer HTKS in Spanish.</td>
<td></td>
<td>Assessor will administer HTKS in Spanish.</td>
<td>Assessor will administer HTKS in English.</td>
</tr>
<tr>
<td><strong>2. Woodcock Johnson (WJ-III) applied problems</strong></td>
<td>Administer the practice problems and test items using the child’s dominant language determined through the screening process. The assessor should NOT switch languages during this assessment.</td>
<td>Assessor will administer the practice problems and assessment items from the Batería. The child may respond in either Spanish or English to receive credit for correct responses.</td>
<td>Assessor will administer the practice problems and assessment items from the English version. The child may respond in either English or Spanish to receive credit for correct responses.</td>
</tr>
<tr>
<td>Assessor will administer the practice problems and assessment items from the Batería. The child may respond in either Spanish or English to receive credit for correct responses.</td>
<td></td>
<td>Assessor will administer the practice problems and assessment items from the English version. The child may respond in either English or Spanish to receive credit for correct responses.</td>
<td></td>
</tr>
<tr>
<td><strong>3. Peabody Picture Vocabulary Test (PPVT-4)</strong></td>
<td>All children will be assessed in English ONLY for the PPVT-4. The assessor will read the instructions in English ONLY.</td>
<td>The assessor should start off by saying, “I just wanted to ask you a couple of questions in English again.” Administer the PPVT-4 in English.</td>
<td>Administer the PPVT-4 in English.</td>
</tr>
<tr>
<td><strong>4. ONLY FOR ONE DIVISION THAT DOES NOT USE PALS-K FOR KINDERGARTEN CHILDREN</strong></td>
<td>All children will be assessed in English ONLY for two PALS-K tasks: 1) lowercase alphabet recognition and 2) letter sound awareness.</td>
<td>The assessor will say, “Now I’m going to ask you some different questions in English.” Administer both PALS-K tasks in English.</td>
<td>Assessor will administer the two PALS-K tasks in English.</td>
</tr>
<tr>
<td>Phonological Awareness Literacy Screening-Kindergarten (PALS-K)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5. Motor assessment</strong></td>
<td>Only administered in preschool (not kindergarten)</td>
<td>Assessor will administer the motor assessment in Spanish.</td>
<td>Assessor will administer the motor assessment in English.</td>
</tr>
<tr>
<td>Assessor will administer the motor assessment in Spanish.</td>
<td>Assessor will administer the motor assessment in English.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>