Impact of the Tennessee Voluntary Prekindergarten Program on Children's Literacy, Language, and Mathematics Skills: Results from a Regression-Discontinuity Design

Georgine M. Pion

Mark W. Lipsey

Vanderbilt University

Working Paper

August 2021

Note: In press, AERA Open

PEABODY research institute

VANDERBILT PEABODY COLLEGE

Acknowledgements:

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305E090009 to Vanderbilt University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Without the partnership of the Tennessee Department of Education, this study would not have been possible. We are especially grateful to Connie Casha and Bobbi Lussier, TN-DOE administrators who provided critical support when the project was launched. The Tennessee Education Research Alliance at Vanderbilt has provided invaluable assistance for obtaining and interpreting data from the Tennessee education data system.

We would also like to gratefully acknowledge the numerous members of the research team who have contributed greatly to this project over many years: Dale Farran (Co-Principal Investigator), Kerry Hofer, Carol Bilbrey, Janie Hughart, Ilknur Sekmen, Nianbo Dong, and the many data collectors who worked across the state.

Abstract

The Tennessee Voluntary Pre-K Program (TN-VPK) is statewide full-day program that gives priority to children from low-income families. A regression-discontinuity design with a statewide probability sample of 155 TN-VPK classrooms and 5,189 children participating across two pre-k cohorts found positive effects at kindergarten entry with the largest effects for literacy skills and the smallest for language skills. The results contribute to the growing body of regression-discontinuity studies of state and local pre-k programs and affirm the statewide generalizability of analogous prior findings from a more specialized subsample in the parent Tennessee Pre-k Study. Further, the respective effect sizes compared favorably with those found in other regression-discontinuity studies of public pre-k on the same outcome measures, providing one index of the quality of the TN-VPK program.

Impact of the Tennessee Voluntary Prekindergarten Program on Children's Literacy, Language, and Mathematics Skills: Results from a Regression-Discontinuity Design

Prior to 1980, only two states funded pre-k programs; currently 44 states and the District of Columbia do so (Friedman-Krauss et al., 2020). The objectives of the advocacy groups and state legislatures promoting this expansion vary, but a common theme is enhancing the school readiness of children prior to kindergarten entry, especially children from economically disadvantaged families. Alongside that short-term goal, however, is an expectation that pre-k will have longer-term effects on cognitive and academic skills that will help close the achievement gap between children from low-income families and their more advantaged counterparts (Phillips et al., 2017).

Unfortunately questions about both short and long-term educational effects are difficult to address in a methodologically rigorous way for contemporary state-funded pre-k programs. A randomized study would require assembling a sample of pre-k eligible children before the beginning of the school year whose parents consented to having them randomly assigned to attend the state pre-k or not. Few families would allow researchers to make that decision for their children as they consider the available range of options, e.g., Head Start, private preschool, and home-based care in addition to state pre-k. As a result, most studies of the immediate effects of state pre-k programs, and all but one study of their longer-term effects, have used nonrandomized designs despite the inherent parental discretion involved and the associated potential for selection bias (Farran & Lipsey, 2016).

One notable circumstance amenable to a randomized design, however, is when a number of pre-k sites receive more applicants than they are able to serve, and thus of

necessity must turn some away. Randomization in this situation provides an equitable and transparent way to allocate the available seats. This approach was used in the Head Start Impact Study (Puma et al., 2012) and in the Tennessee Pre-K Study that is the topic of this paper.

The Tennessee Pre-K Study

The Tennessee Pre-K Study is an evaluation of the Tennessee Voluntary Prekindergarten Program (TN-VPK) designed to assess effects on cognitive skills and academic performance at the end of the pre-k year, and the extent to which those effects are sustained after the pre-k year. This study had several components, the primary one being an RCT in which nearly 3000 applicants to oversubscribed program sites were randomized to offers of admission or a wait list control. A second component (the *intensive substudy*) used a subsample of consented children to compare pre-k participants and nonparticipants on a battery of researcher-collected outcome measures to track achievement indicators in the years before scores on the third grade state achievement tests became available.

These two components of the study produced two especially noteworthy findings relevant to the present paper (Lipsey, Farran, & Durkin, 2018). First, the intensive substudy showed moderately large positive effects at the end of the pre-k year on Woodcock Johnson literacy, language, and math measures and on kindergarten teachers' ratings at the beginning of the kindergarten year for various school readiness skills and behaviors. In this regard, the results of the Tennessee study mirrored the finding of other studies of state pre-k programs that, almost without exception, have shown positive effects on cognitive and early achievement measures at the end of the pre-k year (Duncan and Magnuson, 2013).

Second, the positive effects favoring pre-k participants diminished after the pre-k year and, by the end of kindergarten, there were no significant differences on any of the outcomes measured for the intensive substudy sample. The full RCT sample, in turn, showed that there was also no advantage for pre-k participants on the state achievement tests administered in third grade. This "fadeout" of sustained effects on cognitive and academic performance outcomes after positive pre-k impact is well known as a typical finding of pre-k studies with longer-term follow-up (Bailey et al., 2017).

A striking and unexpected finding of the Tennessee study, however, was a negative effect on the follow-up cognitive and achievement outcomes. Through third grade this pattern was seen in the direction of effects on the measures used in the intensive substudy and in the state achievement test scores analyzed for the full RCT, where they were large enough to reach statistical significance for math and science. Further follow-up has revealed that this pattern has continued through sixth grade with scores on the reading achievement test now also showing statistically significant negative effects (manuscript under review).

The Regression-Discontinuity Study Component

The component of the Tennessee study that is the focus of this report used the pre-k age-cutoff version of the regression-discontinuity (RD) design first described by Gormley and colleagues (2005) in their evaluation of the pre-k program in Tulsa. This design compares outcomes for two cohorts of pre-k participants differentiated by an age cutoff for eligibility to enroll during a given school year. Children in the first cohort are old enough to enroll in the program during that school year and, as the treatment group, have their outcomes measured at the beginning of the next school year shortly after they enter kindergarten. Children in the second cohort are too young by a year or less and

are not enrolled in the pre-k program during that given school year. They constitute the control group and their outcomes are measured at the beginning of the next year shortly after they enter the pre-k program for which they are then eligible. These two cohorts, of course, differ on age, but at the age cutoff that difference is negligibly small. The comparison of outcomes can be extended to children whose age is further away from the cutoff by using age as a statistical control on the assumption that only age has differentiated the two cohorts. To further match the children, statistical controls with demographic and other relevant variables are often used. Of importance in this context, well-executed RD designs generally yield intervention effect estimates with relatively high internal validity (Cook and Wong, 2008).

The results of the RD component augment the findings of the intensive substudy and full RCT components of the Tennessee study in three ways that motivate the analyses and interpretations reported here.

First, the RD results contribute estimates of school readiness effects on widely used measures for another state with a state-funded pre-k program. State programs vary on many dimensions and characterizing their effectiveness requires, at minimum, a body of research in which many state programs are represented. The age-cutoff RD design is uniquely appropriate for building that body of research. It is widely applicable to state programs, virtually all of which have an age eligibility cutoff, and has in fact already been used in quite a number of studies of such programs (citations and a summary of findings are in Appendices G-1 and G-2). These studies not only have a common design, but most use the same or very similar literacy, language, and math outcome measures. The present report adds Tennessee results to this growing body of RD evidence about the school readiness effects of state pre-k programs.

Second, recall that the intensive substudy found positive effects on cognitive and achievement outcomes at the end of pre-k. Those are important findings both in the context of research on the immediate effects of state pre-k programs and the objective of such programs to enhance school readiness for kindergarten. But the intensive substudy sample was limited to children whose parents consented to their participation and, further, as a subsample of the RCT, to applicants to oversubscribed program sites. These limitations raise questions about external validity—the extent to which the findings of the intensive substudy for immediate pre-k effects generalize to the statewide population of TN-VPK program sites. The RD component of the Tennessee study addresses this issue by using a statewide probability sample of that population.

Third, by virtue of the comparability of the pre-k effects across states found in age-cutoff RD studies, the Tennessee RD results bear on the null and unexpectedly negative effects found on the achievement outcomes in the years after participation in TN-VPK. Among the attempts to explain those effects by pre-k advocates is the claim that the Tennessee program is of distinctly poorer quality than those of states that have shown longer-term positive results in their studies (e.g., Meloy, Gardner, and Darling-Hammond, 2019). There is no universally recognized measure of pre-k quality, and none that has been widely applied across state pre-k programs in a way that allows meaningful comparison. Arguably, however, a direct measure of quality is the magnitude of the program's effects on representative school readiness outcomes. The general comparability of effects found in age-cutoff RD studies on such outcomes across state programs provides an opportunity for a comparative assessment of the Tennessee program on at least this one indicator of program quality.

The Tennessee Prekindergarten Program

Launched in 2005, TN-VPK is a relatively typical state-funded full-day program that operates on the same calendar as the public schools. By statute, it gives priority to 4-year-old children eligible for the federal free or reduced price lunch programs. Second priority goes to other at-risk children, for example, English language learners, those in state custody, or at risk due to circumstances of neglect or abuse.

The program requires a licensed teacher and an aide in every classroom, a maximum of 20 children per class, and a curriculum chosen from a state-approved list. Contracts from the state support VPK classrooms in the participating school districts. Some districts choose to partner with community agencies and/or Head Start, but the majority of classes are housed in public schools. In 2008-09 when the evaluation reported here first began, nearly every school district in Tennessee offered at least one full-day VPK classroom with a total enrollment statewide of about 18,000 children.

Methods

In the age-cutoff RD design, a pre-specified age determines which children are eligible to enroll for the coming school year. For TN-VPK, children had to be four years old by September 30th of a given year and eligible for kindergarten the following year to enroll. The outcome variables on which VPK participating and nonparticipating children were compared included six tests from the Woodcock-Johnson III Achievement Battery (Woodcock, McGrew, and Mather, 2001). Children in the treatment cohort who had participated in TN-VPK the prior year were assessed shortly after they entered kindergarten. Their counterparts in the control cohort who were not yet four years old by the cutoff date were also assessed at the beginning of the next year, which for them was shortly after they entered TN-VPK and thus before much exposure to the program.

Sample

Children in the treatment and control cohorts came from classrooms chosen via two-stage probability sampling to represent TN-VPK classrooms across the state. With assistance from Tennessee Department of Education (TN-DOE) staff, four geographic regions (West, Central West, Central East, and East) were designated with the school districts and VPK programs within each identified. To be eligible for selection into the sample, a classroom had to have been in place for two successive school years and staffed by the same teacher both years. These criteria were imposed to increase the likelihood that the children in the treatment and control cohorts would be comparable with regard to their community or neighborhood locations and school preferences.

Within each region, VPK classrooms were differentiated by their profiles on four variables related to program administration that were identified by TN-DOE staff as important distinctions between program sites: (1) urban vs. nonurban location; (2) operated in a school or partner community agency; (3) an original pilot program or one added when TN-VPK went to scale; and (4) whether in a high priority school designated as among the lowest performing in the state. Based on these characteristics, the 942 TN-VPK classrooms found statewide were assigned to a specific stratum consisting of some combination of these four program characteristics.

Disproportionate random sampling from these regions and strata was then conducted with sampling fractions created to sample larger proportions of classrooms from smaller strata to ensure that those strata contributed sufficient numbers of classrooms to allow adequate representation in the analysis. Across the four regions, this sampling strategy resulted in a total of 155 classrooms located in 154 schools (one school had two classrooms) spread across 73 school districts and 62 counties. Sampling

weights were assigned that allow the data from the sampled classrooms to be weighted back to the proportions of the full statewide population of classrooms. (Appendix A provides more detail about the sampling procedure). As a practical matter, it was not possible to collect the child-level and classroom-level data required by the design for all the sampled classrooms within the same school year. The data collection therefore was spread over four school years with the sampled classrooms within one of the four regions participating in that process each year (Figure 1 is a schematic of the design).

Once the sample of classrooms was chosen, the children enrolled in those classrooms were identified. The region chosen for the first wave of data collection was Central West. The TN-VPK (treatment) cohort in this region began pre-k in the fall of 2009. After they completed that pre-k year, they were then assessed within the first six weeks of kindergarten the next fall (2010) as long as they were still in the Tennessee public school system (see Figure 1). The corresponding entry years for treatment group pre-k participation in the other regions were fall 2010 (West), fall 2011 (Central East), and fall 2012 (East). The control cohort of children entered those same pre-k classrooms in the fall of the following year—2010 for Central West, 2011 for West, 2012 for Central East, and 2013 for East. Those children were assessed within the first six weeks after they began pre-k, thus during the same school year as the treatment cohort.

To make the two cohorts as comparable as possible, identical eligibility criteria for the treatment and control cohorts were applied to select the children to be included in the analysis. Those criteria require that children be:

- VPK age-eligible; i.e., 4 years old by September 30th of the respective school year.
- Enrolled in a participating VPK classroom within the first six weeks of the system-specific school start date for the respective pre-k year and remain in that

classroom until at least the last six weeks of the system-specific school end date.[1]

- Enrolled in a Tennessee public school kindergarten the year following pre-k within the first six weeks of the system-specific school start date.

These criteria resulted in some children being excluded from the analysis sample because they withdrew from pre-k early, enrolled too late, or moved to a classroom that was not in the sample. Five classrooms in the original sample were dropped: one changed from 4-year-olds to 3-year-olds over the two study years; one was discontinued after the end of the first study year; one was moved to a nonparticipating school; and the teachers in two classrooms moved to a school that served a different population of children in the second study year. Overall, 16% of the children in the originally sampled classrooms became ineligible, including 17% of the treatment and 15% of the control children. Figure 2 diagrams the flow of children through the study.

The final sample consisted of 5,189 children from successive cohorts attending 155 classrooms—2,622 in the treatment cohort and 2,567 in the control cohort. On average, these samples included 17 children from the treatment cohort and 17 children from the control cohort moving through each selected classroom in successive years.

Outcome Measures and Data Collection

A selection of measures from the Woodcock-Johnson III Achievement Battery (McGrew, Schrank, and Woodcock, 2007; Woodcock, McGrew, and Mather, 2001) was used to estimate the effects of TN-VPK on cognitive and achievement skills, as follows:

- Early literacy
    - *Letter-Word Identification*: Ability to identify and pronounce alphabet letters and read words by sight.
    - *Spelling*: Ability to trace letter shapes and write orally presented letters and

words.

- Language
  - *Oral Comprehension*: Ability to listen to an oral passage and provide a missing key word.
  - *Picture Vocabulary*: Extent of expressive vocabulary in response to pictures of objects and actions.
- Early mathematics
  - *Applied Problems*: Ability to apply an appropriate strategy to solve numerical and spatial problems.
  - *Quantitative Concepts*: Ability to understand number identification and sequencing, shapes, and symbols.

An overall index of achievement (WJ Composite) also was constructed by averaging scores across these six scales. All analyses used the Woodcock-Johnson *W* scores, which are equal-interval scores constructed using Rasch modeling to represents both a child's ability and the difficulty of the items. The *W* scores are viewed by the test authors as those most suitable for statistical calculations (Jaffe, 2009).

These measures were administered to the children in the VPK treatment cohort within the first six weeks of kindergarten and, during that same fall, to the children in the control cohort within the first six weeks of pre-k. Trained assessors administered the measures individually to students at their classroom sites in English, the language of instruction for VPK. In addition, demographic information obtained from TN-DOE records included birthday, gender, race/ethnicity, native language, eligibility for free or reduced price lunch (FRPL), and whether the child had an Individual Educational Plan

(IEP) during the pre-k year. There were no missing data in the analytic sample except for FRPL eligibility for one child, which was imputed as eligible in alignment with the status of a large majority of the other children in the sample.

Final Analytic Sample

Of the 5,189 children who met the original sample criteria, 4,144 (79.9%) were FRPL eligible—2,078 in the treatment cohort and 2,066 in the control cohort. The primary analytic sample was restricted to these children for two reasons. First, these economically disadvantaged children constituted the priority target population for TN-VPK. Second, the analytic sample for the RCT component of the overall Tennessee study included only FRPL-eligible children, so this selection facilitated comparison of the results from the RD with those from the RCT. Four classrooms had no FRPL-eligible children, reducing the number of classrooms in the sample from 155 to 151.

Data Analysis

For the RD design to produce valid estimates of treatment effects, at least three conditions must be satisfied (Jacob et al., 2012). First, there must be a clear cut point on the assignment variable, i.e., a distinct value that determines whether a child is in the treatment or control group. Second, there should be no categorical differences between the treatment and control groups across the cut point on any characteristics other than treatment vs. control status that might be related to the outcomes. Finally, the functional relationship between the assignment variable and the outcome variable must be correctly specified. Each of these is examined below for the TN-VPK data.

A Clear Assignment Variable

The exogenous determinant of eligibility for TN-VPK was age—whether a child was 4 years old by September 30th. Even one day on either side of that cutoff was

defined by TN-DOE as sufficient to differentiate eligible from ineligible children at the beginning of the school year. No violations of this requirement were found in the sample data.

Equivalence between Treatment and Control Groups

Our application of the age-cutoff RD design required that children in each of the two cohorts enroll in the same pre-k classrooms with the same teacher in the same schools, as well as in a Tennessee public school kindergarten the next year. This was to increase the probability that children in the two age cohorts would be comparable on characteristics related to the local context of the respective VPK programs. This requirement was adhered to when constructing the analytic sample with one exception—the teacher changed between cohorts prior to outcome measurement in 25 (16.6%) of the 151 classrooms in the analytic sample. There were no other changes in the location and student population and these classrooms were kept in the sample.

To compare the cohorts on their baseline characteristics, each of the available baseline variables was used as the sole dependent variable in multilevel models with different bandwidths (± months around the cutpoint) and with children nested within classrooms and cohort. As Table 1 reports, for the 12-month and 6-month bandwidths, the two cohorts were similar with regard to gender, race/ethnicity, language spoken at home, and IEP placement, but the control cohort did have a slightly larger, and marginally significant proportion of Hispanic children in the 12-month bandwidth (.054 vs. .044).[2] Age, of course, differed as an intrinsic feature of the RD design. The cohorts also differed on the timing of the outcome assessments, with the time between the start of school and administration of the WJ tests averaging a few days longer for the treatment cohort. Children in the control cohort were all located in their known pre-k

classrooms while the kindergarten classrooms the treatment cohort children attended after pre-k had to be located via information from teachers and TN-DOE records.

The 3-month bandwidth sample not only differed on test lag, but also on race/ethnicity and primary language spoken at home. The proportion of Black children was larger in the treatment cohort whereas the proportions of Hispanic children and those from non-English speaking households were larger in the control cohort. All the variables shown in Table 1 were included as covariates in the analyses of VPK effects to adjust for any differences between the cohorts on these variables.

Specification of the Relationship between Age and Outcome

To determine the appropriate functional form, the approach outlined by Jacob et al. (2012) was used. First, graphs of the relationship between scores on each outcome measure and age (days before and after the cutoff date) were examined for discontinuity at the cut point and any other place (Lee and Lemieux, 2010). Second, a variety of functional forms (e.g., linear, cubic, linear interactions) for the relationships between the assignment variable (age) and the outcome variables were tested. Following Lee and Lemieux (2010) and Jacob et al. (2012), two regressions were run for each outcome. The first regressed the outcome on treatment condition, the assignment variable, and the respective term(s) for the functional form being tested. The second used these variables plus indicator variables for the intervals used in the graphical displays. The F-test procedure from Jacob et al. (2012) was used to compare the two models with results that were not statistically significant, indicating no unexplained variability in the simple model that was captured by the more complex model. These explorations established that simple linear models for the age assignment variable predicting each of the outcome measures were adequate to account for their relationships.

Estimation of TN-VPK Effects

Analyses were conducted via SAS Proc Mixed with weighted multilevel regression models that had children nested in classrooms. The weighting function was based on the sampling fractions from the stratified sample and adjusted the classroom representation to the statewide proportions. The formal representation of the analytic model is $Y_{ij} = \beta_0 + \beta_1 A_{ij} + \beta_2 T_{ij} + \sum_{k=1}^{n} \beta_k X_{kij} + u_j + r_{ij}$ where Y_{ij} is an outcome variable measured for child i in classroom j; A_{ij} is age at the cutoff date in days before (minus) and after (plus); T_{ij} is the treatment variable (cohort 1 vs cohort 2); X_{kij} are covariates, u_j is the Level 2 (classroom) residual, and r_{ij} is the Level 1 (child) residual. The covariates are shown in Table 1 (Appendix B describes how these variables were coded). These analyses were run for ±12 month, ±6 month, and ±3 month bandwidths around the cutoff date. As a robustness check, the effects were re-estimated after excluding the outermost 1%, 5%, and 10% of the data points with the highest and lowest age values. If the functional form was not misspecified, the impact estimates should not markedly change with these outermost data points dropped (Jacob et al., 2012).

Effects for Subgroups

Of particular interest were differences in the effects on any outcome measures related to gender, race/ethnicity, or primary language. To investigate such differential effects, the multilevel models were repeated for the Woodcock Johnson composite measure while including the respective subgroup membership variable and the term for its interaction with the treatment condition.

Results

Effects on Literacy, Language, and Math Skills

Figures 3-9 show graphs of the relationships between the observed outcome

measures and the age span before and after the pre-k eligibility cut point for the FRPL analytic sample. These data are not adjusted for covariate influence nor weighted to represent the statewide population of VPK classrooms. For each outcome variable, these plots show the mean of the outcome at the midpoint of equal-sized intervals on the age-based assignment variable. The interval size was determined via the statistical tests recommended by Jacobs and colleagues (2012). The assignment variable represents the number of days from when the child turned 4 years old, centered on the cutoff of September 30th for the respective year. The general linearity of the age-outcome relationships is evident in these graphs as well as discontinuities at the cut point for all the outcome measures except Oral Comprehension.

Table 2 summarizes the results from the weighted multilevel analysis models fit to these data (full results in Appendix Tables C-1 to C-7). Participation in TN-VPK showed statistically significant effects at all three bandwidths for the WJ composite and all the individual Woodcock-Johnson tests except the 3-month bandwidth for Picture Vocabulary. The effect sizes for the composite measure were relatively large—.85, .83, and .91 for the 12-, 6-, and 3-month bandwidths respectively. The largest effects were found for early literacy (Letter-Word and Spelling; effect sizes from .94 to 1.11). The smallest effects were for language skills (Oral Comprehension and Picture Vocabulary; effect sizes from .22 to .34). The effect sizes for mathematics were in between (Applied Problems and Quantitative Concepts, effect sizes from .47 to .71.). Across the bandwidths, the effect size estimates for each measure were generally similar. Notably, those for the 3-month bandwidth, where the estimates should be best, were comparable to or larger than those for the wider bandwidths except for Picture Vocabulary.

As a further robustness check, each of these analyses was repeated with the

outermost 1%, 5%, and 10% of the youngest and oldest children excluded. The resulting effect sizes did not vary greatly from those for the untrimmed samples (Appendix Table D-1). The average difference for the most trimmed sample was .019 for the 12-month bandwidth (ranging from .004 to .038), .078 for the 6-month bandwidth (.041 to .118), and .084 for the 3-month bandwidth (.033 to .131).

Effects for Subgroups

Differential effects were estimated with the same models used for the main effects analyses, but with interaction terms for the treatment condition and the respective subgroup of interest. Statistically significant differential effects were found for only two of the subgroup comparisons. As Table 3 reports, the effects of TN-VPK on the overall WJ composite measure were larger for Hispanic children and for children whose primary language was not English (Appendix Tables E-1 to E-4 report results on the individual tests for the different subgroups). Because the majority (76%) of children whose primary language was not English were also Hispanic, these comparisons overlap. To better disentangle these two variables, the analyses were repeated with the addition of the three-way interactions that differentiated Hispanic and non-Hispanic children according to their native language. These interactions were statistically significant for the 12-month and 6-month bandwidths, though not for the 3-month one (reported in the lower rows of Table 3 for the WJ Composite; results for the individual tests in Appendix E-5). The patterns across the marginal means show that, although the effects of TN-VPK were positive and relatively large in magnitude for all subgroups, the smallest impact appeared for children who were not Hispanic and whose primary language was English. The largest impact was for the small group of children whose primary language was not English but who were not Hispanic.

Effects for Non-FRPL Eligible Children

As described earlier, the analytic sample for the results reported above includes only children eligible for the federal free or reduced price lunch programs (FRPL). This departs from some previous pre-k RD studies that included children regardless of FRPL status. There were some differences between FRPL-eligible and non-eligible children in the overall sample (Table 4). FRPL-eligible children were more likely to be Black or Hispanic and have a non-English primary language. Although these differences were relatively small, further analyses explored their influence on the VPK effects. Multilevel models were first estimated for the full sample with FRPL eligibility as a covariate. Table 5 summarizes the results (full model results are in Appendix Tables F-1 to F-7), which were quite similar to those obtained for only FRPL-eligible children. Statistically significant differences favoring VPK were found regardless of bandwidth or outcome measure with the exception of Picture Vocabulary with the 3-month bandwidth

A second set of models added the interaction of FRPL eligibility and treatment condition with results reported in Table 6. For the 12-month bandwidth, those interactions showed that the VPK effects were significantly larger for FRPL-eligible children on the WJ composite and the individual tests for literacy and math skills. However, no differences reached statistical significance in the analyses with the narrower bandwidths.

Discussion and Conclusions

A strength of the age-cutoff RD design is the relative ease of implementing it in pre-k settings, but it does have limitations. Most notably, this design is essentially a wait-list control that limits assessment of pre-k effects to those appearing shortly after the end of the pre-k year, so the question of sustained effects cannot be investigated.

Further, although regression-discontinuity designs in general are capable of yielding unbiased causal impact estimates, a host of methodological criteria must be satisfied for this to be accomplished. In addition, the pre-k age-cutoff version of the RD design has some distinctive characteristics that impose additional demands (Lipsey, Weiland, et al., 2015). These stem primarily from the comparison of two discrete cohorts separated by time as well as age. This circumstance opens the door to cohort differences that may bias the effect estimates, most obviously the possibility of changes in the demographic mix of the children who enroll in pre-k in successive years. Other differences may also be in play, for example, the consistency with which outcome measures are operationalized and administered across the two cohorts, one tested in pre-k, the other in kindergarten, and the possibility of differential attrition in obtaining those measures.

In this RD study, we attempted to address these issues to the extent possible within the practical limitations of field-based research. Meeting the widely recognized formal criteria for regression-discontinuity designs was relatively straightforward. The strict age cutoff for VPK eligibility provided a well-defined and exogenously imposed cutpoint for differentiating pre-k participants from nonparticipants. A thorough exploration of the functional form of the relationship between age and the outcome measures identified a simple linear model as a good fit for all outcomes. The effect estimates were repeated with different bandwidths around the cutpoint and varying amounts of data trimming at the extremes of the age range to assess their robustness and neither showed enough variation to call the overall findings into question.

More challenging was the need to ensure the equivalence of the children compared across the cohorts on characteristics potentially related to the outcomes of interest. We were limited to the small set of baseline demographic variables available in

the state data system, but included those as covariates in the analytic models to statistically adjust for any differences between cohorts on those variables. In addition, we tried to maximize the comparability of the treatment and control cohorts by imposing identical study eligibility criteria. Children were only included in the analytic sample if they were in the same VPK classrooms with the same teacher the year before for the treatment cohort and at the beginning of the current year for the control cohort, save for a few exceptions that arose for practical reasons. Moreover, the children in both cohorts had to have been enrolled within the beginning weeks of their pre-k year, remain enrolled through nearly the end of that school year, and attend a Tennessee public school the year after their pre-k year. The administration of the outcome measures was made as systematic and consistent across cohorts as possible and the elapsed time between the start of the school year and testing was used as a covariate in the analyses. Despite these efforts, a threat to the internal validity of the effect estimates remains in the form of unobserved and uncontrolled differences between the cohorts that would be capable of biasing those estimates.

Immediate Effects of TN-VPK

The findings reported here add to the accumulating body of research on the effects of state pre-k programs on cognitive skills and early educational achievement measured shortly after pre-k participation. Overall, that research has shown almost universally positive effects (Duncan and Magnuson, 2013) with age-cutoff RD designs especially common. The results of the present study fall in line with that overall pattern. The effect size estimates on the WJ Composite measure that averaged scores from the individual tests ranged from .83 to .91 across the different bandwidths. Effect sizes in standard deviation units can be easily translated into percentile differences when the

outcome data are normally distributed, as these are. With the mean of the control group score set at the 50th percentile, an effect size of .85, the middle value across the bandwidths, places the mean for TN-VPK participants at the 80th percentile. This represents a rather considerable relative improvement in the measured cognitive skills of the children who participated in TN-VPK program.

The TN-VPK effects, however, varied across the tests measuring literacy, language, and mathematics skills. The largest effects were found for literacy (WJ Letter Word and Spelling tests). The smallest effects by a substantial margin were found on the language measures (WJ Oral Comprehension and Picture Vocabulary). The effect sizes for the mathematics measures (WJ Applied Problems and Quantitative Concepts) fell in between. In addition, moderator analysis revealed that VPK effects were larger for Hispanic children than non-Hispanic children, and for children whose primary language was not English. The largest effects were found for the relatively small subgroup of children who were not Hispanic but also did not have English as their primary language. The smallest, but still positive effects were found for non-Hispanic native English speaking children.

While the focus of this study was on children who qualified for free or reduced price lunch and were thus the priority target group for TN-VPK, the full sample of both FRPL and non-FRPL eligible children included about 25% more children than the FRPL-only sample. Nonetheless, the pattern of results was similar for that more inclusive sample, although the magnitude of the effects was somewhat smaller than for the FRPL-only sample. Taken altogether, the subgroup analyses indicated that VPK effects were larger for children from low-income families (FRPL-eligible) and notably larger for Hispanic children and the overlapping group of children for whom English

was not their native language, whether Hispanic or not.

Comparison with the Effects Found in the Parallel RCT

As noted earlier, the RD design was one component of the larger Tennessee Pre-K Study that also included a randomized control trial of outcomes for FRPL-eligible applicants offered admission to oversubscribed VPK program sites with those for children waitlisted and ultimately not offered admission. One part of that RCT, the intensive substudy (ISS), used the same Woodcock Johnson achievement measures as the RD study with a sample of consented children tested at the beginning and end of the pre-k year. The ISS was analyzed as a quasi-experiment, but had the advantage of an extensive set of baseline measures used as covariates to adjust for the few initial differences between the treatment and control groups. However, as a consented subsample of applicants to oversubscribed program sites, the external validity of those results is uncertain, i.e., their generalization to the statewide population of VPK program sites.

An important feature of the RD component of the Tennessee study is that it is based on a probability sample of the full population of TN-VPK program sites as it existed at the time the study began. In addition, the same Woodcock-Johnson outcome measures used in the ISS were used in the RD study. These circumstances make it possible to directly compare the effect estimates from the two study components. A similar pattern of positive results would be indicative of the external validity of the ISS estimates and any such convergence across the two distinct study components would also be an indication of the robustness of that pattern.

The effect sizes on each of the WJ measures common to the ISS and the RD are compared in Table 7, with the 12-month bandwidth shown for the RD. Because the RD

samples represent children who actually participated in VPK, the ISS effect estimates are from the treatment-on-the-treated analysis that defines treatment and control groups in terms of actual participation irrespective of the condition to which the children were originally assigned. Both the RD and ISS samples showed positive VPK effects on all the measures, although that for Oral Comprehension in the ISS is quite small and not statistically significant. Moreover, the effect estimates in both samples are largest for the literacy outcomes with those for the language outcomes among the smallest, and the math outcomes in between. There is thus notable mutual confirmation of the generally positive VPK effects and their pattern across outcome measures between these two study components.

At the same time, there is a rather large difference between the RD and the ISS results in the magnitude of the effect estimates. The RD effect sizes are larger on every measure—more than twice as large for the WJ Composite and some of the individual tests. It is difficult to interpret that contrast given the differences between the sources of the respective estimates. The RD used a probability sample of statewide TN-VPK programs whereas the ISS used a consented subsample of the overall RCT. Also, outcome assessment in the RD occurred at the beginning of the kindergarten year, approximately 12 months after initial pre-k enrollment, but at the end of the pre-k school year in the ISS, approximately 9 months after initial enrollment. Moreover, outcome measurement for the control group in the RD came at the beginning of the pre-k year, making the prior 3-year old period the counterfactual condition. In contrast, outcome measurement for the ISS control group at the end of the pre-k school year made the 4-year old period the counterfactual condition. The implications of these and other relevant differences between the RD and ISS for the magnitude of the respective

effect size estimates are being explored in a separate report.

TN-VPK Effects Compared with Those from Other Pre-k RD Studies

The findings of this RD study, as well as those from the ISS, demonstrate that TN-VPK has been effective in improving the cognitive skills and early achievement of the low-income students it serves in preparation for their entry into kindergarten. As described in the introduction to this report, however, the longer-term effects found in the ISS and RCT components of the Tennessee study were not so positive—indeed, they diminished sharply after pre-k, becoming null or even somewhat negative. While the pre-k age-cutoff RD design is limited to estimating immediate pre-k effects, its results do bear on the issue of longer-term effects in one important way. In contrast to TN-VPK, studies of the effects of other state pre-k programs have generally found that the positive immediate effects of pre-k were often sustained into the early elementary school years, and even further in some cases (Camilli et al., 2010; Elango et al, 2016).

There are many possible explanations for this contrast, including methodological differences. The longer-term findings in the studies of other state programs are based on nonrandomized designs with only the Tennessee study using a randomized control. One plausible explanation, however, is that the quality of the TN-VPK program is distinctly poorer than that of the pre-k programs in other states. There is no consensus on the definition of pre-k program quality and no measures with strong predictive validity for longer-term effects (Keys et al., 2013; Sabol et al. 2013), so the comparative quality of the Tennessee program is difficult to judge. TN-VPK is substantially similar in structure to most state programs and, at the time this study began, met 9 of the 10 NIEER standards for pre-k programs (Friedman-Krauss et al., 2020).

One revealing way to view the quality of a pre-k program is in terms of its

effectiveness at producing gains over the pre-k year on the cognitive and achievement outcomes widely assessed in pre-k research. The relatively large number of studies of pre-k programs that have used the age-cutoff RD design with some of those widely used outcome measures make it possible to compare their effects. We have identified 18 other pre-k RD studies that can serve as a basis for comparison (details in Appendices G-1 and G-2). These include evaluations of the state-wide programs in Arkansas, California, Georgia, New Mexico, Michigan, New Jersey, North Carolina, Oklahoma, South Carolina, Virginia, and West Virginia; a county-wide program in Michigan; and city-wide programs in Boston, San Francisco, and Tulsa. The studies vary in how the RD design was implemented and analyzed, and in the nature of the programs and samples of children included. Nonetheless, it is informative to examine the distribution of effect sizes found across these similarly designed studies and, in the present context, to observe where the TN-VPK effects fall in that distribution.

Overall, these RD studies show overwhelmingly positive effects for the pre-k programs studied, especially in the commonly measured outcome domains of literacy, language, and math skills. Figure 10 shows the comparison of the VPK effects with those from these other studies for WJ Letter-Word Identification, Figure 11 for PPVT or WJ Picture Vocabulary, and Figure 12 for WJ Applied Problems. The TN-VPK effect sizes shown in these distributions are for the full sample, not the somewhat larger ones for the FRPL-eligible sample that was the focus of this study. Not all the programs represented in these distributions prioritize FRPL-eligible children like TN-VPK does and, indeed, some are universal programs open to all age-eligible children.

Figures 10-12 reveal the generally positive effects on these measures found for all the pre-k programs studied with the age-cutoff RD design. While acknowledging the

methodological and programmatic variation across these studies, it is notable that the TN-VPK effects compare relatively well. TN-VPK is not at the very top of any of these distributions, but it is in the top half in all three, indicating what can be described as an above average performance compared to the peer programs represented. TN-VPK, therefore, cannot be easily discounted as an inferior program despite the disappointing longer-term effects found in the randomized components of the Tennessee study that were able to assess the extent to which the pre-k gains were sustained.

What Happened After Pre-k?

Followed over time in the other components of the Tennessee study, the VPK effects on literacy, language, and math disappeared by the end of kindergarten and turned negative by the end of third grade (Lipsey, Farran, and Durkin, 2018). A common pattern is for pre-k programs to show initial positive effects that then fade out some time afterwards (Bailey, Duncan, Odgers, and Wu, 2017; Bailey et al., 2020). For TN-VPK, however, the pattern is better described as "catch up." The children in the control group showed greater gains in kindergarten than the TN-VPK participants such that their scores on the early achievement measures converged on those of the participants. One possible contributor to this pattern may be the content focus of the instruction the children received in pre-k. The skills emphasized in pre-k generally include directly teachable skills in a finite domain, e.g., letters of the alphabet, geometric shapes, and cardinal numbers. These have been referred to as "constrained skills" (Snow and Matthews, 2016) that a learner can rather completely master. They are also the skills heavily represented in the age appropriate items of the typical pre-k outcome measures, especially the literacy and math measures, somewhat less so in the language measures. It is notable in this regard that the RD results in Figures 10-12 rather

consistently show the largest effects on literacy, smaller ones on math, and the smallest ones on the language measures. Given the nature of these constrained skills, which are also emphasized in kindergarten, it is not surprising that children show similar mastery after the kindergarten year whether or not they participated in a pre-k program.

Continued emphasis on constrained skills in kindergarten that are redundant with what children experienced in pre-k (Engel, Claessens, and Finch, 2013; Claessens, Engel, and Curran, 2014) risks boredom and disengagement by those students. For the TN-VPK randomized sample, we see some indication of this in teacher ratings of feelings about school that are significantly lower for the VPK participants than the nonparticipants at the end of first grade. This is a possible contributor to the slide into negative effects in the later grades for these students.

Unconstrained skills such as vocabulary, comprehension, and mathematical reasoning, by contrast, receive less attention in pre-k and kindergarten but over time become increasingly important for academic performance (Snow and Matthews, 2016). However, the modest enhancement of those skills found for TN-VPK on the language and math measures would only facilitate positive achievement gains in the later grades if they were supported and sustained in those later grades. For this to happen, the students must experience high quality learning environments that provide such support.

Pearman et al. (2020) examined this "sustaining environments" hypothesis (Bailey et al., 2017) for the TN-VPK randomized sample by investigating differential achievement effects for students attending higher quality elementary schools and/or being exposed to higher quality teachers between kindergarten and third grade. Using Tennessee school and teacher evaluation data, they found that VPK participants enrolled in higher quality elementary schools *and* exposed to a succession of especially

effective teachers showed positive effects on the state language arts and math achievement tests relative to their counterparts in the control group. Conversely, the VPK participants without such sustaining environments had lower scores than their counterparts on those achievement tests, i.e., showed the negative effects that appeared for the overall sample. The predominance of those negative effects follows from the fact that, sadly, only 12% of these low income students in the overall sample experienced both high quality schools and highly effective teachers. Here also we have a possible clue to the source of the unexpectedly negative longer-term effects of VPK participation.

Conclusions

What is most clearly demonstrated in this RD study is that the TN-VPK program, viewed statewide, has positive effects on cognitive skills that are generally viewed as supporting school readiness. Moreover, the effects compare favorably with those found in pre-k RD studies of programs in other states and localities. These findings do not support the view that the longer term negative effects on achievement found in the randomized component of the Tennessee study stem from a distinctively poor quality pre-k program. We can only speculate about the nature of the interaction between the TN-VPK experience and the experience those children had in later grades that resulted in the longer term negative effects. The evidence available to date highlights redundant content in kindergarten that diminishes positive feelings about school combined with the lower quality schools and teachers most of these students experience in the later grades as likely factors.

Notes

1. In 5 instances, a student who was enrolled in one participating TN-VPK classroom withdrew within the first six weeks of school and enrolled in another participating TN-VPK classroom; those students were matched with the latter classroom.
2. The race/ethnicity categories for this sample are not mutually exclusive; 10% of the children were identified in school records as mixed race with less than 1% something other than White, Black, Hispanic, or some mix of those. The mixed race children are coded in all the categories for their respective mix. In the unadjusted data that coding identifies 64.2% as White, 36.1% as Black, and 8.5% as Hispanic.

References

Bailey, D. H., Duncan, G. J., Odgers, C., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*, 7–39.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21(2)*, 55-97.

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

Chin, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19, 3127-3131.

Claessens, A., Engel, M., & Curran, F. C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, *51(2)*, 403-434.

Cook, T. D, & Wong, V. C. (2008). Empirical tests of the validity of the regression discontinuity design. *Annals of Economics and Statistics, Econometric Evaluation of Public Policies: Methods and Applications*, *91/92*, 127-150.

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, *27(2)*, 109-132.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2016). Early childhood education. In Moffitt, R. A., (ed), *Economics of means-tested transfer programs in the United States* (v. 2, 235-297). University of Chicago Press.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already

know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35(2)*, 157-178.

Farran, D. C., & Lipsey, M. W. (2016). Evidence for the benefits of state prekindergarten programs: Myth & misrepresentation. *Behavioral Science & Policy*, *2(1)*, 9-18.

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G. & Gardiner, B. A. (2020). *The State of Preschool 2019: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, *41(6)*, 872-884.

Jacob, R., Zhu, P. Somers, M-A., & Bloom, H. (2012). *A practical guide to regression discontinuity*. New York, NY: MDRC.

Jaffe, L. E. (2009). *Development, interpretation, and application of the W score and the relative proficiency index* (Woodcock-Johnson III Assessment Service Bulletin No. 11). Rolling Meadows, IL: Riverside Publishing.

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., Ruzek, E. A., & Howes, C. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development 84(4)*, 1171–1190.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48(2)*, 281-355.

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

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

McGrew, K. S, Schrank, K. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III normative update technical manual*. Rolling Meadows, IL: Riverside.

Meloy, B., Gardner, M., & Darling-Hammond, L. (2019). *Untangling the evidence on preschool effectiveness: Insights for policymakers*. Palo Alto, CA: Learning Policy Institute.

Pearman II, F. A., Springer, M. P., Lipsey, M., Lachowicz, M., Swain, W., & Farran, D. (2020). Teachers, schools, and pre-k effect persistence: An examination of the sustaining environment hypothesis. *Journal of Research on Educational Effectiveness, 13(4)*, 547-573.

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., Duncan, G. J., Dynarski, M., Magnuson, K. A., & Weiland, C. (2017). *Puzzling It Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects: A Consensus Statement*. Washington, DC: Brookings Institution, 2017.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, D., et al. (2012). *Third grade follow-up to the Head Start Impact study final report*, OPRE Report # 2012–45. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families U.S. Department of Health and Human Services.

Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-k programs predict children's learning? *Science, 341(23)*, 845-846.

Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The*

Future of Children, 26(2), 57-74.

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

Table 1
Comparison of TN-VPK and No Pre-K Cohorts on Baseline Characteristics: 12-, 6-, and 3-Month Bandwidth Samples

Variable and Bandwidth	Means[a]		Odds Ratio	TN-VPK vs. No Pre-K Difference (Pooled SD)	*p*-value	Effect Size[b]
	TN-VPK	No Pre-K				
12-month bandwidth	(*N*=2078)	(*N*=2066)				
Gender (Male)	.495	.494	1.01		.915	.004
White	.788	.809	.88		.220	-.073
Black	.216	.201	1.10		.368	.052
Hispanic	.044	.054	.80		.063	-.127
Native language not English	.032	.027	1.19		.221	.095
Has an IEP	.090	.084	1.07		.512	.039
Age at testing (months)	65.5	53.4		12.08 (6.93)	.001	1.743
Test lag (days)	35.2	32.8		2.41 (5.72)	.001	.421
6-month bandwidth	(*N*=1099)	(*N*=1058)				
Gender (Male)	.493	.477	1.07		.462	.037
White	.782	.810	.84		.262	-.096
Black	.236	.196	1.27		.110	.132
Hispanic	.047	.056	.83		.326	-.103
Native language not English	.038	.043	.89		.560	-.064
Has an IEP	.096	.078	1.26		.144	.128
Age at testing (months)	62.4	56.4		5.97 (3.42)	.001	1.747
Test lag (days)	35.2	33.1		2.12 (5.48)	.001	.387
3-month bandwidth	(*N*=519)	(*N*=531)				
Gender (Male)	.484	.496	.95		.695	-.028
White	.726	.783	.73		.114	-.174
Black	.292	.216	1.50		.037	.224
Hispanic	.037	.074	.49		.010	-.394
Native language not English	.032	.060	.51		.024	-.372
Has an IEP	.091	.089	1.03		.882	.016
Age at testing (months)	60.9	57.8		3.10 (1.75)	.001	1.771
Test lag (days)	36.2	32.3		3.97 (5.32)	.001	.746

Note. Only children identified as eligible for free or reduced lunch were included in the analysis. Analyses were weighted to project the estimates to the statewide population of VPK classrooms. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing.

[a] Estimated means from the multilevel models with children nested within classrooms; membership in the treatment group as the only predictor with the respective baseline variable as the dependent variable.

[b] For binary variables, odds ratios were converted into effect sizes, using Chin's (2000) method. For continuous variables effect sizes were calculated by dividing the TN-VPK vs. No Pre-K difference by the pooled standard deviation of the two groups.

Table 2
Estimates of the Pre-K Effect for the WJ Composite and Individual Tests

Outcome and Bandwidth	Means[a]		TN-VPK vs. No Pre-K Difference[b]	*p*-value	TN-VPK Standard Deviation	Effect Size[c]
	TN-VPK	No Pre-K				
WJ Composite						
± 12 months	411.9	401.0	10.93	.001	12.92	.85
± 6 months	412.0	400.2	10.74	.001	12.66	.83
± 3 months	411.9	400.8	10.84	.001	11.94	.91
Letter Word						
± 12 months	346.7	324.7	22.02	.001	21.15	1.04
± 6 months	346.8	325.0	21.83	.001	19.71	1.11
± 3 months	345.4	325.6	19.85	.001	19.17	1.04
Spelling						
± 12 months	379.5	359.2	20.35	.001	20.82	.98
± 6 months	379.7	360.4	19.28	.001	20.62	.94
± 3 months	379.3	359.5	19.76	.001	20.43	.97
Oral Comprehension						
± 12 months	451.5	448.3	3.25	.001	13.71	.24
± 6 months	451.3	448.4	2.91	.009	13.08	.22
± 3 months	452.1	448.1	4.03	.012	12.87	.31
Picture Vocabulary						
± 12 months	464.8	461.1	3.69	.001	10.72	.34
± 6 months	464.9	461.2	3.69	.002	10.89	.34
± 3 months	464.5	461.8	2.69	.107	10.41	.26
Applied Problems						
± 12 months	408.0	399.7	8.37	.001	17.74	.47
± 6 months	409.4	399.6	9.80	.001	18.12	.54
± 3 months	409.8	398.4	11.36	.001	16.08	.71
Quantitative Concepts						
± 12 months	420.7	412.8	7.88	.001	13.99	.56
± 6 months	419.9	412.8	7.08	.001	13.43	.53
± 3 months	419.7	412.3	7.45	.001	12.49	.60

Note. The sample sizes for the TN-VPK and No Pre-K groups were: 2,078 and 2,066 for ±12 months; 1009 and 1,058 for ±6 months; and 519 and 531 for ±3 months. Only children eligible for free or reduced price lunch were included. All analyses were weighted to project the estimates to the statewide population of TN-VPK classrooms.

[a] Estimated marginal means from the multilevel analysis model.

[b] Estimates based on multilevel models with children nested within pre-k classrooms. In addition to condition and days from the age cutoff (centered at zero), covariates included: (1) region; (2) whether the child was male; (3) whether the child was Black; (4) whether the child was Hispanic; (5) whether the child's native language was not English; (6) whether the child had a special education placement (IEP); and (7) the number of days between the WJ testing date and the start of school (centered at the grand mean). Previous analyses showed that the functional form for all models was linear. Appendix Tables C1 to C7 present the results for the full models for each outcome and bandwidth.

[c] Standardized mean difference effect sizes calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the TN-VPK treatment group, recognizing that group as more analogous to the posttest outcomes typically used to compute effect sizes.

Table 3

TN-VPK Effect Estimates on the WJ Composite for Hispanic and Native Language Subgroups of Children

Subgroup and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
12-month bandwidth								
Ethnicity						<.001	12.92	
Hispanic	406.5	168	386.8	186	19.72			1.53
Not Hispanic	412.4	1910	402.3	1880	10.15			.79
Native language						<.001		
English	412.0	1914	401.9	1926	10.07			.78
Not English	410.1	164	389.0	140	21.12			1.64
Ethnicity and native language						<.001		
Hispanic and English	408.5	49	387.3	72	21.18			1.64
Hispanic and not English	398.8	119	379.7	114	19.10			1.48
Not Hispanic and English	412.8	1865	403.0	1854	9.81			.76
Not Hispanic and not English	413.8	45	384.7	26	29.12			2.25
6-month bandwidth								
Ethnicity						<.001	12.66	
Hispanic	405.1	78	384.8	87	20.31			1.60
Not Hispanic	412.6	931	402.7	971	9.97			.79
Native language						<.001		
English	412.0	936	402.1	980	9.90			.78
Not English	411.3	73	390.3	78	21.02			1.66
Ethnicity and native language						<.001		
Hispanic and English	407.8	25	381.9	27	25.89			2.05
Hispanic and not English	397.7	53	379.9	60	17.81			1.41
Not Hispanic and English	412.9	911	403.4	953	9.53			.75
Not Hispanic and not English	416.3	20	384.6	18	31.70			2.50

Continued

Table 3 (*continued*)
TN-VPK Effect Estimates on the WJ Composite for Hispanic and Native Language Subgroups of Children

Subgroup and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
3-month bandwidth								
Ethnicity						.007	11.94	
Hispanic	404.1	35	385.0	41	19.08			1.60
Not Hispanic	412.5	484	402.2	490	10.26			.86
Native language						.003		
English	412.0	487	402.0	492	10.01			.84
Not English	409.1	32	389.0	39	20.13			1.69
Ethnicity and native language						.440		
Hispanic and English	401.0	11	385.0	12	15.99			1.34
Hispanic and not English	396.7	24	376.8	29	19.90			1.67
Not Hispanic and English	413.0	476	403.1	480	9.92			.83
Not Hispanic and not English	408.2	8	386.6	10	21.52			1.80

Note. Only children eligible for free or reduced price lunch were included. Previous analyses showed that the functional form was linear. All analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Marginal means from the multilevel model.

[b] To facilitate comparison of these effect sizes with those for the overall VPK effects, they were calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the treatment (TN-VPK) group (see Table 2).

[c] For the two-way interactions these are the *p*-values for the interaction terms for ethnicity by treatment condition and native language by treatment condition. For the three-way interactions, these are the *p*-values for the interaction terms for ethnicity by native language by treatment condition. Estimates were based on multilevel models with children nested within pre-k classrooms. In addition to treatment condition and days from the age cutoff (centered at zero), covariates included: (1) Region; (2) whether the child was Male; (3) whether the child was Black; (4) whether the child was Hispanic; (5) whether the child's native language was not English; (6) whether the child had an IEP placement in the pre-k year; (7) the number of days elapsed between the Woodcock-Johnson testing date and the start of school (centered at the grand mean); and (8) the appropriate interaction term(s). Appendix Tables E-1 to E-5 present the results for the individual WJ tests for all the demographic subgroups.

Table 4
Comparison of Baseline Characteristics by Eligibility for Free or Reduced Price Lunch: 12-, 6-, and 3-Month Bandwidth Samples

Variables and Bandwidth	Means[a]		Odds Ratio	TN-VPK vs. No Pre-K Difference (Pooled SD)	*p*-value	Effect Size[b]
	FRPL Eligible	Not Eligible				
12-month bandwidth	(N=4144)	(N=1044)				
Gender (Black)	.494	.499	.98		.785	-.01
White	.800	.867	.61		.002	-.27
Black	.204	.118	1.92		<.001	.36
Hispanic	.050	.029	1.75		.002	.31
Native language not English	.031	.019	1.67		.016	.28
Has an IEP	.088	.097	.90		.370	-.06
Age at testing (months)	59.4	59.8		-.39 (6.96)	.104	-.06
Test lag (days)	33.9	34.7		-.72 (5.99)	.002	-.12
6-month bandwidth	(N=2067)	(N=535)				
Gender (Black)	.484	.481	1.02		.872	.01
White	.800	.848	.72		.095	-.18
Black	.208	.143	1.57		.024	.25
Hispanic	.052	.033	1.63		.047	.27
Native language not English	.040	.026	1.56		.560	.25
Has an IEP	.087	.103	.83		.254	-.10
Age at testing (months)	59.3	59.5		-.26 (3.45)	.109	-.08
Test lag (days)	33.9	34.6		-.69 (5.81)	.031	-.12
3-month bandwidth	(N=1050)	(N=284)				
Gender (Black)	.489	.450	1.17		.248	.09
White	.756	.844	.57		.025	-.31
Black	.251	.154	1.84		.014	.34
Hispanic	.059	.038	1.56		.162	.25
Native language not English	.047	.031	1.57		.204	.25
Has an IEP	.090	.118	.74		.168	-.17
Age at testing (months)	59.3	59.6		-.25 (1.78)	.047	.14
Test lag (days)	33.9	34.7		-.85 (5.78)	.059	-.15

Note. Children regardless of their free or reduced lunch eligibility were included in the analysis. Analyses were weighted to represent the statewide population of TN-VPK classrooms. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing.

[a]Estimated means from the multilevel models with children nested within classrooms, eligibility for free or reduced price lunch (FRPL) as the only predictor, and the respective baseline variable as the dependent variable.

[b]For binary variables, odds ratios were converted into effect sizes, using Chin's (2000) method. For continuous variables, effect sizes were calculated by dividing the Eligible vs. Not Eligible difference by the pooled standard deviation.

Table 5
Estimates of the Pre-K Effect for the WJ Composite and Individual Tests: Sample of All Children Regardless of Eligibility for Free or Reduced Price Lunch

Outcome and Bandwidth	Means[a]		TN-VPK vs. No Pre-K Difference[b]	*p*-value	TN-VPK Standard Deviation	Effect Size[c]
	TN-VPK	No Pre-K				
WJ Composite						
± 12 months	412.9	402.0	10.90	.001	13.19	.83
± 6 months	413.2	402.4	10.82	.001	12.93	.84
± 3 months	412.9	402.6	10.30	.001	12.21	.84
Letter Word						
± 12 months	347.8	326.2	21.66	.001	21.79	.99
± 6 months	348.0	326.8	21.25	.001	20.69	1.03
± 3 months	346.7	327.6	19.08	.001	20.14	.95
Spelling						
± 12 months	380.6	360.3	20.35	.001	21.06	.97
± 6 months	381.3	361.3	20.01	.001	20.76	.96
± 3 months	380.8	361.0	19.87	.001	20.34	.98
Oral Comprehension						
± 12 months	452.8	448.9	3.86	.001	13.88	.28
± 6 months	452.8	449.3	3.50	.001	13.64	.26
± 3 months	453.3	449.6	3.71	.009	13.28	.28
Picture Vocabulary						
± 12 months	465.3	461.9	3.40	.001	10.73	.32
± 6 months	465.4	462.2	3.26	.002	11.00	.30
± 3 months	465.4	462.8	2.65	.063	10.36	.26
Applied Problems						
± 12 months	409.5	401.0	8.53	.001	18.01	.47
± 6 months	411.0	401.1	9.93	.001	17.98	.55
± 3 months	410.6	401.0	9.52	.001	16.16	.59
Quantitative Concepts						
± 12 months	421.5	413.9	7.55	.001	14.40	.52
± 6 months	420.9	413.9	7.02	.001	13.92	.50
± 3 months	420.5	413.7	6.83	.001	13.25	.52

Note. The sample sizes for the TN-VPK and No Pre-K groups were: 2,621 and 2,567 for ±12 months; 1,280 and 1,322 for ±6 months; and 661 and 673 for ±3 months. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Full results of the multilevel models are reported in Appendix Tables F-1 to F-7.

[a] Estimated marginal means.

[b] Estimates based on a multilevel model with children nested within pre-k classroom. In addition to days from the eligibility cutoff date (centered at zero), covariates included: (1) Region; (2) Male; (3) Black; (4) Hispanic; (5) non-native English; (6) IEP placement; (7) FRPL eligible; and (8) days between the WJ testing and the start of school (grand mean centered).

[c] Effect sizes are the TN-VPK vs. No Pre-K difference divided by the standard deviation of the TN-VPK treatment group.

Table 6
TN-VPK Effect Estimates for WJ Scores by Children's Free or Reduced Lunch Eligibility

Outcome and Sample	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	Pooled TN-VPK Standard Deviations	Effect Size[b]
	Mean	*N*	Mean	*N*				
WJ Composite								
12-month bandwidth						.005	18.93	
Eligible	412.3	2078	400.8	2066	11.52			.61
Not eligible	415.6	543	406.7	501	8.87			.47
6-month bandwidth						.358	16.74	
Eligible	412.1	1009	401.0	1058	11.12			.66
Not eligible	417.6	271	407.6	264	9.96			.60
3 -month bandwidth						.644	16.08	
Eligible	411.4	519	401.3	531	10.11			.63
Not eligible	418.3	142	407.3	142	10.91			.68
Letter-Word								
12-month bandwidth							29.32	
Eligible	346.8	2078	324.3	2066	22.52	.015		.77
Not eligible	352.2	543	333.3	501	18.82			.64
6-month bandwidth						.406	27.15	
Eligible	346.3	1009	324.6	1058	21.67			.80
Not eligible	354.9	271	335.0	264	19.92			.73
3 -month bandwidth						.834	26.54	
Eligible	344.8	419	325.5	531	19.24			.73
Not eligible	353.9	142	335.3	142	18.60			.70
Spelling								
12-month bandwidth						.428	29.96	
Eligible	379.7	2078	359.1	2066	20.60			.69
Not eligible	384.5	543	365.0	501	19.49			.65
6-month bandwidth						.708	26.45	
Eligible	379.9	1009	359.7	1058	20.18			.76
Not eligible	386.6	271	367.2	264	19.43			.74
3 -month bandwidth						.253	25.24	
Eligible	378.7	519	359.6	531	19.10			.76
Not eligible	388.6	142	366.2	142	22.36			.89
Oral Comprehension								
12-month bandwidth							15.67	
Eligible	452.2	2078	448.3	2066	3.89	.887		.25
Not eligible	455.1	543	451.4	501	3.77			.24
6-month bandwidth						.505	14.57	
Eligible	451.8	1009	448.5	1058	3.31			.23
Not eligible	456.6	271	452.5	264	4.12			.28
3 -month bandwidth						.092	14.32	
Eligible	452.0	419	448.9	531	3.04			.21
Not eligible	458.1	142	452.2	142	5.90			.41

Continued

Table 6 (*continued*)
TN-VPK Effect Estimates for WJ Scores by Children's Free or Reduced Lunch Eligibility

Outcome and Sample	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	Pooled TN-VPK Standard Deviations	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Picture Vocabulary								
12-month bandwidth						.004	15.21	
Eligible	465.1	2078	461.1	2066	3.98			.26
Not eligible	466.3	543	464.8	501	1.51			.10
6-month bandwidth						.225	14.81	
Eligible	465.0	1009	461.4	1058	3.61			.24
Not eligible	467.3	271	465.2	264	2.13			.14
3 -month bandwidth						.421	14.92	
Eligible	464.8	519	461.8	531	2.97			.20
Not eligible	467.8	142	466.2	142	1.60			.11
Applied Problems								
12-month bandwidth							26.04	
Eligible	409.2	2078	399.3	2066	9.97	<.001		.38
Not eligible	411.2	543	407.4	501	3.81			.15
6-month bandwidth						.215	22.86	
Eligible	410.0	1009	399.5	1058	10.48			.46
Not eligible	415.1	271	406.9	264	8.14			.36
3 -month bandwidth						.747	22.29	
Eligible	409.0	519	399.7	531	9.32			.42
Not eligible	416.3	142	406.1	142	10.15			.46
Quantitative Concepts								
12-month bandwidth						.007	17.27	
Eligible	420.9	2078	412.8	2066	8.11			.47
Not eligible	423.9	543	418.1	501	5.73			.33
6-month bandwidth						.296	15.46	
Eligible	420.0	1009	412.6	1058	7.33			.47
Not eligible	424.7	271	418.7	264	6.01			.39
3 -month bandwidth						.956	14.91	
Eligible	419.4	519	412.6	531	6.80			.46
Not eligible	424.8	142	417.9	142	6.90			.46

Note. Children regardless of their free or reduced lunch eligibility were included in the analysis. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Estimated marginal means from the multilevel analysis model.

[b] Effect sizes are the TN-VPK vs. No Pre-K difference divided by the standard deviations of the TN-VPK treatment groups pooled for the Eligible and Not Eligible subgroups.

[c] *p*-value for the interaction term for condition (TN-VPK vs. No Pre-K) and FRPL eligibility. Estimates were based on a multilevel model with children nested within pre-k classroom. In addition to days from the eligibility cutoff date (centered at zero), covariates included: (1) Region; (2) Male; (3) Black; (4) Hispanic; (5) non-native English; (6) IEP placement; (7) FRPL eligible; and (8) days between the WJ testing and the start of school (grand mean centered); and (9) the interaction term.

Table 7
Effect Sizes from the RDD Compared with Those from the ISS

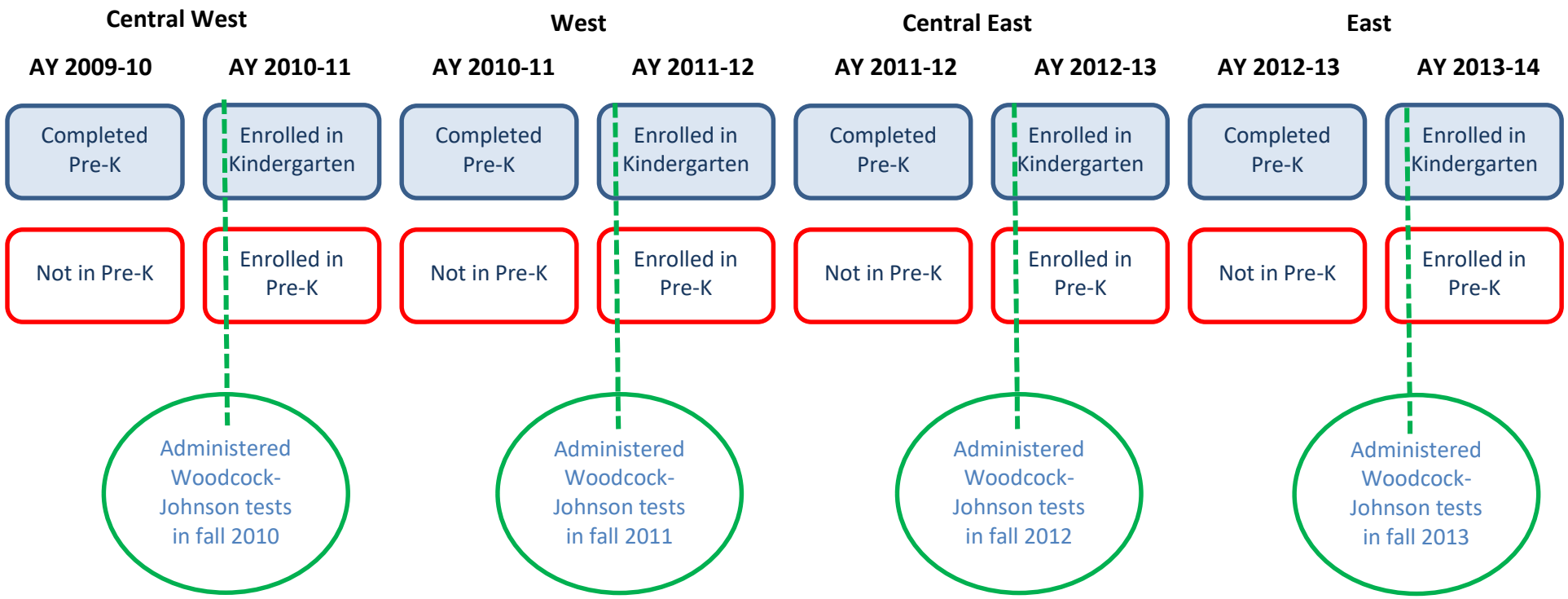
Outcome Measure	Effect Size Estimates	
	RDD[a]	ISS[b]
WJ Composite	.85	.40
Literacy		
Letter-Word Identification	1.04	.47
Spelling	.98	.38
Language		
Oral Comprehension	.24	.07
Picture Vocabulary	.34	.32
Mathematics		
Applied Problems	.47	.26
Quantitative Concepts	.56	.33

Notes: Estimates for both the RDD and ISS are from samples of FRPL eligible children. All these effect sizes are statistically significant except for ISS Oral Comprehension.

[a] 12-month bandwidth estimates for the RDD study reported here (see Table 2).

[b] Treatment-on-the-treated estimates from the Intensive Substudy Sample of the RCT component of the overall study (see Lipsey, Farran, & Durkin, 2018).

Figure 1: Structure of the Regression-Discontinuity Design



Note. AY refers to Academic Year. The blue outlined and shaded boxes represent the TN-VPK or treatment cohort; the red outlined and unshaded boxes refer to the No Pre-K control cohort.

Figure 2: Consort Chart for the Achieved Sample of Children

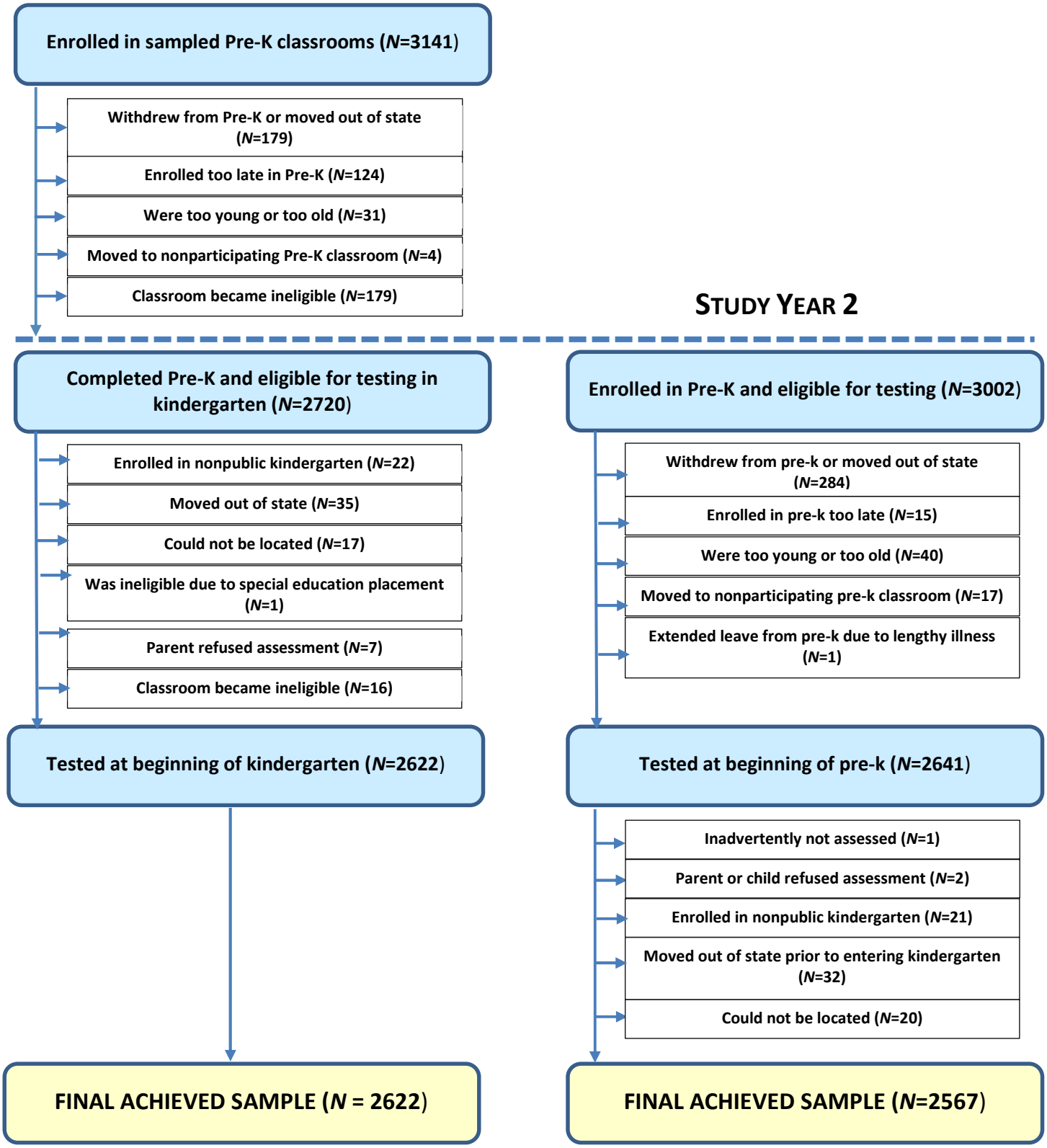
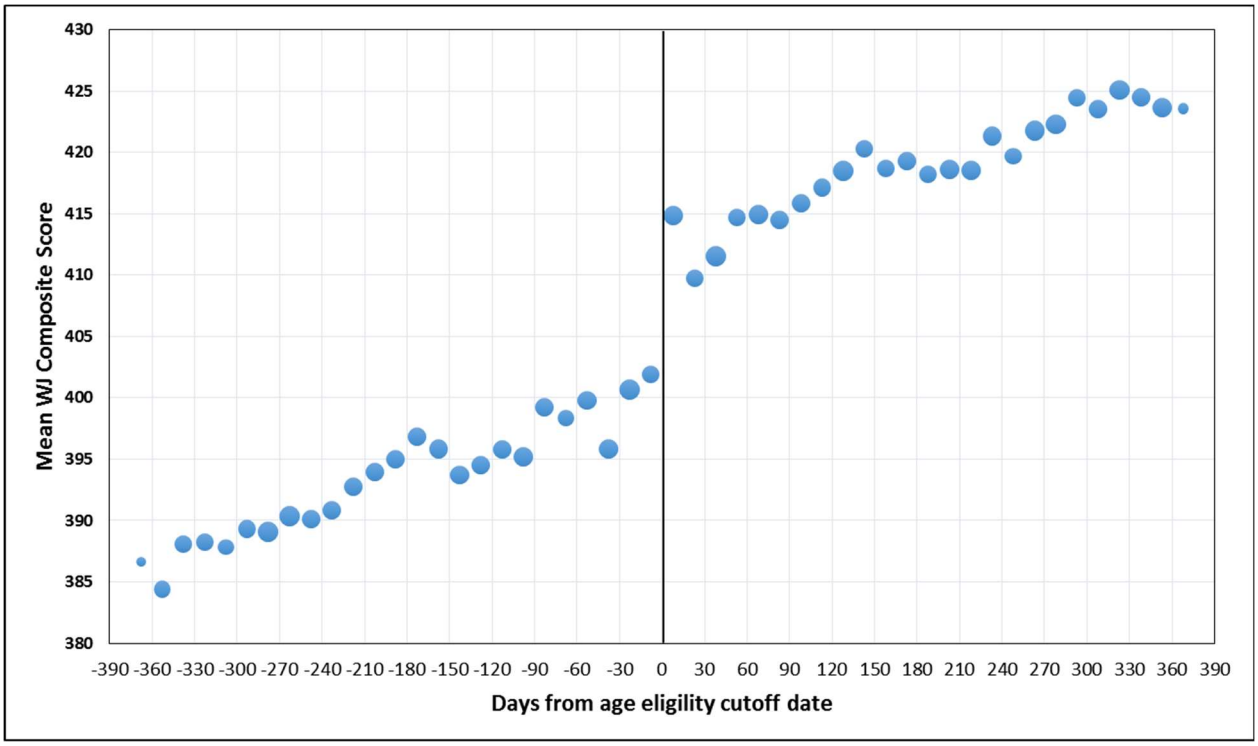
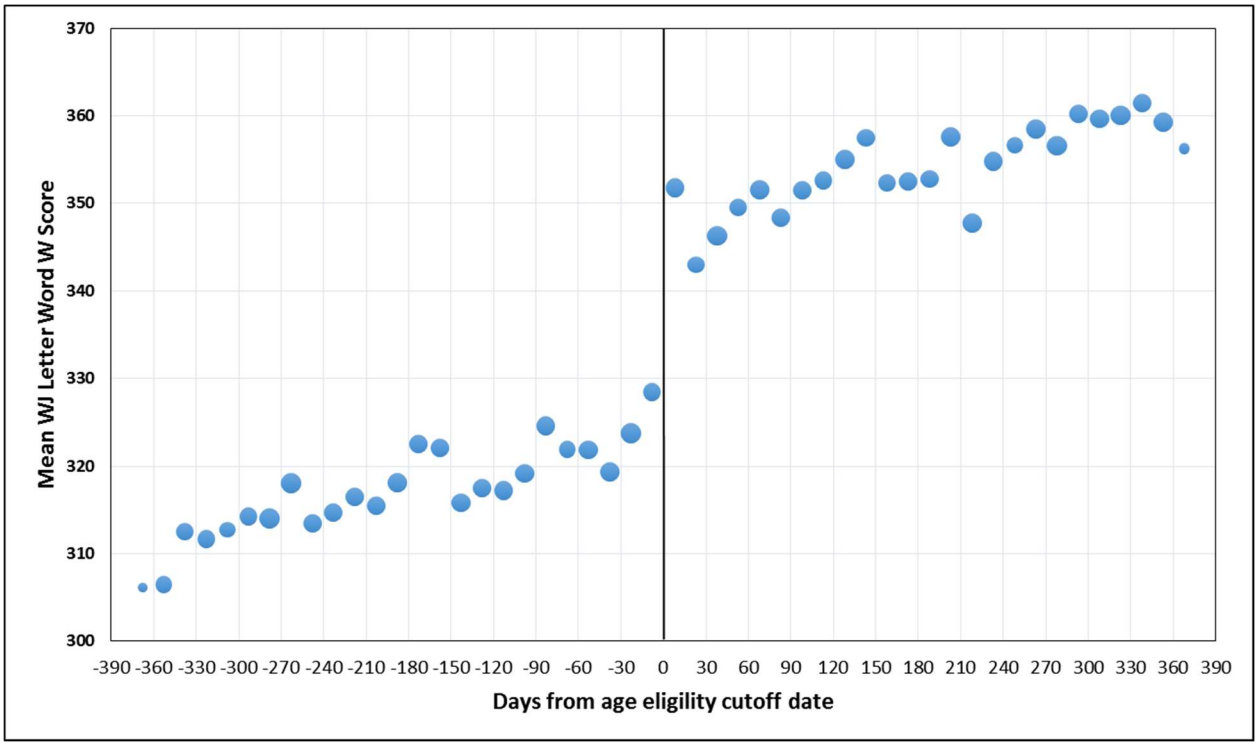


Figure 3: Unadjusted WJ Composite Score by Days from Age Eligibility Cutoff



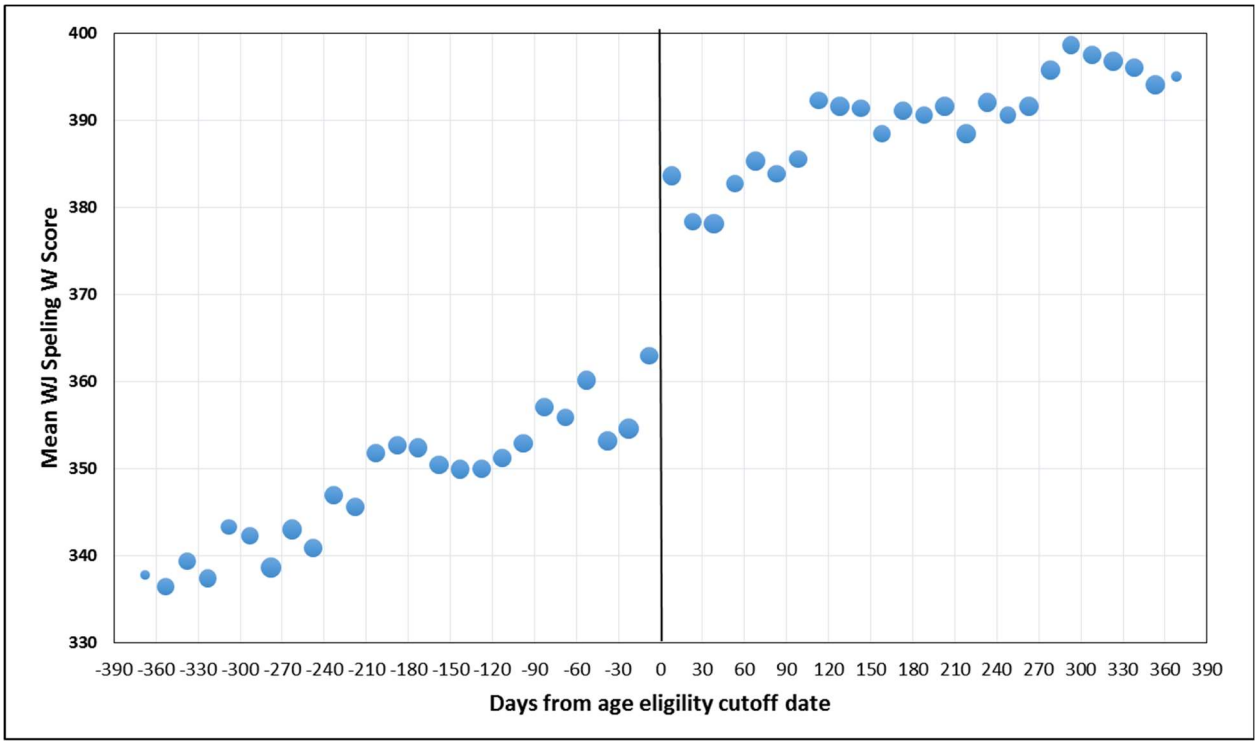
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 4: Unadjusted Letter Word Score by Days from Age Eligibility Cutoff



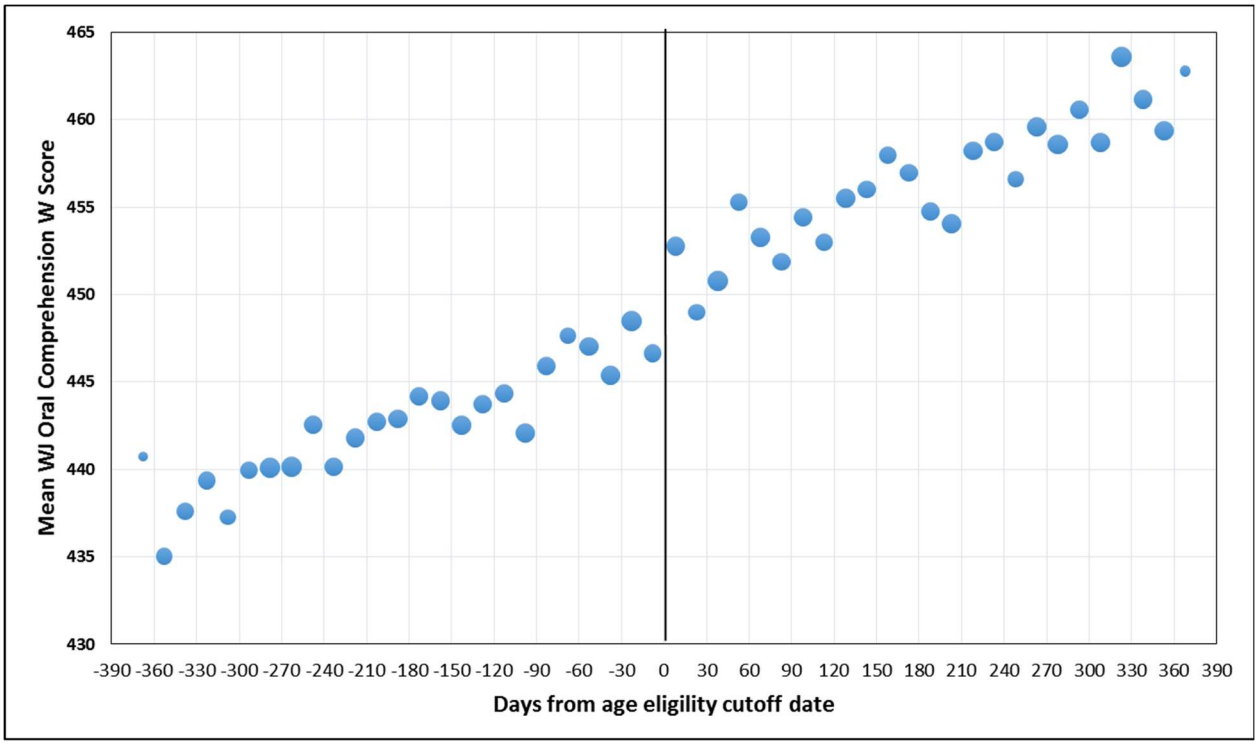
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 5: Unadjusted Spelling Score by Days from Age Eligibility Cutoff



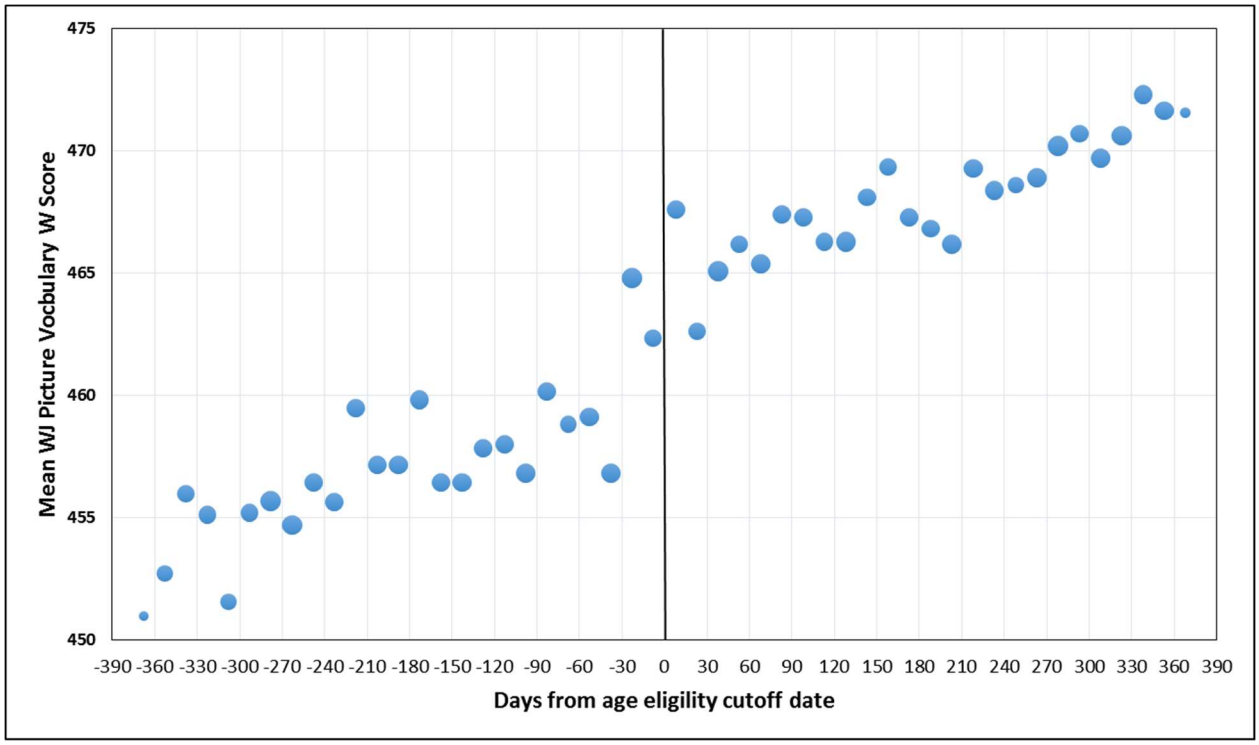
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 6: Unadjusted Oral Comprehension Score by Days from Age Eligibility Cutoff



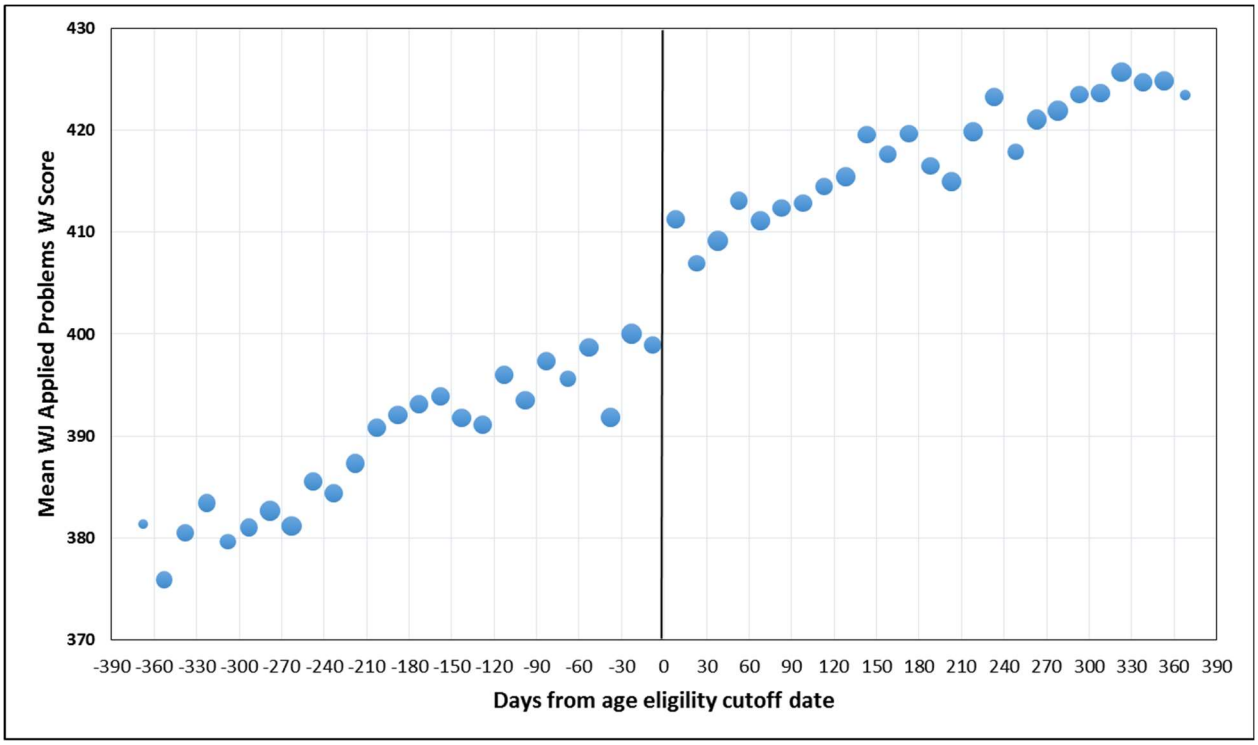
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 7: Unadjusted Picture Vocabulary Score by Days from Age Eligibility Cutoff



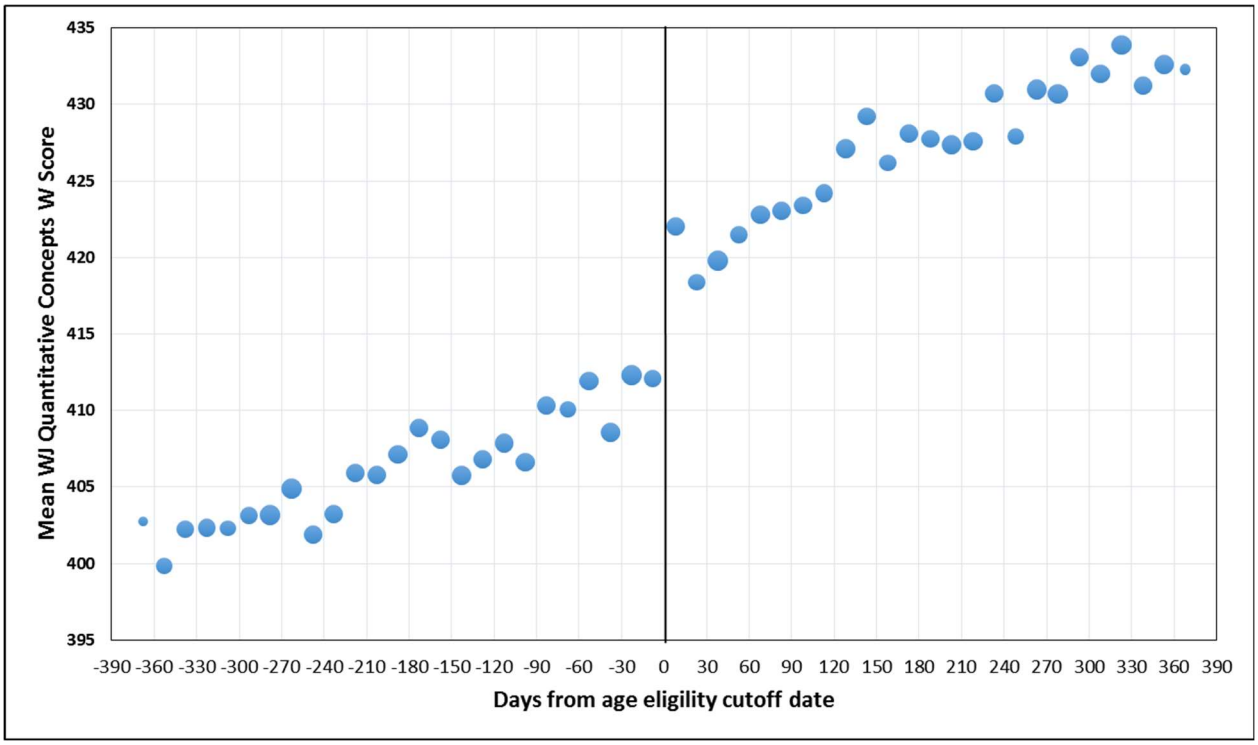
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 8: Unadjusted Applied Problems Score by Days from Age Eligibility Cutoff



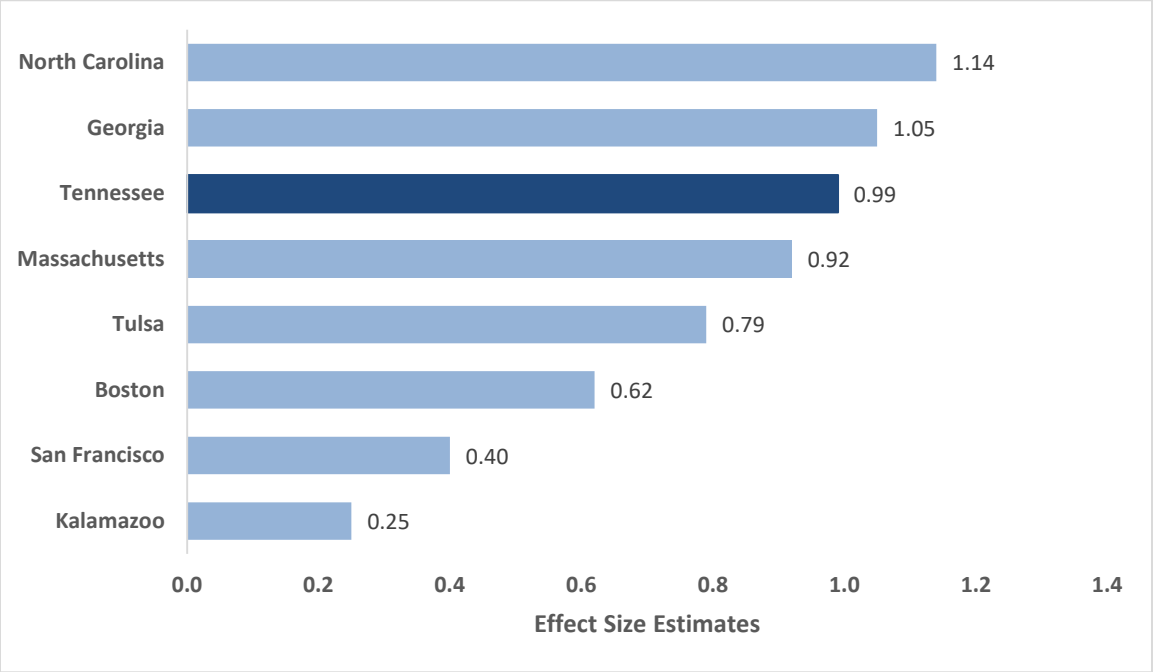
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 9: Unadjusted Quantitative Concepts Score by Days from Age Eligibility Cutoff



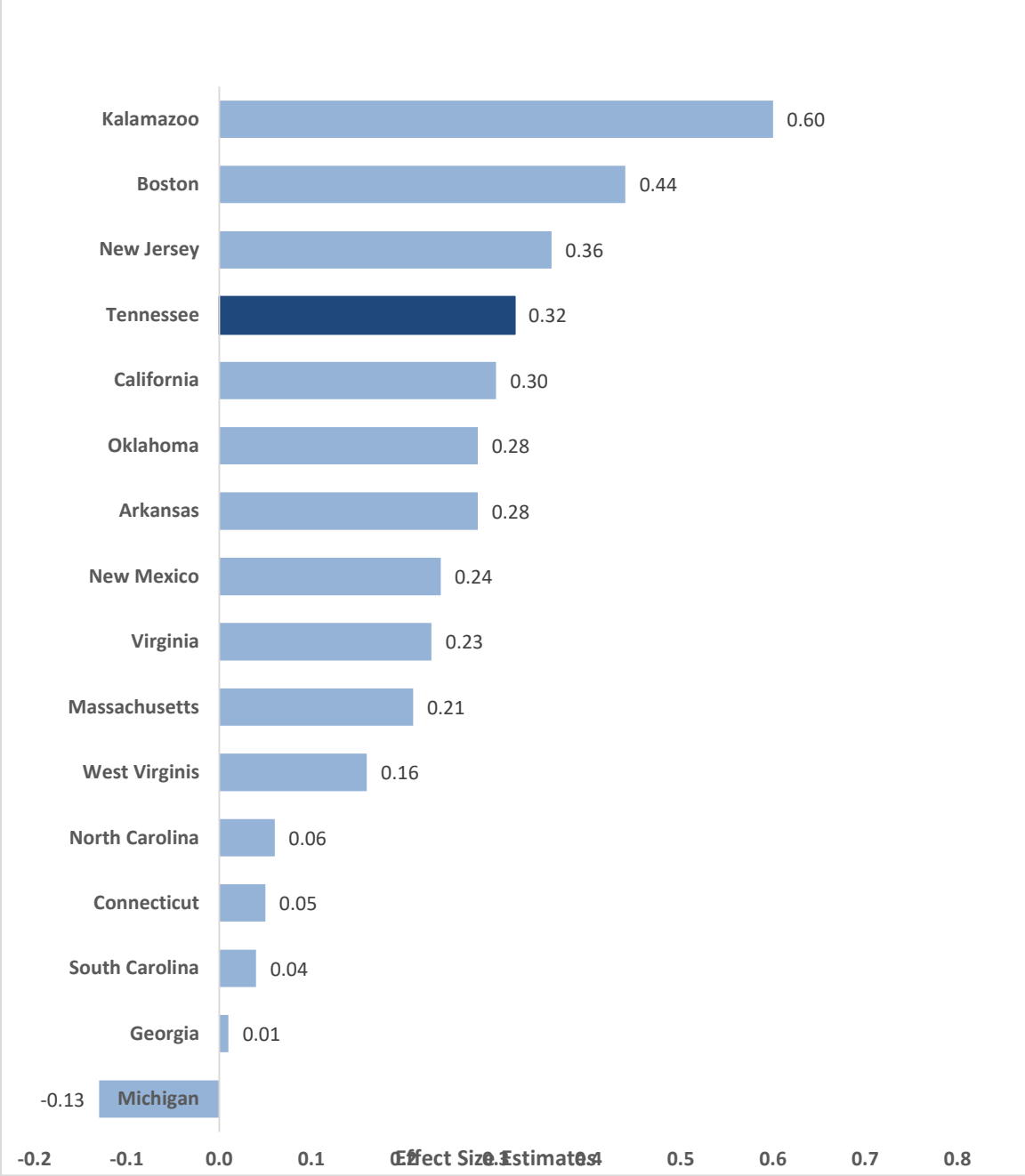
Notes. FRPL sample, N=4144. Unadjusted for covariates or sample weights. Means for 15-day intervals based on the tests described by Jacob, Zhue, Somers, and Bloom (2012). The size of each dot reflects the number of observations in that data point.

Figure 10: Effect Size Estimates for WJ Letter-Word Identification from Age-Cutoff Regression-Discontinuity Studies of Pre-K Programs



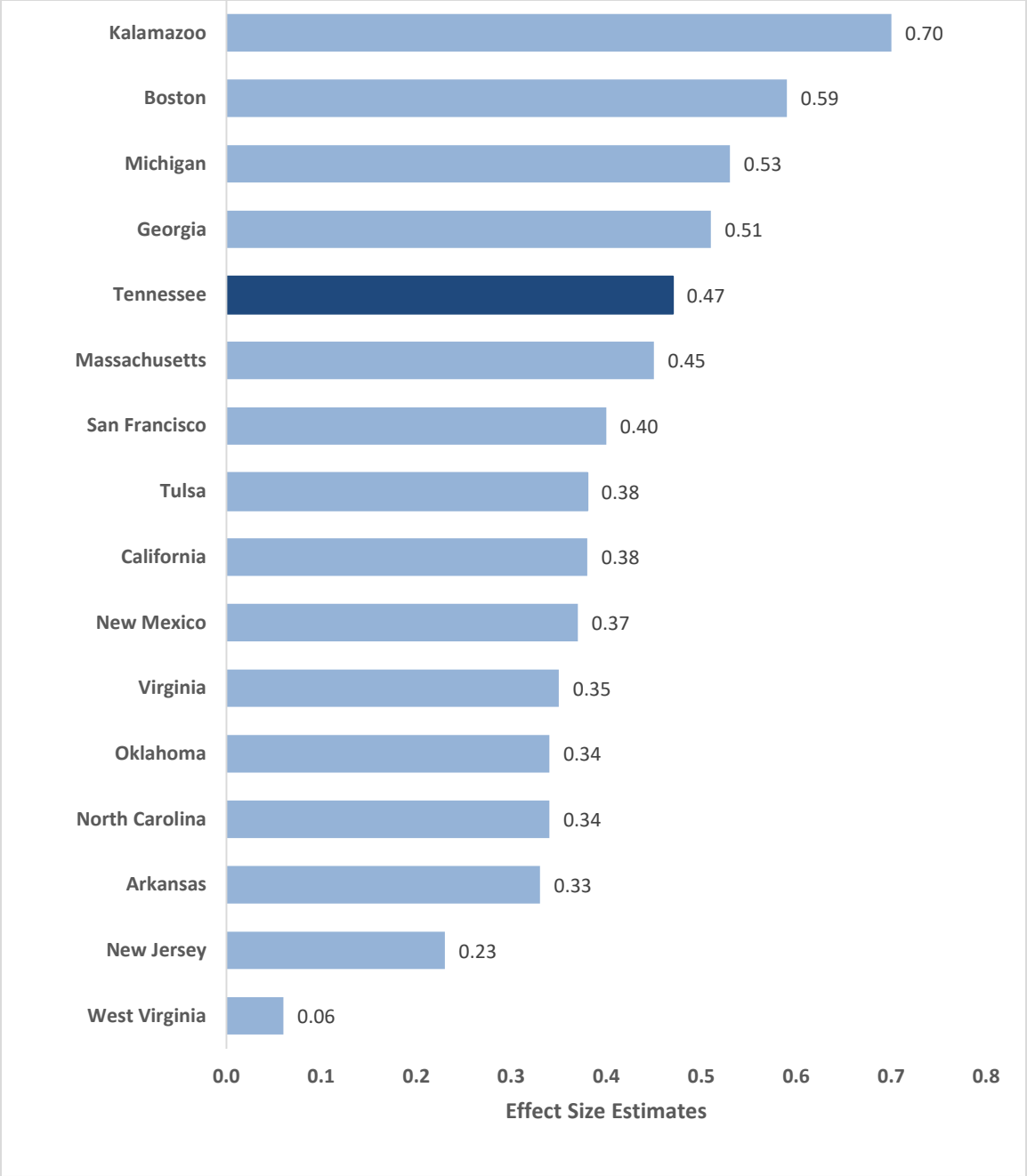
Note. Appendix G-1 identifies the reported effect size estimates for all outcomes examined in each study. References for the source studies are in Appendix G-2.

Figure 11: Effect Size Estimates for PPVT or WJ Picture Vocabulary from Age-Cutoff Regression-Discontinuity Studies of Pre-K Programs



Note. Appendix G-1 identifies the reported effect size estimates for all outcomes examined in each study. References for the source studies are in Appendix G-2.

Figure 12: Effect Size Estimates for WJ Applied Problems from Age-Cutoff Regression-Discontinuity Studies of Pre-K Programs



Note. Appendix G-1 identifies the reported effect size estimates for all outcomes examined in each study. References for the source studies are in Appendix G-2.

Appendices

Appendix A. Regression-Discontinuity Sampling Strategy

A probability sampling strategy was used to construct a sample representative of TN-VPK classrooms in Tennessee. The state was first divided into four geographic regions. Although three "Grand Divisions" (East, Middle, and West) are widely recognized in Tennessee geography, we split the Middle Grand Division into two separate regions (Central West and Central East) due to its size. The school districts within each of these regions were enumerated and schools with a TN-VPK program were identified within each district. Overall, there were 132 districts that included at least one school with a TN-VPK classroom and a total of 942 TN-VPK classrooms were identified within those districts and schools (see Appendix Table A-1).

Each classroom within each region was then assigned to a stratum based on its profile on four school characteristics that Tennessee Department of Education personnel identified as important distinctions within the TN-VPK program—urban/nonurban, partnership sites, pilot sites, and AYP priority schools. Urban/nonurban distinguished programs in the major urban areas. Partnership sites were those where the VPK programs were administratively associated with specific schools, but operated by a community center or Head Start agency. Pilot sites were the locations of the original and oldest VPK programs implemented during the initial pilot phase of the TN-VPK roll out. AYP priority schools were those designated as underperforming by the state and thus eligible for a range of supports and accountability monitoring.

As Appendix Table A-1 reports, the combinations of these characteristics produced 11 distinct strata within regions, with each region having between 6 and 11 strata. In three of the resulting strata there was at least one classroom in the sampling frame but no sample was drawn from that stratum. These strata were merged within region with the next most similar stratum, with similarity determined by allowing, first, differences on AYP priority, then on pilot site status.

To be eligible for selection into the final sample, a TN-VPK classroom had to be in place for both the 2009-10 and 2010-11 school year and staffed by the same teacher both years. The requirement that both age cohorts of children be enrolled in the same VPK classrooms was imposed to help make the children in the two cohorts as comparable as possible on whatever unmeasured variables were associated with the local school catchment areas and the process by which children were sorted into classrooms.

Within each region, a disproportionate random sample of eligible classrooms was selected with the aim of representing every strata but with larger sampling proportions for the smaller strata and smaller sampling proportions for the larger ones. Sampling weights were generated for use in the analysis that differentially weighted the classrooms in the different strata to make the resulting proportions match those for the statewide population of TN-VPK classrooms. The original sampling frame included 942 classrooms. Across the four regions, the final sample totaled 155 classrooms from 154 schools (one school contributed 2 classrooms) and 73 school districts.

In constructing this final sample, various sampling-related decisions were made in the different regions. These were as follows:

- **West.** The West region consists of 36 school districts and includes the city of Memphis. Forty VPK programs in 16 of the 36 school districts were drawn for the sample. All 40 programs contributed to the final participating sample in this region.
- **Central West.** The Central West region has 28 middle Tennessee school districts, including the city of Nashville. VPK programs in 17 of the districts in this region were drawn for the sample. Originally, 39 programs were selected from those districts, but three were dropped because changes at the host schools prevented the second cohort of children (the control group) from enrolling in TN-VPK classrooms comparable to the ones in which the first cohort (the treatment group) had participated the year before., leaving 36 programs in the sample from this region.
- **Central East**. This region includes 32 districts and is the region in which Chattanooga is located. The initial sample had 40 VPK programs in 20 of these districts. However, only 39 programs remained in the final sample; one was excluded due to the lack of comparability between the first and second cohort of children.
- **East.** In the East region, which includes 36 school districts and the city of Knoxville, 20 districts were sampled, which housed 41 VPK programs. Forty of the 41 programs remained in the final sample; one program had to be dropped because it did not have a comparable second cohort of children.

Construction of sampling weights. As noted, the original sampling frame included 942 TN-VPK classrooms, the full population of TN-VPK classrooms located by the research team at the time this study began. Of those, 155 classrooms were selected for the final sample. The sampling fraction for each stratum was computed as the number of classrooms sampled from that stratum divided by the total number of classrooms in that stratum in the original sampling frame. Thus, the overall sampling fraction was 155/942=.1645.

The weight to be assigned to each classroom in the analyses was calculated as the overall sampling fraction (.1645) divided by the sampling fraction for the respective stratum. Multiplying that weight by the number of cases sampled in each stratum yielded the number of cases from the population represented by that stratum. These summed to 155 as expected, i.e., the total sample *N* remained the same with the weighting.

Appendix Table A-1
Sampling Frame for VPK Classrooms, Strata, and Sampling Fractions

Region	Number of Districts		Strata				Number of Classrooms		Sampling Fraction
	Total	Sample	Urban	Partnership	Pilot	AYP Priority	Total	Sampled	
Central West	28	17	X				47	6	0.1277
			X			X	6	3	0.5000
			X		X		6	3	0.5000
			X	X			5	3	0.6000
			X	X	X		4	1	0.2500
						X	2	1	0.5000
					X		47	7	0.1489
				X			3	1	0.3333
							117	11	0.0940
Central West Totals	**28**	**17**					**237**	**36**	
West	36	16	X				47	6	0.1277
			X			X	11	3	0.2727
			X		X		9	3	0.3333
			X	X			39	6	0.1538
			X	X	X		5	2	0.4000
						X	6	2	0.3333
					X		33	3	0.0909
					X	X	15	3	0.2000
				X			14	3	0.2143
				X	X		1	1	1.0000
							101	8	0.0792
West Totals	**36**	**16**					**281**	**40**	

Continued

Appendix Table A-1 (*continued*)
Sampling Frame for VPK Classrooms, Strata, and Sampling Fractions

Region	Number of Districts		Strata				Number of Classrooms		Sampling Fraction
	Total	Sample	Urban	Partnership	Pilot	AYP Priority	Total	Sampled	
Central East	32	20	X				9	3	0.3333
			X			X	1	1	1.0000
			X		X		4	1	0.2500
			X	X			13	4	0.3077
						X	4	2	0.5000
					X		16	5	0.3125
				X			8	4	0.5000
							153	19	0.1242
Central East Totals	**32**	**20**					**208**	**39**	
East	36	20	X				7	4	0.5714
						X	3	2	0.6667
					X		31	9	0.2903
				X			18	6	0.3333
				X	X		10	3	0.3000
							147	16	0.1088
East Totals	**36**	**20**					**216**	**40**	
Grand Totals	**132**	**73**					**942**	**155**	**.1645**

Note. The overall sampling fraction was 0.1645, which was divided by the sampling fraction of the respective stratum to derive the overall weight for that stratum. The regions are listed in the order in which data collection occurred. That is, outcome data were collected for Central West schools in the fall of 2010, for West schools in fall 2011, for Central East schools in fall 2012, and for East schools in fall 2013.

Appendix B. Coding of Variables in the Analysis Models

- *Age*. Age was centered at the cutoff for pre-k eligibility; children with birthdays before September 30th received positive values for the number of days before that date, those with birthdays after that date received corresponding negative values.
- *Condition*. Children who completed TN-VPK and were enrolled in TN public schools during the subsequent year were assigned a "1" as the treatment cohort, with a "0" given to children who had just begun TN-VPK, the control cohort.
- *Region*. Because the sampling design had region as a stratification variable with data collected over successive years, it was represented as a fixed effects factor in the analysis. Three dummy variables, one each for the West, Central East, and East regions, were used with Central West omitted as the reference value.
- *Covariates*. To adjust for any baseline differences (see Table 1) and improve statistical power, the baseline demographic characteristics and the timing-of-assessment variable were included as covariates. Demographic characteristics included: gender (male=1, female=0); Black (Yes=1, No=0); Hispanic (Yes=1, No=0); language other than English spoken at home (Yes=1, No=0); and whether the child had an IEP (Yes=1, No=0). The number of days between administration of the Woodcock Johnson tests and the start of the school year also was included and centered at the grand mean. Note that the sample included mixed race children whose race/ethnicity was coded in each category represented thus the respective codes are not mutually exclusive.

Appendix Table C-1
Statistical Models and TN-VPK Effects for the WJ Composite

Independent Variable	WJ Composite 12 mo Bandwidth			WJ Composite 6 mo Bandwidth			WJ Composite 3 mo Bandwidth		
	b	**SE**	***p***	***b***	**SE**	***p***	***b***	**SE**	***p***
Intercept	405.85	.80	<.001	405.77	1.04	<.001	406.40	1.435	<.001
Days from age cutoff	.04	<.01	<.001	.04	.01	<.001	.03	.02	.079
Region: West	-1.34	.87	.127	-1.70	1.06	.111	-2.56	1.36	.061
Region: Central East	.19	.91	.834	.43	1.11	.700	-.92	1.45	.527
Region: East	-.06	.93	.947	.60	1.11	.591	.24	1.45	.866
Male	-2.33	.39	<.001	-1.68	.58	.004	-2.25	.82	.007
Black	-3.04	.59	<.001	-2.27	.80	.005	-2.08	1.07	.052
Hispanic	-11.59	1.00	<.001	-13.46	1.54	<.001	-14.24	2.32	<.001
Native language not English	-6.69	1.11	<.001	-6.66	1.63	<.001	-9.02	2.44	<.001
Test lag	.13	.02	<.001	.18	.03	<.001	.18	.04	<.001
Has an IEP	-10.56	.69	<.001	-11.06	.99	<.001	-11.71	1.42	<.001
TN-VPK participation	**10.93**	.79	**<.001**	**10.74**	1.15	**<.001**	**10.84**	1.66	**<.001**

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing.

Analyses were weighted to represent the statewide population of TN-VPK classrooms. The composite is an average of the W-scores for the six subtests used to measure outcomes: Letter-Word, Spelling, Oral Comprehension, Picture Vocabulary, Applied Problems, and Quantitative Concepts. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table C-2
Statistical Models and TN-VPK Effects for WJ Letter-Word

Independent Variable	WJ Letter-Word 12 mo Bandwidth			WJ Letter-Word 6 mo Bandwidth			WJ Letter-Word 3 mo Bandwidth		
	b	**SE**	***p***	***b***	**SE**	***p***	***b***	**SE**	***p***
Intercept	328.77	1.41	<.001	329.04	1.78	<.001	331.26	2.47	<.001
Days from age cutoff	.04	<.01	<.001	.04	.01	<.001	.05	.03	.052
Region: West	-.47	1.56	.762	-1.62	1.85	.381	-2.73	2.42	.263
Region: Central East	-.66	1.63	.686	-1.19	1.93	.538	-2.68	2.58	.299
Region: East	-1.71	1.66	.304	-.50	1.94	.796	-1.30	2.58	.613
Male	-4.02	.68	<.001	-4.32	.96	<.001	-6.14	1.38	<.001
Black	1.48	1.02	.148	3.84	1.36	.005	3.41	1.84	.064
Hispanic	-11.26	1.72	<.001	-13.04	2.58	<.001	-10.68	3.91	.006
Native language not English	-2.71	1.91	.156	-4.59	2.73	.093	-6.38	4.10	.121
Test lag	.24	.04	<.001	.32	.05	<.001	.38	.07	<.001
Has an IEP	-10.23	1.18	<.001	-11.47	1.65	<.001	-11.74	2.39	<.001
TN-VPK Participation	**22.02**	1.36	**<.001**	**21.83**	1.92	**<.001**	**19.85**	2.79	**<.001**

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table C-3
Statistical Models and TN-VPK Effects for WJ Spelling

Independent Variable	WJ Spelling 12 mo Bandwidth			WJ Spelling 6 mo Bandwidth			WJ Spelling 3 mo Bandwidth		
	b	**SE**	***p***	***b***	**SE**	***p***	***b***	**SE**	***p***
Intercept	362.69	1.20	<.001	362.87	1.67	<.001	362.83	2.35	<.001
Days from age cutoff	.06	<.01	<.001	.06	.01	<.001	.04	.03	.133
Region: West	1.57	1.27	.219	.61	1.69	.717	-1.41	2.24	.530
Region: Central East	.62	1.33	.640	1.69	1.76	.340	-.57	2.38	.810
Region: East	2.33	1.36	.088	4.15	1.78	.021	3.22	2.38	.178
Male	-5.98	.64	<.001	-4.93	.93	<.001	-5.35	1.36	<.001
Black	-1.57	.91	.086	-.01	1.29	.993	1.87	1.77	.291
Hispanic	-4.14	1.60	.010	-5.18	2.49	.038	-6.21	3.83	.106
Native language not English	.43	1.77	.809	.95	2.63	.717	-.58	4.02	.884
Test lag	.22	.04	<.001	.27	.05	<.001	.31	.07	<.001
Has an IEP	-11.32	1.10	<.001	-12.37	1.60	<.001	-12.79	2.35	<.001
TN-VPK Participation	**20.35**	1.28	**<.001**	**19.28**	1.86	**<.001**	**19.76**	2.75	**<.001**

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table C-4
Statistical Models and TN-VPK Effects for WJ Oral Comprehension

Independent Variable	WJ Oral Comprehension 12 mo Bandwidth			WJ Oral Comprehension 6 mo Bandwidth			WJ Oral Comprehension 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	454.54	0.83	<.001	454.08	1.08	<.001	454.27	1.40	<.001
Days from age cutoff	.03	<.01	<.001	.03	.01	<.001	.01	.01	.414
Region: West	-3.39	.93	<.001	-3.01	1.16	.010	-3.10	1.36	.024
Region: Central East	1.32	.97	.175	1.79	1.21	.140	.61	1.44	.673
Region: East	-.87	.99	.377	-.88	1.22	.472	-.66	1.45	.648
Male	-1.53	.39	<.001	-.57	.56	.305	-.44	.79	.572
Black	-5.45	.60	<.001	-5.22	.81	<.001	-5.71	1.04	<.001
Hispanic	-11.09	1.00	<.001	-13.86	1.50	<.001	-15.39	2.23	<.001
Native language not English	-12.35	1.11	<.001	-11.33	1.59	<.001	-12.34	2.34	<.001
Test lag	.05	.02	.028	.09	.03	.005	.07	.04	.102
Has an IEP	-10.02	.68	<.001	-9.58	.96	<.001	-10.67	1.36	<.001
TN-VPK Participation	**3.25**	.79	**<.001**	**2.91**	1.11	**.009**	**4.03**	1.59	**.012**

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table C-5
Statistical Models and TN-VPK Effects for WJ Picture Vocabulary

Independent Variable	WJ Picture Vocabulary 12 mo Bandwidth			WJ Picture Vocabulary 6 mo Bandwidth			WJ Picture Vocabulary 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	465.47	.81	<.001	465.42	1.06	<.001	466.94	1.44	<.001
Days from age cutoff	.02	<.01	<.001	.02	.01	<.001	.02	.02	.159
Region: West	-1.87	.89	.038	-1.59	1.10	.150	-2.02	1.38	.146
Region: Central East	-.31	.93	.741	-.31	1.15	.785	-1.78	1.47	.229
Region: East	-.55	.95	.562	-.55	1.16	.638	-.98	1.47	.506
Male	.44	.40	.268	1.33	.57	.020	1.32	.82	.109
Black	-1.78	.59	.003	-2.28	.81	.005	-3.03	1.08	.005
Hispanic	-18.54	1.00	<.001	-20.87	1.54	<.001	-20.85	2.33	<.001
Native language not English	-13.84	1.11	<.001	-13.57	1.63	<.001	-18.25	2.44	<.001
Test lag	.03	.02	.249	.04	.03	.182	.03	.04	.548
Has an IEP	-7.92	.69	<.001	-7.60	.99	<.001	-7.57	1.43	<.001
TN-VPK Participation	**3.69**	.80	**<.001**	**3.69**	1.14	**.001**	2.69	1.67	.107

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table C-6
Statistical Models and TN-VPK Effects for WJ Applied Problems

Independent Variable	WJ Applied Problems 12 mo Bandwidth			WJ Applied Problems 6 mo Bandwidth			WJ Applied Problems 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	407.91	1.26	<.001	407.52	1.55	<.001	407.86	2.16	<.001
Days from age cutoff	.05	<.01	<.001	.04	.01	<.001	.02	.02	.451
Region: West	-3.03	1.36	.027	-3.17	1.52	.039	-3.50	2.06	.092
Region: Central East	.10	1.42	.943	.54	1.59	.735	-.22	2.20	.920
Region: East	-.01	1.45	.995	-.28	1.60	.862	-.81	2.20	.712
Male	-2.02	.64	.002	-.56	.90	.532	-1.97	1.24	.114
Black	-8.07	.94	<.001	-8.16	1.21	<.001	-7.98	1.62	<.001
Hispanic	-17.66	1.61	<.001	-19.09	2.40	<.001	-24.34	3.51	<.001
Native language not English	-9.90	1.79	<.001	-9.95	2.53	<.001	-12.58	3.68	<.001
Test lag	.14	.04	<.001	.20	.05	<.001	.14	.06	.023
Has an IEP	-15.93	1.11	<.001	-16.87	1.54	<.001	-18.60	2.15	<.001
TN-VPK Participation	**8.37**	1.28	**<.001**	**9.80**	1.79	**<.001**	**11.36**	2.51	**<.001**

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table C-7
Statistical Models and TN-VPK Effects for WJ Quantitative Concepts

Independent Variable	WJ Quantitative Concepts 12 mo Bandwidth			WJ Quantitative Concepts 6 mo Bandwidth			WJ Quantitative Concepts 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	415.58	0.78	<.001	415.42	1.03	<.001	415.21	1.39	<.001
Days from age cutoff	.04	<.01	<.001	.04	.01	<.001	.02	.02	.132
Region: West	-.77	.84	.360	-1.43	1.03	.169	-2.75	1.28	.034
Region: Central East	.22	.88	.801	.20	1.08	.853	-.82	1.37	.551
Region: East	.63	.90	.483	1.72	1.08	.114	1.97	1.37	.153
Male	-.91	.40	.021	-1.00	.58	.087	-.85	.83	.309
Black	-2.66	.58	<.001	-1.48	.80	.064	-.93	1.05	.380
Hispanic	-6.86	1.00	<.001	-8.78	1.56	<.001	-8.12	2.33	<.001
Native language not English	-1.81	1.11	.103	-1.29	1.65	.433	-3.96	2.45	.106
Test lag	.10	.02	<.001	.14	.03	<.001	.16	.04	<.001
Has an IEP	-7.94	.69	<.001	-8.55	1.00	<.001	-9.16	1.44	<.001
TN-VPK Participation	**7.88**	.80	**<.001**	**7.08**	1.16	**<.001**	**7.45**	1.68	**<.001**

Note. Sample sizes are 4,144 for the 12-month bandwidth, 2,067 for the 6-month bandwidth, and 1,050 for the 3-month bandwidth. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Analyses were weighted to represent the statewide population of TN-VPK classrooms. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the appropriate functional form for all models was linear.

Appendix Table D-1
Robustness Checks for Estimating the TN-VPK Effects

Outcome and Trim	12 mo Bandwidth			6 mo Bandwidth			3 mo Bandwidth		
	(1) SD[a]	(2) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ2)/(1)[b]	(3) SD[a]	(4) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ4)/(3)[b]	(5) SD[a]	(6) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ6)/(5)[b]
Composite									
All cases	19.34	10.93 (.79)		16.97	10.74 (1.15)		16.50	10.84 (1.66)	
Outermost 1% trimmed		10.97 (.79)	-.002		10.30 (1.16)	.026		11.36 (1.67)	-.032
Outermost 5% trimmed		10.93 (.81)	.000		9.70 (1.20)	.061		10.50 (1.77)	.021
Outermost 10% trimmed		10.59 (.83)	.018		9.18 (1.28)	.093		9.17 (1.85)	.101
Letter-Word									
All cases	29.71	22.02 (1.36)		27.24	21.83(1.92)		26.95	19.85 (2.79)	
Outermost 1% trimmed		22.11 (1.37)	-.003		21.17 (1.94)	.024		20.56 (2.80)	-.026
Outermost 5% trimmed		22.25 (1.40)	.008		20.34 (2.02)	.055		19.26 (2.96)	.022
Outermost 10% trimmed		21.82 (1.43)	.007		18.62 (2.16)	.118		16.33 (3.10)	.131
Spelling									
All cases	30.00	20.34 (1.28)		26.61	19.28 (1.86)		25.73	19.76 (2.75)	
Outermost 1% trimmed		20.52 (1.28)	.011		18.58 (1.87)	.026		20.76 (2.75)	-.039
Outermost 5% trimmed		20.20 (1.31)	.021		17.99 (1.95)	.048		19.32 (2.89)	.017
Outermost 10% trimmed		19.99 (1.34)	.028		17.63 (2.08)	.062		18.91 (3.03)	.033

Continued

Appendix Table D-1 (*continued*)
Robustness Checks for Models Estimating the TN-VPK Effects

Outcome and Trim	12 mo Bandwidth			6 mo Bandwidth			3 mo Bandwidth		
	(1) SD[a]	(2) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ2)/(1)[b]	(3) SD[a]	(4) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ4)/(3)[b]	(5) SD[a]	(6) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ6)/(5)[b]
Oral Comprehension									
All cases	16.45	3.25 (.79)		15.19	2.91 (1.11)		14.85	4.03 (1.59)	
Outermost 1% trimmed		3.17 (.79)	.005		2.90 (1.13)	<.001		4.48 (1.61)	-.030
Outermost 5% trimmed		3.15 (.81)	.006		2.47 (1.17)	.029		3.93 (1.69)	.007
Outermost 10% trimmed		2.95 (.83)	.018		2.20 (1.24)	.047		3.09 (1.79)	.063
Picture Vocabulary									
All cases	16.34	3.69 (.80)		15.92	3.69 (1.14)		16.33	2.69 (1.67)	
Outermost 1% trimmed		3.80 (.80)	-.007		3.36 (1.16)	.021		3.03 (1.67)	-.039
Outermost 5% trimmed		3.89 (.82)	-.012		2.46 (1.22)	.077		1.91 (1.80)	.048
Outermost 10% trimmed		3.75 (.84)	-.004		2.05 (1.31)	.103		1.06 (1.90)	.100
Applied Problems	27.18			23.87			23.56		
All cases		8.37 (1.28)			9.80 (1.79)			11.36 (2.51)	
Outermost 1% trimmed		8.36 (1.28)	<.001		9.24 (1.80)	.023		11.61 (2.52)	-.011
Outermost 5% trimmed		8.23 (1.32)	.006		8.75 (1.89)	.044		11.19 (2.66)	.007
Outermost 10% trimmed		7.80 (1.34)	.021		8.83 (2.00)	.041		9.29 (2.83)	.088

Continued

Appendix Table D-1 (*continued*)
Robustness Checks for Models Estimating the TN-VPK Effects

Outcome and Trim	12 months			6 months			3 months		
	(1) SD[a]	(2) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ2)/(1)[b]	(3) SD[a]	(4) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ4)/(3)[b]	(5) SD[a]	(6) Coefficient for TN-VPK vs. No Pre-K (SE)	Difference in Estimates (Δ6)/(5)[b]
Quantitative Concepts									
All cases	17.32	7.88 (.80)		15.41	7.08 (1.16)		14.89	7.45 (1.68)	
Outermost 1% trimmed		7.84 (.80)	.002		6.67 (1.17)	.027		7.70 (1.70)	-.017
Outermost 5% trimmed		7,87 (.82)	<.001		6.29 (1.22)	.051		7.30 (1.77)	.010
Outermost 10% trimmed		7.23 (.84)	.038		5.84 (1.30)	.080		6.34 (1.86)	.075
Average effect size for:[c]									
All cases vs. 1% trimmed		.004			.021			.028	
All cases vs. 5% trimmed		.008			.052			.019	
All cases vs. 10% trimmed		.019			.078			.084	

Note. Only children eligible for free or reduced price lunch were included.

[a] Standard deviation for the TN-VPK sample, i.e., children who had completed TN-VPK and were tested at the beginning of kindergarten.

[b] Effect sizes for the difference between the coefficients of the sample with all cases and the respective trimmed sample. The coefficient differences are computed from columns (2), (4), and (6) respectively by subtracting each trimmed value from the all cases value in turn. For comparison purposes, those differences (Δ) are standardized into an effect size metric by dividing each by the standard deviation for all cases shown in columns (1), (3), and (5) respectively. Those standardized differences are then reported in the "Difference in Estimates" columns.

[c] This is the average of the absolute values of the "Difference in Estimates" effect sizes for the respective comparison of the two samples.

Appendix Table E-1
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Male and Female Children

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
WJ Composite								
12-month bandwidth						.464	12.92	
Female	412.90	1044	402.25	1044	10.65			.824
Male	410.85	1034	399.63	1022	11.22			.868
6-month bandwidth						.984	12.66	
Female	412.78	511	402.05	553	10.73			.848
Male	411.11	498	400.36	505	10.75			.849
3 -month bandwidth						.593	11.94	
Female	412.66	267	402.25	181	10.41			.872
Male	410.87	252	399.58	259	11.29			.946
Letter-Word								
12-month bandwidth							21.15	
Female	348.84	1044	326.60	1044	22.24	.731		1.052
Male	344.58	1034	322.81	1022	21.77			1.029
6-month bandwidth						.144	19.71	
Female	349.65	511	326.44	553	23.21			1.178
Male	343.88	498	323.46	505	20.42			1.036
3 -month bandwidth						.645	19.17	
Female	348.74	267	328.27	181	20.47			1.068
Male	341.93	252	322.74	259	19.19			1.001
Spelling								
12-month bandwidth						.335	20.82	
Female	382.77	1044	361.84	1044	20.93			1.005
Male	376.19	1034	356.47	1022	19.72			.947
6-month bandwidth						.181	20.62	
Female	382.74	511	362.23	553	20.51			.995
Male	376.53	498	358.50	505	18.03			.874
3 -month bandwidth						.322	20.43	
Female	382.56	267	361.49	272	21.07			1.031
Male	375.80	252	357.43	259	18.37			.899
Oral Comprehension								
12-month bandwidth						.079	13.71	
Female	451.96	1044	449.37	1044	2.59			.189
Male	451.12	1034	447.16	1022	3.96			.289
6-month bandwidth						.004	13.08	
Female	450.53	511	449.54	553	.99			.076
Male	451.98	498	447.10	505	4.88			.373
3 -month bandwidth						.052	12.87	
Female	451.57	267	449.03	272	2.54			.197
Male	452.73	252	447.13	259	5.60			.435

Continued

Appendix Table E-1 (*continued*)
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Male and Female Children

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Picture Vocabulary								
12-month bandwidth						.572	10.72	
Female	464.42	1044	460.94	1044	3.48			.325
Male	465.08	1034	461.16	1022	3.92			.366
6-month bandwidth						.459	10.89	
Female	464.00	511	460.73	553	3.27			.300
Male	465.77	498	461.66	505	4.11			.377
3 -month bandwidth						.568	10.41	
Female	463.60	267	461.37	272	2.23			.214
Male	465.42	252	462.24	259	3.18			.305
Applied Problems								
12-month bandwidth						.096	17.74	
Female	408.51	1044	401.16	1044	7.35			.414
Male	407.55	1034	398.08	1022	9.47			.534
6-month bandwidth						.767	18.12	
Female	409.48	511	399.94	553	9.54			.526
Male	409.19	498	399.13	505	10.06			.555
3 -month bandwidth						.075	16.08	
Female	409.58	267	400.39	272	9.19			.572
Male	409.94	252	396.31	259	13.63			.848
Quantitative Concepts								
12-month bandwidth						.146	13.99	
Female	420.69	1044	413.56	1044	7.13			.510
Male	420.55	1034	412.08	1022	8.47			.605
6-month bandwidth						.807	13.43	
Female	420.26	511	413.33	553	6.93			.516
Male	419.41	498	412.20	505	7.21			.537
3 -month bandwidth						.596	12.49	
Female	419.89	267	412.87	272	7.02			.562
Male	419.51	252	411.60	259	7.91			.633

Note. Only children eligible for free or reduced price lunch were included. Previous analyses showed that the functional form was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Marginal means from the multilevel model.

[b] To facilitate comparison of these effect sizes with those reported for the overall pre-k effects, they were calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the treatment (TN-VPK) group.

[c] This is the *p*-value for the interaction term for condition (TN-VPK vs. No Pre-K) and whether the child was Male or Female. Estimates were based on a multilevel model with children nested within pre-k classroom. In addition to condition, gender, and days from the age cutoff (centered at zero), covariates included: (1) Region; (2) whether the child was Black; (3) whether the child was Hispanic; (4) whether the child's native language was not English; (5) whether the child had an IEP placement; (6) the number of days elapsed between the Woodcock-Johnson testing date and the start of school (centered at the grand mean); and (7) the interaction term.

Appendix Table E-2
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Black and Non-Black Children

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
WJ Composite								
12-month bandwidth						.813	12.92	
Black	410.01	750	398.94	748	11.07			.857
Not Black	412.95	1328	402.08	1318	10.87			.841
6-month bandwidth						.931	12.66	
Black	410.56	372	399.74	376	10.82			.855
Not Black	412.78	737	402.06	682	10.72			.847
3 -month bandwidth						.779	11.94	
Black	410.28	197	399.81	181	10.47			.877
Not Black	412.62	322	401.63	350	10.99			.920
Letter-Word								
12-month bandwidth							21.15	
Black	347.20	750	326.11	748	21.09	.371		.977
Not Black	346.39	1328	323.99	1318	22.40			1.059
6-month bandwidth						.228	19.71	
Black	348.37	372	328.33	376	20.04			1.017
Not Black	345.83	637	323.26	682	22.57			1.145
3 -month bandwidth						.486	19.17	
Black	346.88	197	328.56	181	18.32			.956
Not Black	344.52	322	324.08	350	20.44			1.066
Spelling								
12-month bandwidth						.156	20.82	
Black	379.22	750	357.49	748	21.73			1.043
Not Black	379.79	1328	360.01	1318	19.78			.950
6-month bandwidth						.518	20.62	
Black	380.16	372	359.96	376	20.20			.980
Not Black	379.50	637	360.60	682	18.90			.917
3 -month bandwidth						.920	20.43	
Black	380.36	197	360.82	181	19.54			.956
Not Black	378.64	322	358.80	350	19.84			.971
Oral Comprehension								
12-month bandwidth						.606	13.71	
Black	448.21	750	444.65	748	3.56			.260
Not Black	453.44	1328	450.32	1318	3.12			.228
6-month bandwidth						.632	13.08	
Black	448.15	372	444.82	376	3.33			.255
Not Black	453.07	637	450.33	682	2.74			.209
3 -month bandwidth						.701	12.87	
Black	448.71	197	444.20	181	4.51			.350
Not Black	454.09	322	450.25	350	3.84			.298

Continued

Appendix Table E-2 (*continued*)
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Black and Non-Black Children

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Picture Vocabulary								
12-month bandwidth						.036	10.72	
Black	462.95	750	460.54	748	2.41			.225
Not Black	465.66	1328	461.45	1318	4.21			.393
6-month bandwidth						.052	10.89	
Black	462.55	372	460.58	376	1.97			.181
Not Black	466.07	637	461.67	682	4.40			.404
3 -month bandwidth						.025	10.41	
Black	461.16	197	461.39	181	-.23			-.022
Not Black	466.20	322	462.37	350	3.83			.368
Applied Problems								
12-month bandwidth						.498	17.74	
Black	403.21	750	394.17	748	9.04			.510
Not Black	410.80	1328	402.70	1318	8.10			.457
6-month bandwidth						.405	18.12	
Black	404.72	372	393.77	376	10.95			.604
Not Black	412.04	637	402.72	682	9.32			.514
3 -month bandwidth						.870	16.08	
Black	404.81	197	393.13	181	11.68			.726
Not Black	412.57	322	401.34	350	11.23			.698
Quantitative Concepts								
12-month bandwidth						.261	13.99	
Black	419.37	750	410.80	748	8.57			.613
Not Black	421.53	1328	413.93	1318	7.60			.543
6-month bandwidth						.101	13.43	
Black	419.65	372	411.11	376	8.54			.636
Not Black	420.06	637	413.60	682	6.46			.481
3 -month bandwidth						.201	12.49	
Black	419.91	197	410.78	181	9.13			.731
Not Black	419.67	322	412.89	350	6.78			.543

Note. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. Previous analyses showed that the functional form was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Marginal means from the multilevel model.

[b] To facilitate comparison of these effect sizes with those reported for the overall pre-k effects, they were calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the treatment (TN-VPK) group.

[c] This is the *p*-value for the interaction term for condition (TN-VPK vs. No Pre-K) and whether the child was Black or not Black. Estimates were based on a multilevel model with children nested within pre-k classroom. In addition to condition, Black, and days from the age cutoff (centered at zero), covariates included: (1) Region; (2) whether the child was Male; (3) whether the child was Hispanic; (4) whether the child's native language was not English; (5) whether the child had an IEP placement; (6) the number of days elapsed between the Woodcock-Johnson testing date and the start of school (centered at the grand mean); and (7) the interaction term.

Appendix Table E-3
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Hispanic and Non-Hispanic Children

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
WJ Composite								
12-month bandwidth						<.001	12.92	
Hispanic	406.49	168	386.77	186	19.72			1.526
Non-Hispanic	412.43	1910	402.28	1880	10.15			.786
6-month bandwidth						<.001	12.66	
Hispanic	405.13	78	384.82	87	20.31			1.604
Non-Hispanic	412.63	931	402.66	971	9.97			.788
3 -month bandwidth						.011	11.94	
Hispanic	404.09	35	385.01	41	19.08			1.598
Non-Hispanic	412.50	484	402.24	490	10.26			.859
Letter-Word								
12-month bandwidth							21.15	
Hispanic	339.68	168	312.18	186	27.50	.038		1.300
Non-Hispanic	347.42	1910	325.89	1880	21.53			1.018
6-month bandwidth						.028	19.71	
Hispanic	339.93	78	309.28	87	30.65			1.555
Non-Hispanic	347.48	931	326.36	971	21.12			1.072
3 -month bandwidth						.151	19.17	
Hispanic	341.01	35	312.90	41	28.11			1.466
Non-Hispanic	345.87	484	326.62	490	19.25			1.004
Spelling								
12-month bandwidth						.192	20.82	
Hispanic	376.90	168	354.58	186	23.32			1.072
Non-Hispanic	379.77	1910	359.59	1880	20.18			.969
6-month bandwidth						.725	20.62	
Hispanic	376.06	78	354.83	87	21.23			1.030
Non-Hispanic	380.03	931	360.91	971	19.12			.927
3 -month bandwidth						.333	20.43	
Hispanic	370.84	35	355.03	41	15.81			.774
Non-Hispanic	379.87	484	359.84	490	20.03			.980
Oral Comprehension								
12-month bandwidth						.068	13.71	
Hispanic	443.15	168	436.93	186	6.22			.454
Non-Hispanic	452.33	1910	449.35	1880	2.98			.217
6-month bandwidth						.015	13.08	
Hispanic	441.53	78	433.41	87	8.12			.621
Non-Hispanic	452.15	931	449.66	971	2.49			.190
3 -month bandwidth						.303	12.87	
Hispanic	440.20	35	432.67	41	7.53			.585
Non-Hispanic	453.11	484	449.33	490	3.78			.294

Continued

Appendix Table E-3 (*continued*)
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Hispanic and Non-Hispanic Children

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Picture Vocabulary								
12-month bandwidth						<.001	10.72	
Hispanic	459.56	168	436.00	186	23.56			2.198
Non-Hispanic	465.34	1910	463.41	1880	1.93			.180
6-month bandwidth						<.001	10.89	
Hispanic	457.79	78	433.09	87	24.70			2.268
Non-Hispanic	465.61	931	463.64	971	1.97			.181
3 -month bandwidth						<.001	10.41	
Hispanic	460.06	35	435.02	41	25.04			2.405
Non-Hispanic	465.10	484	464.03	490	1.07			.103
Applied Problems								
12-month bandwidth						<.001	17.74	
Hispanic	402.93	168	375.90	186	27.03			1.524
Non-Hispanic	408.60	1910	401.88	1880	6.72			.379
6-month bandwidth						<001	18.12	
Hispanic	401.47	78	374.99	87	26.48			1.461
Non-Hispanic	410.16	931	401.69	971	8.47			.467
3 -month bandwidth						<.001	16.08	
Hispanic	398.78	35	370.07	41	28.71			1.785
Non-Hispanic	410.84	484	400.73	490	10.11			.629
Quantitative Concepts								
12-month bandwidth						.009	13.99	
Hispanic	416.64	168	405.05	186	11.59			.828
Non-Hispanic	421.12	1910	413.56	1880	7.56			.540
6-month bandwidth						.043	13.43	
Hispanic	414.02	78	403.07	87	10.95			.815
Non-Hispanic	420.39	931	413.62	589	6.77			.504
3 -month bandwidth						.371	12.49	
Hispanic	413.51	35	404.07	41	9.44			.756
Non-Hispanic	420.21	484	412.91	490	7.30			.584

Note. Only children eligible for free or reduced price lunch and who were either Hispanic or White were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. Previous analyses showed that the functional form was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Marginal means from the multilevel model.

[b] To facilitate comparison of these effect sizes with those reported for the overall pre-k effects, they were calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the treatment (TN-VPK) group.

[c] This is the *p*-value for the interaction term for condition (TN-VPK vs. No Pre-K) and whether the child was Hispanic or white. Estimates were based on a multilevel model with children nested within pre-k classroom. In addition to condition, Hispanic, and days from the age cutoff (centered at zero), covariates included: (1) Region; (2) whether the child was male; (3) whether the child was Black; (4) whether the child's native language was not English; (5) whether the child had an IEP placement; (6) the number of days elapsed between the Woodcock-Johnson testing date and the start of school (centered at the grand mean); and (7) the interaction term.

Appendix Table E-4
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Children Whose Native Language was English or Another Language

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
WJ Composite								
12-month bandwidth						<.001	12.92	
English	411.96	1914	401.89	1926	10.07			.779
Not English	410.12	164	389.00	140	21.12			1.635
6-month bandwidth						<.001	12.66	
English	412.04	936	402.14	980	9.90			.782
Not English	411.30	73	390.28	78	21.02			1.660
3 -month bandwidth						.003	11.94	
English	411.98	487	401.97	492	10.01			.838
Not English	409.09	32	388.96	39	20.13			1.686
Letter-Word								
12-month bandwidth						<.001	21.15	
English	346.55	1914	325.33	1926	21.22			1.003
Not English	348.35	164	316.85	140	31.50			1.489
6-month bandwidth						.028	19.71	
English	346.67	936	325.82	980	20.85			1.058
Not English	348.92	73	315.18	78	33.74			1.712
3 -month bandwidth						.016	19.17	
English	345.27	487	326.58	492	18.69			.975
Not English	347.17	32	314.80	39	32.37			1.689
Spelling								
12-month bandwidth						.006	20.82	
English	379.24	1914	359.40	1926	19.84			.953
Not English	382.60	164	356.12	140	26.48			1.272
6-month bandwidth						.044	20.62	
English	379.36	936	360.63	980	18.73			.908
Not English	384.19	73	358.21	78	25.98			1.260
3 -month bandwidth						.944	20.43	
English	379.29	487	359.57	492	19.72			.965
Not English	378.95	32	358.83	39	20.12			.985
Oral Comprehension								
12-month bandwidth						.213	13.71	
English	452.37	1914	449.27	1926	3.10			.226
Not English	440.85	164	435.86	140	4.99			.264
6-month bandwidth						.132	13.08	
English	451.98	936	449.31	980	2.67			.204
Not English	442.37	73	436.46	78	5.91			.452
3 -month bandwidth						.807	12.87	
English	452.94	487	448.98	492	3.96			.308
Not English	441.08	32	436.32	39	4.76			.370

Continued

Appendix Table E-4 (*continued*)
Estimated Impact of TN-VPK on the WJ Composite and Individual Tests: Children Whose Native Language was English or Another Language

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Picture Vocabulary								
12-month bandwidth						<.001	10.72	
English	465.04	1914	462.85	1926	2.20			.205
Not English	459.72	164	438.14	140	21.58			2.013
6-month bandwidth						<.001	10.89	
English	465.22	936	462.82	980	2.40			.220
Not English	460.68	73	441.32	78	19.36			1.778
3 -month bandwidth						<.001	10.41	
English	464.84	487	463.89	492	.95			.091
Not English	459.26	32	437.42	39	21.84			2.098
Applied Problems								
12-month bandwidth						<.001	17.74	
English	407.95	1914	401.26	1926	6.69			.377
Not English	407.54	164	379.24	140	28.30			1.595
6-month bandwidth						<001	18.12	
English	409.34	936	401.02	980	8.32			.459
Not English	409.88	73	381.93	78	27.95			1.542
3 -month bandwidth						<.001	16.08	
English	409.79	487	400.08	492	9.71			.604
Not English	409.20	32	379.66	39	29.54			1.837
Quantitative Concepts								
12-month bandwidth						.001	13.99	
English	420.61	1914	413.24	1926	7.37			..527
Not English	421.68	164	407.74	140	13.94			.996
6-month bandwidth						.003	13.43	
English	419.69	936	413.13	980	6.56			.488
Not English	422.02	73	408.68	78	13.34			.993
3 -month bandwidth						.123	12.49	
English	419.76	487	412.74	492	7.02			.562
Not English	419.01	32	406.69	39	12.32			.986

Note. Only children eligible for free or reduced price lunch were included. Previous analyses showed that the functional form was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Marginal means from the multilevel model.

[b] To facilitate comparison of these effect sizes with those reported for the overall pre-k effects, they were calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the treatment (TN-VPK) group.

[c] This is the *p*-value for the interaction term for condition (TN-VPK vs. No Pre-K) and whether the child's native language was English or not. Estimates were based on a multilevel model with children nested within pre-k classroom. In addition to condition, native language, and days from the age cutoff (centered at zero), covariates included: (1) Region; (2) whether the child was male; (3) whether the child was Black; (4) whether the child was Hispanic; (5) whether the child had an IEP placement; (6) the number of days elapsed between the Woodcock-Johnson testing date and the start of school (centered at the grand mean); and (7) the interaction term.

Appendix Table E-5
Estimated Impact of TN-VPK on the WJ Tests: Children Who Were or Were Not Hispanic and Whose Native Language was English or Another Language

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
WJ Composite								
12-month bandwidth						<.001	12.92	
Hispanic & English	408.5	49	387.3	72	21.18			1.64
Hispanic & not English	398.8	119	379.7	114	19.10			1.48
Not Hispanic & English	412.8	1865	403.0	1854	9.81			.76
Not Hispanic & not English	413.8	45	384.7	26	29.12			2.25
6-month bandwidth						<.001	12.66	
Hispanic & English	407.8	25	381.9	27	25.89			2.05
Hispanic & not English	397.7	53	379.9	60	17.81			1.41
Not Hispanic & English	412.9	911	403.4	953	9.53			.75
Not Hispanic & not English	416.3	20	384.6	18	31.70			2.50
Letter-Word								
12-month bandwidth						<.001	21.15	
Hispanic & English	342.63	49	313.85	72	28.78			1.361
Hispanic & not English	335.37	119	308.04	114	27.33			1.292
Not Hispanic & English	347.37	1865	326.29	1854	21.08			.997
Not Hispanic & not English	357.78	45	310.90	26	46.88			2.217
6-month bandwidth						.043	19.71	
Hispanic & English	340.08	25	310.35	27	29.73			1.508
Hispanic & not English	335.19	53	304.33	60	30.86			1.565
Not Hispanic & English	347.59	911	326.94	953	20.65			1.048
Not Hispanic & not English	355.37	20	311.89	18	43.48			2.206
Spelling								
12-month bandwidth						.202	20.82	
Hispanic & English	376.25	49	357.41	72	18.84			.905
Hispanic & not English	377.10	119	352.84	114	24.26			1.165
Not Hispanic & English	379.56	1865	359.68	1854	19.88			.955
Not Hispanic & not English	388.44	45	354.17	26	34.27			1.646
6-month bandwidth						<.001	20.62	
Hispanic & English	374.88	25	355.99	27	18.89			.916
Hispanic & not English	376.99	53	354.97	60	22.02			1.068
Not Hispanic & English	379.74	911	361.06	953	18.68			.906
Not Hispanic & not English	392.03	20	352.44	18	39.59			1.920

Continued

Appendix Table E-5 (*continued*)
Estimated Impact of TN-VPK on the WJ Tests: Children Who Were or Were Not Hispanic and Whose Native Language was English or Another Language

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Oral Comprehension								
12-month bandwidth						<.001	13.71	
Hispanic & English	447.48	49	435.13	72	12.35			.901
Hispanic & not English	430.20	119	426.93	114	3.27			.239
Not Hispanic & English	453.21	1865	450.33	1854	2.88			.210
Not Hispanic & not English	443.26	45	431.25	26	12.01			.876
6-month bandwidth						<.001	13.08	
Hispanic & English	449.21	25	427.74	27	21.47			1.641
Hispanic & not English	428.03	53	425.46	60	2.57			.196
Not Hispanic & English	452.87	911	450.62	953	2.25			.172
Not Hispanic & not English	448.04	20	431.09	18	16.95			1.296
Picture Vocabulary								
12-month bandwidth						<.001	10.72	
Hispanic & English	461.58	49	434.91	72	26.67			2.488
Hispanic & not English	444.62	119	422.94	114	21.68			2.022
Not Hispanic & English	466.35	1865	464.75	1854	1.60			.149
Not Hispanic & not English	456.84	45	434.40	26	22.44			2.093
6-month bandwidth						<.001	10.89	
Hispanic & English	463.27	25	424.01	27	39.26			3.605
Hispanic & not English	443.04	53	424.64	60	18.40			1.690
Not Hispanic & English	466.52	911	464.93	953	1.59			.146
Not Hispanic & not English	457.55	20	435.88	18	21.67			1.990
Applied Problems								
12-month bandwidth						<.001	17.74	
Hispanic & English	404.99	49	376.91	72	28.08			1.583
Hispanic & not English	391.46	119	364.93	114	26.53			1.495
Not Hispanic & English	409.17	1865	402.97	1854	6.20			.349
Not Hispanic & not English	410.17	45	374.48	26	35.69			2.012
6-month bandwidth						<.001	18.12	
Hispanic & English	403.86	25	371.66	27	32.20			1.777
Hispanic & not English	391.00	53	367.20	60	23.70			1.308
Not Hispanic & English	410.60	911	402.82	953	7.78			.429
Not Hispanic & not English	415.96	20	374.14	18	41.82			2.308

Continued

Appendix Table E-5 (*continued*)
Estimated Impact of TN-VPK on the WJ Tests: Children Who Were or Were Not Hispanic and Whose Native Language was English or Another Language

Outcome and Bandwidth	TN-VPK		No Pre-K		TN-VPK vs. No Pre-K Difference	*p*-value[c]	TN-VPK Standard Deviation	Effect Size[b]
	Mean[a]	*N*	Mean[a]	*N*				
Quantitative Concepts								
12-month bandwidth						<.001	13.99	
Hispanic & English	418.04	49	405.75	72	12.29			.878
Hispanic & not English	414.00	119	402.59	114	11.41			.928
Not Hispanic & English	421.11	1865	413.84	1854	7.27			.520
Not Hispanic & not English	426.82	45	403.27	26	23.55			1.916
6-month bandwidth						<.001	13.43	
Hispanic & English	415.64	25	401.53	27	14.11			1.075
Hispanic & not English	411.83	53	402.36	60	9.47			.705
Not Hispanic & English	420.30	911	413.93	953	6.37			.474
Not Hispanic & not English	429.11	20	402.58	18	26.53			1.975

Note. Only children eligible for free or reduced price lunch were included. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. Previous analyses showed that the functional form was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

[a] Marginal means from the multilevel model.

[b] To facilitate comparison of these effect sizes with those reported for the overall pre-k effects, they were calculated by dividing the TN-VPK vs. No Pre-K difference by the standard deviation of the treatment (TN-VPK) group.

[c] This is the *p*-value for the interaction term for condition (TN-VPK vs. No Pre-K), whether the child was Hispanic, and whether the child's native language was English or another language. Estimates were based on a multilevel model with children nested within pre-k classroom. In addition to condition, Hispanic, Language, and days from the age cutoff (centered at zero), covariates included: (1) Region; (2) whether the child was male; (3) whether the child was Black; (4) whether the child had an IEP placement; (5) the number of days elapsed between the Woodcock-Johnson testing date and the start of school (centered at the grand mean); and (6) the interaction term.

Appendix Table F-1
Statistical Models for the Effects of TN-VPK on the WJ Composite: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Composite 12 mo Bandwidth			WJ Composite 6 mo Bandwidth			WJ Composite 3 mo Bandwidth		
	b	**SE**	***p***	***b***	**SE**	***p***	***b***	**SE**	***p***
Intercept	410.39	.82	<.001	411.75	1.09	<.001	413.03	1.46	<.001
Days from age cutoff	.04	<.00	<.001	.04	<.01	<.001	.04	.01	.008
Region: West	-1.37	.83	.102	-1.53	1.02	.135	-2.63	1.30	.044
Region: Central East	-.04	.84	.960	-.15	1.04	.889	-1.59	1.33	.234
Region: East	-.09	.86	.920	-.05	1.05	.962	-.37	1.35	.787
Male	-2.24	.35	<.001	-1.56	.51	.003	-2.22	.72	.003
Black	-3.35	.54	<.001	-2.57	.74	<.001	-1.73	.98	.076
Hispanic	-11.97	.91	<.001	-13.46	1.39	<.001	-14.22	1.98	<.001
Native language not English	-6.51	1.02	<.001	-6.57	1.48	<.001	-7.65	2.09	<.001
Test lag	.12	.02	<.001	.16	.03	<.001	.16	.04	<.001
Has an IEP	-10.08	.60	<.001	-10.27	.85	<.001	-10.69	1.20	<.001
Eligible for FRPL	-4.50	.46	<.001	-5.98	.65	<.001	-6.42	.90	<.001
TN-VPK participation	**10.90**	.70	**<.001**	**10.82**	1.01	**<.001**	**10.30**	1.45	**<.001**

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The composite is an average of the W-scores for the six tests used to measure outcomes. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates are based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table F-2
Statistical Models for the Effects of TN-VPK on WJ Letter-Word: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Letter-Word 12 mo Bandwidth			WJ Letter-Word 6 mo Bandwidth			WJ Letter-Word 3 mo Bandwidth		
	B	**SE**	***p***	***b***	**SE**	***p***	***b***	**SE**	***p***
Intercept	336.30	1.44	<.001	338.69	1.87	<.001	340.76	2.58	<.001
Days from age cutoff	.04	<.01	<.001	.04	.01	<.001	.06	.02	.013
Region: West	-.67	1.46	.646	-1.58	1.77	.374	-2.56	2.36	.278
Region: Central East	-1.32	1.48	.372	-1.93	1.79	.283	-3.51	2.41	.147
Region: East	-1.79	1.51	.238	-1.71	1.82	.349	-2.13	2.44	.385
Male	-3.98	.61	<.001	-3.60	.86	<.001	-5.40	1.24	<.001
Black	1.02	.95	.283	3.27	1.26	.010	3.75	1.71	.029
Hispanic	-11.24	1.59	<.001	-12.16	2.37	<.001	-9.96	3.41	.004
Native language not English	-2.29	1.77	.196	-4.14	2.52	.101	-5.00	3.60	.166
Test lag	.23	.04	<.001	.29	.05	<.001	.32	.07	<.001
Has an IEP	-10.20	1.05	<.001	-11.28	1.45	<.001	-11.91	2.07	<.001
Eligible for FRPL	-7.15	.81	<.001	-9.44	1.10	<.001	-9.50	1.55	<.001
TN-VPK participation	**21.66**	1.22	**<.001**	**21.26**	1.71	**<.001**	**19.09**	2.50	**<.001**

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates are based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table F-3
Statistical Models for the Effects of TN-VPK on WJ Spelling: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Spelling 12 mo Bandwidth			WJ Spelling 6 mo Bandwidth			WJ Spelling 3 mo Bandwidth		
	b	**SE**	***p***	***b***	**SE**	***p***	***b***	**SE**	***p***
Intercept	368.55	1.26	<.001	370.56	1.77	<.001	371.74	2.39	<.001
Days from age cutoff	.05	<.01	<.001	.06	.01	<.001	.05	.02	.033
Region: West	1.09	1.22	.376	.55	1.66	.742	-1.06	2.12	.618
Region: Central East	.46	1.24	.711	-.06	1.68	.974	-2.29	2.18	.295
Region: East	1.88	1.27	.142	2.37	1.71	.167	1.81	2.21	.415
Male	-6.20	.57	<.001	-5.03	.82	<.001	-5.75	1.18	<.001
Black	-1932	.85	.024	-.85	1.20	.480	1.65	1.60	.303
Hispanic	-4.84	1.47	.001	-5.87	2.25	.009	-7.55	3.25	.020
Native language not English	.23	1.64	.886	1.07	2.40	.655	1.04	3.44	.762
Test lag	.21	.03	<.001	.24	.05	<.001	.27	.06	<.001
Has an IEP	-11.07	.97	<.001	-11.99	1.38	<.001	-11.81	1.98	<.001
Eligible for FRPL	-5.35	.74	<.001	-7.12	1.05	<.001	-8.28	1.48	<.001
TN-VPK participation	**20.35**	1.14	**<.001**	**20.00**	1.63	**<.001**	**19.87**	2.39	**<.001**

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table F-4
Statistical Models for the Effects of TN-VPK on WJ Oral Comprehension: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Oral Comprehension 12 mo Bandwidth			WJ Oral Comprehension 6 mo Bandwidth			WJ Oral Comprehension 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	457.15	.85	<.001	458.22	1.11	<.001	459.27	1.39	<.001
Days from age cutoff	.03	<.01	<.001	.03	<.01	<.001	.02	.01	.103
Region: West	-3.47	.87	<.001	-2.92	1.06	.007	-3.52	1.21	.004
Region: Central East	.95	.88	.285	1.37	1.07	.203	.24	1.24	.848
Region: East	-.81	.90	.374	-.59	1.09	.588	-.73	1.26	.560
Male	-1.42	.35	<.001	-.71	.50	.159	-.63	.70	.372
Black	-5.45	.55	<.001	-5.35	.74	<.001	-5.68	.94	<.001
Hispanic	-11.54	.92	<.001	-14.08	1.38	<.001	-14.60	1.92	<.001
Native language not English	-12.07	1.03	<.001	-11.30	1.47	<.001	-11.55	2.03	<.001
Test lag	.05	.02	.018	.10	.03	<.001	.08	.04	.026
Has an IEP	-9.17	.60	<.001	-8.36	.84	<.001	-9.33	1.17	<.001
Eligible for FRPL	-2.98	.47	<.001	-4.49	.65	<.001	-4.75	.87	<.001
TN-VPK participation	**3.86**	.71	**<.001**	**3.50**	1.00	**<.001**	**3.71**	1.42	**.009**

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table F-5
Statistical Models for the Effects of TN-VPK on WJ Picture Vocabulary: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Picture Vocabulary 12 mo Bandwidth			WJ Picture Vocabulary 6 mo Bandwidth			WJ Picture Vocabulary 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	467.90	.78	<.001	468.58	1.04	<.001	470.51	1.37	<.001
Days from age cutoff	.02	<.01	<.001	0.02	<.01	<.001	.02	.01	.139
Region: West	-1.75	.78	.026	-1.32	.94	.162	-2.10	1.16	.074
Region: Central East	-.44	.79	.578	-0.34	.96	.724	-1.73	1.20	.151
Region: East	-.58	.81	.472	-1.03	.97	.289	-1.60	1.21	.190
Male	0.70	.35	.043	1.32	0.50	.009	1.10	.71	.119
Black	-2.11	.53	<.001	-2.42	0.71	<.001	-2.56	.93	.006
Hispanic	-18.58	.90	<.001	-20.34	1.37	<.001	-20.60	1.92	<.001
Native language not English	-13.49	1.00	<.001	-12.83	1.46	<.001	-15.50	2.03	<.001
Test lag	.03	.02	.187	0.03	0.03	.352	.01	.03	.725
Has an IEP	-7.36	.59	<.001	-6.98	0.84	<.001	-6.72	1.17	<.001
Eligible for FRPL	-2.40	.46	<.001	-3.06	.64	<.001	-3.73	.869	<.001
TN-VPK participation	**3.40**	.70	<.001	**3.26**	1.00	**.002**	2.65	1.42	.063

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table F-6
Statistical Models for the Effects of TN-VPK on WJ Applied Problems: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Applied Problems 12 mo Bandwidth			WJ Applied Problems 6 mo Bandwidth			WJ Applied Problems 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	412.37	1.27	<.001	413.53	1.58	<.001	415.18	2.13	<.001
Days from age cutoff	.05	<.01	<.001	0.04	0.01	<.001	0.04	0.02	.045
Region: West	-2.52	1.25	.045	-2.59	1.40	.068	-3.47	1.86	.064
Region: Central East	-.33	1.26	.793	0.36	1.43	.799	-.86	1.91	.654
Region: East	.27	1.30	.834	-0.52	1.45	.719	-1.19	1.93	.534
Male	-1857	.56	.001	-0.76	0.78	.327	-2.13	1.07	.047
Black	-8.59	.85	<.001	-8.49	1.09	<.001	-7.31	1.43	<.001
Hispanic	-18.67	1.46	<.001	-20.16	2.12	<.001	-24.36	2.93	<.001
Native language not English	-9.50	1.63	<.001	-9.95	2.25	<.001	-11.20	3.09	<.001
Test lag	.12	.03	<.001	0.16	0.04	<.001	0.11	0.05	.036
Has an IEP	-14.30	.96	<.001	-14.03	1.30	<.001	-14.88	1.78	<.001
Eligible for FRPL	-4.92	.74	<.001	-6.25	.98	<.001	-6.83	1.33	<.001
TN-VPK participation	**8.53**	1.13	**<.001**	**9.93**	1.54	**<.001**	**9.52**	2.15	**<.001**

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table F-7
Statistical Models for the Effects of TN-VPK on WJ Quantitative Concepts: Sample of Children Regardless of Eligibility for Free or Reduced Price Lunch (FRPL)

Independent Variable	WJ Quantitative Concepts 12 mo Bandwidth			WJ Quantitative Concepts 6 mo Bandwidth			WJ Quantitative Concepts 3 mo Bandwidth		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Intercept	419.91	.83	<.001	420.93	1.10	<.001	421.03	1.47	<.001
Days from age cutoff	.03	<.01	<.001	0.04	<.01	<.001	.03	.01	.026
Region: West	-.79	.83	.344	-1.19	1.02	.247	-2.85	1.27	.027
Region: Central East	-.03	.84	.970	-.24	1.04	.820	-1.51	1.31	.251
Region: East	.61	.86	.476	.97	1.06	.364	1.30	1.33	.328
Male	-.74	.36	.040	-0.56	0.52	.278	-.49	.74	.511
Black	-3.06	.55	<.001	-1.92	0.75	.011	-.78	.99	.430
Hispanic	-7.15	.93	<.001	-8.59	1.42	<.001	-8.56	2.03	<.001
Native language not English	-1.95	1.04	.061	-2.07	1.51	.170	-3.53	2.14	.100
Test lag	.09	.02	<.001	0.13	0.03	<.001	.15	.04	<.001
Has an IEP	-8.33	.61	<.001	-8.78	.87	<.001	-9.24	1.23	<.001
Eligible for FRPL	-4.06	.47	<.001	-5.41	.66	<.001	-5.45	.92	<.001
TN-VPK participation	**7.55**	.72	**<.001**	**7.02**	1.03	<.001	**6.82**	1.49	**<.001**

Note. Sample sizes are 5,188 for the 12-month bandwidth, 2,602 for the 6-month bandwidth, and 1,334 for the 3-month bandwidth. The race/ethnicity categories are not mutually exclusive; mixed race children are coded in all the categories for their respective mix. IEP indicates special education placement; test lag is time between start of school and testing. Estimates were based on a multilevel model with children nested within pre-k classroom. Previous analyses showed that the functional form for all models was linear. Analyses were weighted to represent the statewide population of TN-VPK classrooms.

Appendix Table G-1
Effect Size Estimates from Age-Cutoff Regression-Discontinuity Studies of State and Locally-Funded Pre-K Programs

State or Locality	Study	Measure	Outcome			
			Literacy	Language	Math	Other
Arkansas	Jung, Barnett, Hustedt, & Francis (2013)	CTOPPP Print Awareness	.82*			
		PPVT		.28*		
		WJ Applied Problems			.33*	
Boston, Massachusetts	Weiland & Yoshikawa (2013)	WJ Letter-Word Identification	.62*			
		PPVT		.44*		
		WJ Applied Problems			.59*	
		Early Math Assessment			.43*	
		Pencil Tap				.21*
		Backward Digit Span				.24*
		Forward Digit Span				.24*
		Dimensional Card Sort				.28*
		TOQ Attention				.11*
California	Barnett, Howes, & Jung (2009)	CTOPPP Print Awareness	1.17*			
		PPVT		.30*		
		WJ Applied Problems			.38*	
Connecticut	Connecticut Academy of Science and Engineering (2016)	WJ Letter-Word Identification & Word Attack (composite)	.69*			
		PPVT		.05		
		WJ Oral Comprehension		.31		
		WJ Applied Problems, Calculations, & Math Fluency (composite)			.48*	
		Behavior Assessment Scale for Children (BASC-3)				n.s
Georgia	Peisner-Feinberg, Schaaf, LaForett, Hildebrandt, & Sideris (2014)	Letter Knowledge	.89*			
		WJ Letter-Word Identification	1.05*			
		WJ Sound Awareness	.59*			
		WJ Word Attack	1.20*			
		WJ Picture Vocabulary		.01		
		WJ Applied Problems			.51*	
		Counting			.86*	
		Social Awareness				.43*
		SSiS Social Skills				.23
		SSiS Problem Behavior				.10

continued

Appendix Table G-1 (continued*)*
Effect Size Estimates from Age-Cutoff Regression-Discontinuity Studies of State and Locally-Funded Pre-K Programs

State or Locality	Study	Measure	Outcome			
			Literacy	Language	Math	Other
Kalamazoo County Michigan	Bartik (2013)	WJ Letter-Word Identification	.25			
		PPVT		.60*		
		WJ Applied Problems			.70*	
		Devereux Early Childhood Assessment Protective Factors				.51
		Devereux Early Childhood Assessment Behavioral Concerns				-.28
Los Angeles, California	Kyger & Barnhart (2017)	DIBELS Letter Naming Fluency	.83*			
		DIBELS Phoneme Segmentation Fluency		.76*		
		DIBELS Initial Sound Fluency		.43		
		IGDis-EN Oral Counting			.08	
		IGDis-EN Quantity Comparison			-.50	
		IGDis-EN Correspondence Counting			1.83	
Massachusetts	Hofer, Checkoway, Goodson, & Nichols (2018)	WJ Letter-Word Identification	.92*			
		PPVT		.21*		
		WJ Applied Problems			.45*	
		Executive Functioning (hearts & flowers task)				.05
Michigan	Wong, Cook, Barnett, & Jung (2008)	CTOPPP Print Awareness	1.09*			
		PPVT		-.13		
		WJ Applied Problems			.53*	
New Jersey	Wong, Cook, Barnett, & Jung (2008)	CTOPPP Print Awareness	.32*			
		PPVT		.36*		
		WJ Applied Problems			.23*	
New Mexico	Hustedt, Barnett, Jung, & Friedman (2010)	ELSA Early Literacy	1.30*			
		PPVT		.24*		
		WJ Applied Problems			.37*	
North Carolina	Peisner-Feinberg & Schaaf (2011)	Phonological Awareness (TOPEL)	.56*			
		Print Knowledge (TOPEL)	1.16*			
		WJ Letter-Word Identification	1.14*			
		PPVT		.06		
		WJ Applied Problems			.34*	
		Counting			.81*	
Oklahoma	Wong, Cook, Barnett, & Jung (2008)	CTOPPP Print Awareness	.42			
		PPVT		.28*		
		WJ Applied Problems			.34	

continued

Appendix Table G-1 (*continued)*
Effect Size Estimates from Age-Cutoff Regression-Discontinuity Studies of State and Locally-Funded Pre-K Programs

State or City	Study	Measure	Outcome			
			Literacy	Language	Math	Other
San Francisco	Applied Survey Research (2013)	WJ Letter-Word Identification	.40*			
		Receptive One-Word Picture Vocabulary (ROWPVT)		.29		
		WJ Applied Problems			.40*	
		Self-Regulation (HTKS)				.51*
South Carolina	Wong, Cook, Barnett, & Jung (2008)	CTOPPP Print Awareness	.78*			
		PPVT		.04		
Tennessee	Pion & Lipsey (2021)	WJ Composite				.83*
		WJ Letter-Word Identification	.99*			
		WJ Spelling	.97*			
		WJ Oral Comprehension		.28*		
		WJ Picture Vocabulary		.32*		
		WJ Applied Problems			.47*	
		WJ Quantitative Concepts			.52*	
		Note: Estimates for full sample				
Tennessee	Coburn (2009)	Brigance Preschool & K-1 Screen				NA
Tulsa, Oklahoma	Gormley, Gayer, Philips, & Dawson (2005)	WJ Letter-Word Identification	.79*			
		WJ Spelling	.64*			
	Gormley & Gayer (2005)	WJ Applied Problems			.38*	
	Gormley, Phillips, & Gayer (2008)	Cognitive/knowledge				.39*
		Language		.38*		
	Phillips & Meloy (2012)	Motor Skills				.24*
	Gormley (2008)	Social-emotional				n.s.
Virginia	Huang (2017)	PALS Lowercase Letter Recognition	.95*			
		Note: Midpoint of reported range				
Virginia	Gaylor, Golan, Chow, Grindal, Mercier, Williamson, & Tiruke (2019)	PALS Letter Sounds	1.04*			
		PALS Lowercase Letter Recognition	1.04*			
		PPVT		.23*		
		WJ Applied Problems			.35*	
		Self-regulation (HTKS)				.31*
		Note: Mean of 2 estimates				
West Virginia	Wong, Cook, Barnett, & Jung (2008)	CTOPPP Print Awareness	.92*			
		PPVT		.16		
		WJ Applied Problems			.06	

Notes. CTOPPP= Comprehensive Test of Phonological and Print Processing. PALS=Phonological Awareness Literacy Screening. PPVT=Peabody Picture Vocabulary Test. SSiS=Social Skills Improvement System. TOPEL=Test of Preschool Early Literacy. TOQ=Task Orientation Questionnaire. WJ=Woodcock Johnson III Achievement Battery.
* $p<.05$

Appendix G-2: References for Studies in Table G-1

Arkansas

Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). *Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade*. (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Hustedt, J.T., Barnett, W.S., Jung, K., & Thomas, J. (2007). The effects of the Arkansas Better Chance Program on young children's school readiness. (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Boston, Massachusetts

Weiland, C. & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84(6),* 2112–2130.

Weiland, C. (undated). *Impacts of the Boston prekindergarten program on the school readiness of young children with special needs*. Manuscript.

California

Barnett, W. S., Howes, C., & Jung, K. (2009). *California's State Preschool Program: Quality and Effects on Children's Cognitive Abilities at Kindergarten Entry* (Report to the California Children and Families Commission). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Connecticut

Connecticut Academy of Science and Engineering. (2016). *Early Childhood Regression Discontinuity Study*.(Report for the Connecticut General Assembly and the Education Committee). Rocky Hill, CT: Connecticut Academy of Science and Engineering.

Georgia

Peisner-Feinberg, E. S., Schaaf, J. M., LaForett, D. R., Hildebrandt, L.M., & Sideris, J. (2014). *Effects of Georgia's Pre-K Program on Children's School Readiness Skills: Findings from the 2012-2013 Evaluation Study*. Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Kalamazoo County, Michigan

Bartik, T. J. (2013). *Effects of the Pre-K Program of Kalamazoo County Ready 4s on Kindergarten Entry Test Scores: Estimates Based on Data from the Fall of 2011 and the Fall of 2012* (Upjohn Institute Working Paper 13-198). Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Los Angeles, California

Kyger, M., & Barnhart, M. (2017). *Effects of High-Quality Preschool on Early Literacy and Mathematics Skills—A Regression-Discontinuity Approach.* Los Angeles, CA: Los Angeles Universal Preschool, Research and Evaluation.

Massachusetts

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.

Michigan

Lamy, C., Barnett, W. S., & Jung, K. (2005a). *The Effects of the Michigan School Readiness Program on Young Children's Abilities at Kindergarten Entry* (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154. doi:10.1002/pam.2031

New Jersey

Lamy, C., Barnett, W. S., & Jung, K. (2005). *The Effects of New Jersey's Abbott Preschool Program on Young Children's School Readiness* (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154. doi:10.1002/pam.2031

New Mexico

Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010). *The New Mexico Pre-k Evaluation: Impacts from the Fourth Year (2008-2009) of New Mexico's State-funded Pre-k Program* (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

North Carolina

Peisner-Feinberg, E. S., & Schaaf, J. M. (2011). *Summary of Key Findings: Effects of the North Carolina More at Four Prekindergarten Program on Children's School Readiness Skills*. Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Oklahoma

Lamy, C., Barnett, W. S., & Jung, K. (2005). *The Effects of Oklahoma's Early Childhood Four-year-old Program on Young Children's School Readiness* (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154. doi:10.1002/pam.2031

San Francisco

Applied Survey Research (2013, August). *Evaluating Preschool for All Effectiveness*. Research Brief. San Jose, CA: Applied Survey Research.

South Carolina

Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of South Carolina's Early Childhood Programs on Young Children's School Readiness* (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154. doi:10.1002/pam.2031

Tennessee

Pion, G. M., & Lipsey, M. W. (2021). Impact of the Tennessee Voluntary Prekindergarten Program on children's literacy, language, and mathematics skills: Results from a regression-discontinuity design. *AERA Open*, this issue.

Coburn, J.L. (2009). *The Effects of Tennessee's Prekindergarten Programs on Young Children's School Readiness Skills: A Regression-Discontinuity Design.* Dissertation. Tennessee Technological University.

Tulsa, Oklahoma

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology, 41(6),* 872-884.

Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program. *The Journal of Human Resources*, *40(3),* 533-558.

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320(June 27)*, 1723-1724.

Phillips, D.A., & Meloy, M.C. (2012). High-quality school-based pre-k can boost early learning for children with special needs. *Exceptional Children, 78(4),* 471-490.

Gormley, W.T. (2008). The effects of Oklahoma's pre-k program on Hispanic children. *Social Science Quarterly, 89(4)*, 916-936.

Virginia

Huang, F. L. (2017). Does attending a state-funded preschool program improve letter name knowledge? *Early Childhood Research Quarterly, (38),* 116-126.

Gaylor, E., Golan, S., Chow, K., Grindal, T., Mercier, B., Williamson, C., & Tiruke, T. (2019). *Virginia Preschool Initiative Plus (VPI+): Final evaluation report*. Prepared for The Virginia Department of Education. Menlo Park, CA: SRI International.

West Virginia

Lamy, C., Barnett, W. S., & Jung, K. (2005). *The Effects of West Virginia's Early Education Program on Young Children's School Readiness* (Research Report). New Brunswick, NJ. Rutgers University, National Institute for Early Education Research.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154. doi:10.1002/pam.2031