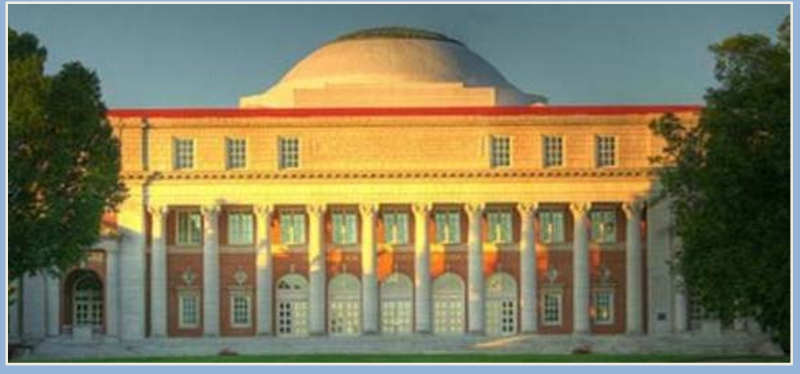


Evaluation of the Tennessee Voluntary Prekindergarten Program: End of Pre-K Results from the Randomized Control Design

Research Report

Mark W. Lipsey, Ph.D.
Kerry G. Hofer, Ph.D.
Nianbo Dong, Ph.D.
Dale C. Farran, Ph.D.
Carol Bilbrey, Ph.D.

May 2013

PEABODY research institute

VANDERBILT PEABODY COLLEGE

Director
Mark W. Lipsey, Ph. D.

Senior Associate Director
Dale C. Farran, Ph. D.

Associate Director
Sandra Jo Wilson, Ph. D.

Peabody Research Institute

Vanderbilt University

The mission of the Peabody Research Institute is to conduct research aimed at improving the effectiveness of programs for children, youth, and families. Using field research, program evaluation, and research synthesis (meta-analysis), our faculty and staff help determine which programs are actually making a difference in the lives of the people they serve.

Recommended Citation:

Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013). *Evaluation of the Tennessee Voluntary Prekindergarten Program: End of Pre-K Results from the Randomized Control Design* (Research Report). Nashville, TN: Vanderbilt University, Peabody Research Institute.

Funding Source:

The research reported here was supported by Grant #R305E090009 from the Institute of Education Sciences, U. S. Department of Education for the study titled "Evaluating the Effectiveness of Tennessee's Voluntary Pre-Kindergarten Program."

Contact Us:
Phone: 615.322.8540
Fax: 615.322.0293

Mailing Address:
Peabody Research Institute
230 Appleton Place
PMB 181
Nashville, TN 37203-5721

Delivery Address:
Peabody Research Institute
1930 South Drive
Room 410A
Nashville, TN 37212

http://peabody.vanderbilt.edu/research/pri

Staff and Contact Information

Peabody Research Institute, Vanderbilt University

Principal Investigator: Mark W. Lipsey, Director, Peabody Research Institute; Mark.Lipsey@vanderbilt.edu.
Co-Principal Investigator: Dale C. Farran, Senior Associate Director, Peabody Research Institute, and Professor, Department of Teaching and Learning; Dale.Farran@vanderbilt.edu.
Program Manager: Carol Bilbrey, Research Associate; Carol.Bilbrey@vanderbilt.edu.
Data Managers: Kerry G. Hofer, Senior Research Associate, and Nianbo Dong, Research Associate.
Project Coordinators: Janie Hughart and Patricia Abelson.
Regional Coordinator: Diane Hughes.
Research Analysts: Chris Ham, Rick Feldser, Ilknur Sekmen, and Laura Kerr.
Doctoral Student: Sascha Mowrey.
Assessors and Observers across Tennessee.

With thanks for assistance from the Tennessee Department of Education

Bobbi Lussier, Executive Director, Office of Student Teaching/Teacher Licensure, Middle Tennessee State University; former Assistant Commissioner of Special Populations.
Connie Casha, Director, Office of Early Learning, Division of Curriculum and Instruction; Connie.Casha@tn.gov.
Robert Taylor, Consultant and former Superintendent of Bradley County Schools; taylor_robtl@yahoo.com.

Evaluation of the Tennessee Voluntary Prekindergarten Program: End of Pre-K Results from the Randomized Control Design

Table of Contents

List of Tables

List of Figures

Evaluation of the Tennessee Voluntary Prekindergarten Program: End of Pre-K Results from the Randomized Control Design

Executive Summary

In 2009, Vanderbilt University's Peabody Research Institute, with the assistance of the Tennessee Department of Education's Division of Curriculum and Instruction, initiated a rigorous, independent evaluation of the state's Voluntary Prekindergarten program (TN-VPK). TN-VPK is a full-day prekindergarten program for four-year-old children expected to enter kindergarten the following school year. The program in each participating school district must meet standards set by the State Board of Education that require each classroom to have a teacher with a license in early childhood development and education, an adult-student ratio of no less than 1:10, a maximum class size of 20, and an approved age-appropriate curriculum.

TN-VPK is an optional program that is focused on the neediest children in the state. It uses a tiered admission process with children from low-income families who apply to the program admitted first. If there are remaining seats in a given location, they are allocated to otherwise at-risk children including those with disabilities and limited English proficiency.

The evaluation was funded by a grant from the U. S. Department of Education's Institute of Education Sciences. It was designed to determine whether the children who participate in the TN-VPK program make greater academic and behavioral gains in areas that prepare them for later schooling than comparable children who do not participate in the program.

Do children who participate in the TN-VPK program make greater academic and behavioral gains than comparable children who do not participate in the program?

Research design. There are several different parts of the research design for this evaluation. The part reported here, and the one that provides the strongest test of the effects of TN-VPK, is a randomized control trial (RCT) in which children applying to TN-VPK are admitted to the program on a random basis. The TN-VPK programs that participated in the RCT were among those where more eligible children were expected to apply for the program than there were seats available. Under such circumstances, only some of the applicants can be admitted and, of necessity, some must be turned away. The participating programs agreed to make this decision on the basis of chance, a process rather like randomly selecting names out of a hat, to determine which children would be admitted. This procedure treats every applicant equally and, as a result, no differences are expected on average between the characteristics of those children admitted and those not admitted. Comparing their academic and behavioral outcomes at the end of the school year, then, provides a direct indication of the effects of the TN-VPK program on the children who were admitted.

To implement this procedure, TN-VPK programs across Tennessee that expected more applicants than they could accommodate and were willing to participate in the evaluation submitted lists of eligible applicants to the researchers at the Peabody Research Institute. The research team then shuffled each list into a random order and the TN-VPK program staff were asked to fill the available seats by first offering admission to the child at the top of the list and then going down the list in order until all the available seats were filled. Once a program had admitted enough children to fill its seats, any remaining children were put on a waiting list and were admitted, in order, if an additional seat became available. Those on the waiting list who were not admitted to TN-VPK became the control group for the study.

This procedure was used for two cohorts of children, TN-VPK applicants for the 2009-10 and 2010-11 school years, and resulted in more than 3000 randomly assigned children. Both the children who participated in TN-VPK and those who did not participate will be tracked through the state education database until their third grade year. Information from that database on attendance, disciplinary actions, special education placements, grade retention, and state standardized test scores is being collected each year to determine the long-term impact of the TN-VPK program. In addition, we obtained parental consent for a portion of this randomized sample, referred to as the Intensive Substudy. Children in the Intensive Substudy, including some who participated in TN-VPK and some who did not, are being directly assessed by the research team and rated by their teachers in each year of the study.

This report presents the findings from the Intensive Substudy portion of the randomized control design for the outcomes at the end of the prekindergarten year. The two cohorts of consented children in the Intensive Substudy provided a combined sample of more than 1,000 children who applied to 58 TN-VPK programs across 21 school districts in Tennessee. Though the consent procedures differed for the two cohorts and resulted in different participation rates in the Intensive Substudy, the two cohorts were similar in terms of demographic characteristics and initial academic skills. Likewise, although the consent rate for the children admitted to TN-VPK was higher than that for the children who were not admitted, there were few significant differences between them at the time of pretest. To ensure that no baseline differences between children who participated in TN-VPK and those who did not could bias the estimates of the effects of TN-VPK on the outcome measures, the baseline variables were used as statistical controls in the analyses. Using this large Intensive Substudy sample, two questions were investigated for this report:

1. Does participation in TN-VPK improve the school readiness of the economically disadvantaged children eligible for the program?

2. What are the characteristics of the children who benefit the most from TN-VPK?

Outcome measures. The academic gains of the children in the Intensive Substudy sample were measured with a selection of standardized tests from the Woodcock Johnson III Achievement Battery that were individually administered at the beginning and end of the school year. These tests assessed early literacy, language, and math skills:

Literacy

- *Letter-Word Identification*: Assesses the ability to identify and pronounce alphabet letters and read words.
- *Spelling*: Assesses prewriting skills, such as drawing lines and tracing, writing letters, and spelling orally presented words.

Language

- *Oral Comprehension*: Assesses listening ability and understanding by asking the child to complete analogies and provide words with similar or different meanings from key words.
- *Picture Vocabulary*: Assesses early language development and lexical knowledge by asking the child to name objects presented in pictures and point to the picture that goes with a word.

Math

- *Applied Problems*: Assesses the ability to solve small numerical and spatial problems presented verbally with accompanying pictures of objects.
- *Quantitative Concepts*: Assesses quantitative reasoning and math knowledge by asking the child to point to or state answers to questions on number identification, sequencing, shapes, symbols, and the like.

The scores on all of the above tests were summarized in a composite measure (WJ Composite) that averaged them together to create an overall measure representing children's combined achievement in literacy, language, and math.

In addition, reports of the children's academic skills and behavior were collected from their kindergarten teachers early in the fall of the school year after pre-k. Two teacher rating instruments were used for this purpose:

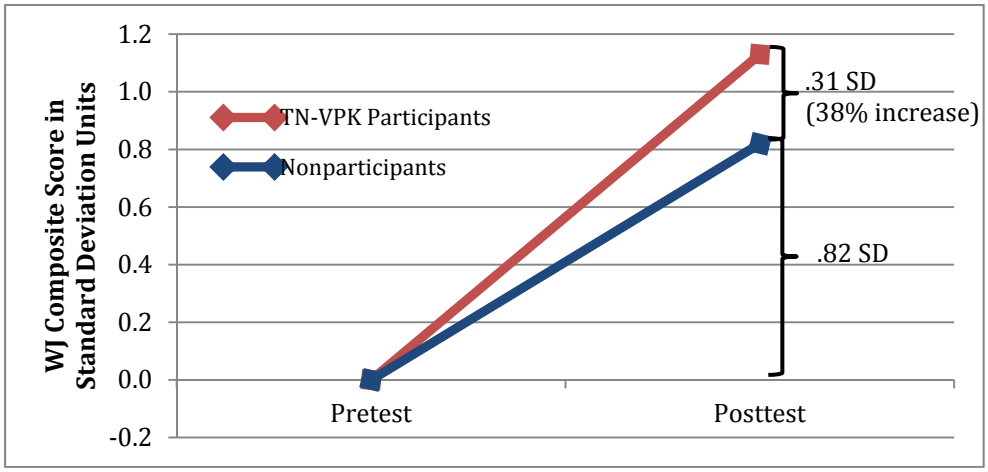
- *Cooper-Farran Behavioral Rating Scales*: Teacher ratings for each child on two scales:
  - *Work-Related Skills*: The ability to work independently, listen to the teacher, remember and comply with instructions, complete tasks, function within designated time periods, and otherwise engage appropriately in classroom activities.
  - *Social Behavior*: Social interactions with peers including appropriate behavior while participating in group activities, play, and outdoor games; expression of feelings and ideas; and response to others' mistakes or misfortunes.
- *Academic Classroom and Behavior Record*: Teacher ratings for each child on three scales:
  - *Readiness for Kindergarten*: How well prepared the child is for kindergarten in literacy, language, and math skills, and social behavior.
  - *Liking for School*: The child's liking or disliking for school, having fun at school, enjoying and engaging in classroom activities, and seeming happy at school.
  - *Behavior Problems*: Whether the child has shown explosive or overactive behaviors, attention problems, physical or relational aggression, social withdrawal or anxiety, motor difficulties, and the like.

Findings. All the children included in the Intensive Substudy sample met the requirement that they qualify for the federal Free or Reduced Price Lunch programs. The children who participated in TN-VPK attended an average of 149 days during the school year. In contrast, more than half of the children who were not admitted to TN-VPK stayed home with a parent or other guardian and only 27% were enrolled in Head Start or private center-based childcare.

Though the skills of all children improved over the course of the pre-k year, TN-VPK resulted in an average 38% improvement in gain on academic skills over and above that made by children who did not attend the program.

During the course of the pre-k school year, the academic skills of all the children improved. However, the children who participated in TN-VPK gained significantly more on all the direct assessments of academic skills than the children who did not attend. In standard deviation units, the standardized mean difference effect size for the WJ Composite scale was .31, and the effect sizes for the individual literacy, language, and math scales ranged from .12 to .46. These standard deviation units allow the proportionately greater gain for TN-VPK participants relative to nonparticipants to be represented as a percentage increase. This is shown in the graph below for the WJ Composite outcome measure.

Gain from the Beginning to End of Pre-K on the Summary Achievement Measure for Children Who Participated in TN-VPK Compared to Children Who Did Not Participate



As shown in this graph, TN-VPK resulted in a gain on the WJ Composite measure that was 38% greater than the gain made by the children who did not attend TN-VPK. The analogous improvements for the TN-VPK participants relative to the nonparticipants on the individual academic achievement measures were as follows:

Literacy: *Letter-Word Identification*, 92%; *Spelling*, 32%.
Language: *Oral Comprehension*, 31%; *Picture Vocabulary*, 74%.
Math: *Applied Problems*, 21%; *Quantitative Concepts*, 50%.

Positive effects of TN-VPK were also found on the kindergarten teachers' ratings of children's preparedness for kindergarten and, to a lesser extent, on their ratings of the children's classroom work behavior and social behavior.

We also investigated whether the TN-VPK program was differentially effective for different subgroups of children. The program's positive effects were not different for boys compared to girls, but there were larger effects on the academic skills of children who were not native English speakers than for those who were. Most of these English Language Learners were Hispanic, so no separate analysis was done for differences among ethnic groups.

The findings described in this report show that the Tennessee Voluntary Prekindergarten program produces significant improvements in the academic skills generally regarded as important for school readiness compared to the gains made by comparable children who did not participate in the program. These positive outcomes coincide with the main goal of the TN-VPK program, which is to increase the school readiness of the economically disadvantaged children it serves. The sample of children in the Intensive Substudy whose end of pre-k outcomes are examined here is being followed into subsequent grades in Tennessee public schools. Further assessments of the academic skills and classroom behavior of these children will be made each year until at least third grade to investigate the strength and longevity of these effects. The findings from those follow-ups, and from other components of the evaluation research design, will be the subject of later reports.

Evaluation of the Tennessee Voluntary Prekindergarten Program: End of Pre-K Results from the Randomized Control Design

The Peabody Research Institute at Vanderbilt University received a grant in 2009 from the U. S. Department of Education Institute of Education Sciences to conduct a study of the effectiveness of the Tennessee Voluntary Prekindergarten program (TN-VPK). This study was undertaken with the assistance of the Division of Curriculum and Instruction in the Tennessee Department of Education. The support and active participation of the state staff, and their commendable openness to a fair and probing investigation of the effectiveness of TN-VPK for improving the school readiness of the children it serves, has been critical to the successful implementation of the study. However, the design of the study, the collection and analysis of data, and the description of the results have been done entirely by the research team at the Peabody Research Institute of Vanderbilt University as an independent evaluation.

This five-year project includes two distinct studies with different research designs:

- *A randomized controlled trial (RCT)*. Schools across the state with more TN-VPK applicants than places in their classrooms were asked to admit children in order from randomized lists for school years 2009-10 and/or 2010-11. Individual assessments of literacy, language, and math skills at the beginning and end of the pre-k year were conducted on the subset of these children with parental consent (referred to as the Intensive Substudy). In addition, their kindergarten teachers rated these children early the next fall on their preparedness for kindergarten and other related characteristics. Data on the outcomes for the children enrolled in TN-VPK and those who applied for TN-VPK but were not admitted will be collected through third grade when the mandatory Tennessee state achievement tests are administered. Data collection and analysis for the outcomes at the end of the pre-k year for the consented children in the Intensive Substudy are now complete, and the results are presented in this report.
- *A regression discontinuity design (RDD)*. Two cohorts of children in a representative sample of what will eventually be about 160 schools across the state of Tennessee are participating in this phase of the study. Individual assessments on the same literacy, language, and math measures used in the RCT are being administered at the beginning of kindergarten for the cohort of children who were enrolled in TN-VPK in one of these schools during the previous school year. The scores for those children will then be compared with scores obtained from the cohort of children just entering TN-VPK in those same schools that same school year. With the difference in age accounted for statistically, this comparison will provide another set of estimates of the effects of TN-VPK. Although to date complete data have been collected from only two of four state regions, the data from all four regions will provide a representative sample of TN-VPK programs statewide. These data will provide the state the opportunity to investigate whether programmatic characteristics are linked to differential child outcomes. The results from this study will be presented later in a separate report.

The present report provides the results from the RCT Intensive Substudy only. As further data and analyses become available, later reports will present findings from the follow-up of the Intensive Substudy sample into later grades, the outcomes from the state education

database for the full randomized sample, and the results from the regression-discontinuity design.

The Tennessee Voluntary Pre-K program

Tennessee Voluntary Pre-K (TN-VPK) is a statewide program administered by the Division of Curriculum and Instruction in the Tennessee Department of Education. It began as a pilot program in 1998 and expanded rapidly after 2005 with legislation that increased its funding by providing money from the state lottery (additional state funds were added in 2006-07 and 2007-08). The program operates through competitive grants to local school systems who apply for approval and funding of one or more TN-VPK classrooms. However, those grants support only a portion of the actual cost; the balance must come from other sources. This arrangement permits and encourages collaboration between school systems and other organizations. In this "collaboration model," school districts may, at their option, operate their TN-VPK programs through collaborative agreements with local non-profit and for-profit child care providers or Head Start programs so long as those agencies have attained the highest rating from the licensing system administered by the Tennessee Department of Human Services and their programs meet the State standards for TN-VPK.

TN-VPK is a full-school day prekindergarten program for four-year-old children expected to enter kindergarten the following school year. By statute, the program gives priority to children eligible for the federal free or reduced price lunch programs and, secondarily, to students with disabilities, identified as English Language Learners (ELL), or otherwise at-risk. The program in each participating school district must meet standards set by the State Board of Education that require each classroom to have a licensed teacher with an Early Development and Learning Prekindergarten-K Endorsement or a Prekindergarten-3 or Prekindergarten-4 license, an adult-student ratio of no less than 1:10, a maximum class size of 20, and an approved age-appropriate curriculum.

Tennessee invests over $86 million a year (FY 2012-13) in the TN-VPK program. Currently, 934 state-funded TN-VPK classrooms serve more than 18,000 children from economically disadvantaged families across all 95 Tennessee counties.

Randomized Controlled Trial (RCT) Intensive Substudy: Cohorts 1 & 2 (School Years 2009-10 and 2010-11)

This report presents results from the first wave of outcome data collected from the children participating in the Intensive Substudy portion of the overall evaluation research design. It compares outcomes at the end of the pre-k year for children enrolled in TN-VPK with those for comparable children who were not enrolled. Data collection for additional waves that follow these children through successive school years is underway and the results will be reported when that data collection is complete for each follow-up wave.

The results reported here focus on two questions about the effects of the TN-VPK program:

1. Does participation in TN-VPK improve the school readiness of the economically disadvantaged children eligible for the program?
2. What are the characteristics of the children who benefit the most from TN-VPK?

Defining the Sample

There are two cohorts of children in the overall RCT study; Cohort 1 applied to TN-VPK for the 2009-10 school year, and Cohort 2 applied for the 2010-11 school year. The same randomization procedure was used for both cohorts. Prior to the pre-k school year, administrators in schools that expected more applicants to their TN-VPK program than available places were asked to participate in the RCT. When parents applied to pre-k for their children, the participating school personnel added them to an applicant list. When the schools were ready to offer admission, that list was sent to our research team where it was shuffled into a random order using a random number table. The list was then returned to the school so that parents could be contacted and offered admission for their children in the order in which those children appeared on the randomized list. School personnel were asked to make several attempts to contact parents on different days and at different times of the day to determine if admission would be accepted before moving on to the next name on the list. Once the available places were filled with children whose parents accepted admission, the remaining children became a waiting list for any seats that became available. Those remaining when no additional children on the randomized list could be offered admission provided the control group for this study. This procedure gave each child an equal chance to be ranked high enough in the list to be admitted but also, by the same equal chance, left some children too low on the list for a seat in TN-VPK to be available for them.

To be eligible for the randomization at the participating schools, children had to meet certain criteria. These required that they be:

- Assigned to either TN-VPK or the control condition on the basis of their position on the randomized applicant list (i.e., not automatically let in because of siblings in the school, etc.).
- Age-eligible (age and subsequent grade progression indicated that the child was old enough to attend kindergarten the next school year).
- Income-eligible (based on the exclusion of those children who did not meet standards for the Free or Reduced Price Lunch programs).
- Placed in a regular TN-VPK classroom; that is, not a blended or special education classroom.
- Definitively located in the state education database in at least one year following the pre-k year so that an official record was available verifying their enrollment status (enrolled vs. not enrolled) in TN-VPK.

In practice, the location of children on the randomized applicant lists was not a definitive indication of whether they were then enrolled or not enrolled in TN-VPK. Some parents declined when offered admission and others could not be contacted, with some children then offered admission who would not otherwise have been given that opportunity. We therefore determined enrollment status in TN-VPK on the basis of information obtained from the state's Education Information System (EIS) database that tracks all children enrolled in Tennessee public schools. If the State database reported that a child was enrolled in a TN-VPK classroom in the respective 2009-10 (Cohort 1) or 2010-11 (Cohort 2) school year, and was not withdrawn on the same date as initially enrolled, the child was considered to have participated in TN-VPK. Conversely, children who were not reported in

the EIS database as being enrolled in a TN-VPK classroom in the corresponding pre-k year, or had enrollment and withdrawal dates on the same day, were considered to be non-participants in TN-VPK. In addition, in order for the children on a school's randomized applicant list to be eligible for inclusion in the sample, that list had to include at least one eligible child who participated in TN-VPK (though not necessarily at the school that provided the list) and at least one eligible child who did not participate in TN-VPK according to the information reported in the State EIS database.

Across the two cohorts, 80 schools in 29 Tennessee school districts submitted applicant lists for randomization that were eligible for inclusion in the RCT sample under these criteria. These 80 schools were spread across the state with the largest number in the Central West region (Nashville and surrounding counties). Overall, their distribution was:

- 15 schools (8 districts) in the West region (including Shelby County and Memphis);
- 33 schools (10 districts) in the Central West region (including Davidson County and Nashville);
- 14 schools (3 districts) in the Central East region (including Hamilton County and Chattanooga);
- 17 schools (8 districts) in the East region (including Knox County and Knoxville).

These schools included 20 in urban areas (12 in large cities and 8 in mid-size cities), 29 in suburban areas, and 31 in rural areas.

The sample of over 3000 children that resulted from this randomization procedure will continue to be tracked through the State Education Information System database until their third grade year when the mandated state achievement tests (TCAP) are given to all children in Tennessee public schools. Outcome data from that source are not yet available. Results from a more intensive substudy of a subset of the children in this overall sample are available, however, and are reported here.

The Intensive Substudy Sample

The families of the children on the eligible TN-VPK applicant lists were contacted and asked to provide consent for their children to be individually assessed for the Intensive Substudy (ISS). In order for a child to be included in the ISS, he/she had to meet the first five eligibility criteria listed above for the full sample *and* be consented and assessed by PRI staff at least once during the pre-k or kindergarten year. In addition, the consented, eligible children on a school's applicant list were only included if there was at least one consented, assessed child who participated in TN-VPK and at least one consented child who did not. If this latter condition was not met, none of the children on that applicant list were included in the ISS sample.

Rather than rely only on information in the State EIS database to determine whether a child was a participant or non-participant in TN-VPK, as was done in the full randomized sample, the determination for the ISS was made on the basis of all the available information. For the consented children in the ISS sample, we had information from our contact with parents, teachers, and other school staff in addition to the information available from the State EIS database. Further, for the ISS sample, we considered a child to be a non-participant in TN-VPK if he/she attended TN-VPK for fewer than 20 days. Descriptive information and a flow chart are provided in Appendix A3 that identify the number of

eligible and ineligible children across study components, including the Intensive Substudy sample and the analysis sample for this report.

Parental Consent for the Intensive Substudy

In Cohort 1, confidentiality concerns by personnel at the Tennessee Department of Education required us to use an indirect process for obtaining parental consent for children to participate in the Intensive Substudy. In this procedure, we provided consent forms to a staff member at the Department of Education who then mailed them to the parents with an envelope in which they could return them to the research office. The lack of personal contact with parents, the time lag after notification about admission to TN-VPK, obsolete mailing addresses, and the expectable inattention to mailed material resulted in a low consent rate that was quite variable across sites.

After further review of the confidentiality issue, a better procedure was allowed for Cohort 2. Rather than going through the State Department of Education, PRI research staff were able to be directly involved in the application procedure at each school. Staff members attended TN-VPK enrollment sessions at the participating schools, added a consent form to the application package that was given to parents, and were available to answer questions about the study. This improved consent procedure resulted in a much higher consent rate than with Cohort 1. Detailed consent rate summaries for the participating schools and overall are provided in Appendices A1 and A2

The procedure used for Cohort 1 produced an overall consent rate for the ISS of 42%, with a 46% consent rate for TN-VPK participants and 32% for non-participants. The procedure used for Cohort 2 resulted in an overall consent rate of 71%, with 74% of the participants and 68% of the non-participants consenting. The poor to modest consent rates overall, and the differential consent rates for participants and non-participants in each cohort created the potential for the TN-VPK participant and nonparticipant groups to differ on important initial characteristics despite the randomization applied to the applicant lists from which they were drawn. Without comparison groups formed solely and entirely by random assignment (not including consent), it is especially important to assess their similarity on initial baseline characteristics and take any initial differences between the groups into account during the analysis. The baseline comparisons and associated analysis approaches are described in later sections of this report.

Selection of the Intensive Substudy Analysis Sample

The analysis sample for this report began with the 3171 children whose names appeared on the 113 randomized applicant lists that were provided by the 80 schools in 29 districts that participated in the RCT for Cohort 1 and/or Cohort 2. Of those, 76 applicant lists from 58 unique schools in 21 districts across the state included at least one TN-VPK participant and one non-participant whose parents consented to their inclusion in the ISS and who were posttested at the end of the pre-k year. Those applicant lists identified 1079 children, 308 from Cohort 1 and 771 from Cohort 2 (774 TN-VPK participants and 305 non-participants). However, two children were subsequently trimmed from the sample because of their outlier propensity scores (a full description of this procedure is presented later in this report), which left 1077 children in the final analysis sample, 773 TN-VPK

participants and 304 nonparticipants. Appendix A3 provides flow charts that show the derivation of the ISS analysis sample from the initial set of consented Cohort 1 and 2 children. The randomized lists that generated this sample of children were contributed by 10 schools in the West region of the state, 24 schools in the Central West, 12 schools in the Central East, and 12 schools in the East. Nineteen of the schools were near cities (10 large, 7 mid-size, and 2 small), 11 were in the suburbs, 12 were in towns, and 16 were considered rural. Three of them were public pre-k centers affiliated with public schools. None of the 58 schools was a partnership site (i.e. received funding from multiple sources).

Student Measures

Direct Child Assessments. Achievement on emergent literacy, language, and math school readiness measures was assessed for the children in the Intensive Substudy at the beginning and end of the pre-k school year. A timeline of these study events can be found in Appendix A4. The children enrolled in TN-VPK were individually assessed by trained members of the research team on site at the school where they attended. The children who were not enrolled in TN-VPK were assessed in places arranged with the parents/guardians that were convenient for them, e.g., libraries, parks, homes, schools, childcare sites, etc.

The standardized assessments administered to each child included the scales listed below from the Woodcock Johnson III Achievement Battery (*WJ;* Woodcock, McGrew, and Mather, 2001). These instruments are widely used in prekindergarten research to assess emergent literacy, language, and math skills related to young children's readiness to begin kindergarten and have the advantage of being applicable longitudinally through at least the elementary grades.

Literacy

- *Letter-Word Identification*: Assesses children's ability to identify and pronounce alphabet letters and read words.
- *Spelling*: Assesses children's prewriting skills, such as drawing lines and tracing, writing letters, and spelling orally presented words.

Language

- *Oral Comprehension*: Assesses children's ability to complete analogies and provide words with similar or different meanings from key words; it measures their listening ability and understanding.
- *Picture Vocabulary*: Assesses children's ability to name the objects presented in pictures and point to the picture that goes with a word; it measures early language development and lexical knowledge.

Math

- *Applied Problems*: Assesses children's ability to solve small numerical and spatial problems presented verbally with accompanying pictures of objects.
- *Quantitative Concepts*: Assesses children's ability to point to or state answers to questions on number identification, sequencing, shapes, symbols, and the like; it measures aspects of quantitative reasoning and math knowledge.

Composite Measure

- *WJ Composite Score:* A principal components factor analysis revealed that all the scales above were rather intercorrelated with high loadings on a single factor. The W-scores

on those scales were therefore combined to create a composite measure representing children's overall achievement in literacy, language, and math. Because these scales had somewhat different distributions, this composite was created by standardizing and then rescaling each scale to have a common standard deviation of 20, adding back the grand mean, and then averaging the transformed scale scores together.

Examples of the items on each of these scales are provided in Appendix B1. For the Woodcock Johnson tests, the IRT scaled W-scores were used in the analyses unless otherwise indicated. The W-scores are the ones suitable for longitudinal comparisons using any of these scales.

Teacher Reports. In addition to direct assessments of children's achievement, reports of their academic skills and classroom behavior were collected from their kindergarten teachers early in the fall of the school year after the pre-k year. Of the 1077 children in the analysis sample, we were able to obtain teacher reports for 965 (90%). With these reports we can compare the kindergarten teachers' perceptions of the children who participated in TN-VPK the year before and those who did not participate. Among the original schools contributing children to the TN-VPK participant and nonparticipant groups from their randomized applicant lists, however, there were four school lists for which no kindergarten teacher reports were obtained for at least 1 child who had attended TN-VPK and 1 child who had not, and those lists were excluded from the analysis of effects on teacher reports. This left 914 children with kindergarten teacher reports. The two teacher rating instruments used for this purpose were the following:

- *Cooper-Farran Behavioral Rating Scales:* Teacher ratings for each child on items about the child's actions and interactions within the classroom context. Items include questions about a child's ability to pay attention and engage in classroom activities, compliance with instructions, work and play with peers, ability to act independently, and similar topics (Cooper and Farran, 1991). It is comprised of two parts:
  - *Work-Related Skills:* A factor score derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about the child's ability to work independently, listen to the teacher, remember and comply with instructions, complete games and activities, function within designated time periods, and otherwise behave appropriately in relation to classroom work and prescribed activities.
  - *Social Behavior:* A factor score also derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about the child's social interactions with peers including appropriate behavior while participating in group activities, play, and outdoor games, expression of feelings and ideas during group discussions; and response to others' mistakes or misfortunes.
- *Academic Classroom and Behavior Record*: Teacher ratings for each child on items about the child's preparedness for kindergarten work, liking school, behavior problems in the classroom, and relations with peers (Farran, Bilbrey, and Lipsey, 2003). It is comprised of three parts:
  - *Readiness for Kindergarten:* A factor score derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated

items about how well prepared the child is for kindergarten in literacy and language skills, math skills, and social behavior.

- *Liking for School:* A factor score derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about the child liking or disliking school, having fun at school, enjoying and engaging in classroom activities, and seeming happy at school.
- *Behavior Problems:* Whether the child has shown any behavior problems from a list that includes explosive and overactive behaviors, attention problems, physical or relational aggression, social withdrawal or anxiety, motor difficulties, and the like. Teachers' responses were coded into two outcome variables: (a) a dichotomous variable indicating whether any of the behavior problems on the list had been observed (0= no, 1=yes), and (b) the number of behavior problems on the list that were observed.

Examples of the items used in these scales are provided in Appendix B2.

Parent Questionnaires. Questionnaires were administered via telephone to the parents of the consented children. These were conducted during the pre-k year to provide information about child and family characteristics that might impact achievement such as demographic information and activities at home with the child. The parents whose children were not admitted to TN-VPK were also asked about any alternative preschool or childcare arrangements they had made. These were completed with the parent or guardian of 1034 of the 1077 children in the analysis sample (96%). Among the items on that questionnaire were questions about the parents' education, the number of working parents, the home language environment, and various literacy activities of the child or the child and parent at home including library card use and newspaper and magazine subscriptions. A copy of the parent questionnaire is provided in Appendix B3.

Description of the ISS Analysis Sample

As described earlier, the ISS analysis sample consisted of 1077 children with parental consent and end of pre-k posttest assessments who were drawn from randomized applicant lists that included at least one eligible child who participated in TN-VPK and one who did not. The demographic characteristics for this sample are given in Table 1 below, and a breakdown of these numbers by Cohort can be found in Appendix C1.

Table 1. Demographic Characteristics of the Children in the ISS Analysis Sample

Characteristic	Mean
Age in months at start of pre-k year	51.8
White	.57
Black	.23
Hispanic	.19
Asian & others	.01
Male	.48
Native English Speaker	.80

Table 2 below reports descriptive results from a selection of the items about family context that appeared on the parent questionnaire, and a Cohort breakdown of these numbers can be found in Appendix C2.

Table 2. Family Context Information for the ISS Analysis Sample from the Parent Questionnaire

Questionnaire Item	N	Min	Max	Mean	SD
Mother's Education (1-4)	1031	1	4	2.09	0.73
Library Card Use (0-2)	1018	0	2	0.91	0.84
Newspaper Subscriptions (0-3)	1027	0	3	0.36	0.77
Magazine Subscriptions (0-2)	1029	0	2	0.28	0.51
Number of Working Parents	1034	0	2	1.25	0.63
Weekday TV hrs watched (1-4)	1031	1	4	2.01	0.76
Saturday TV hrs watched (1-4)	1030	1	4	2.16	0.93
Sunday TV hrs watched (1-4)	1022	1	4	1.99	0.89
Number children's books (0-5)	1031	0	5	4.07	1.24
How often parents read to child (1-5)	1031	1	5	4.28	1.00
How often parents count (1-5)	1031	1	5	4.17	1.12

Mother's Education (1=less than high school, 2=high school diploma/GED, 3=associate's degree, 4=more than associate's degree); Library Card Use (0=no card/used almost never, 1=used once or twice a year or every few months, 2=used more than once a year or at least weekly); Newspaper Subscriptions (0=0, 1=1, 2=2-3, 3=>3); Magazine Subscriptions (0=0, 1=1-3, 2=>3); TV hours watched (1=< 1, 2=1-3, 3=3-5, 4=>4); Number children's books (0=0, 1=1-5, 2=6-10, 3=11-19, 4=20-50, 5=>50); How often read/count with child (1=Almost never, 2=Not weekly but sometimes, 3=1-3 times/week, 4=4-6 times/week, 5=Daily).

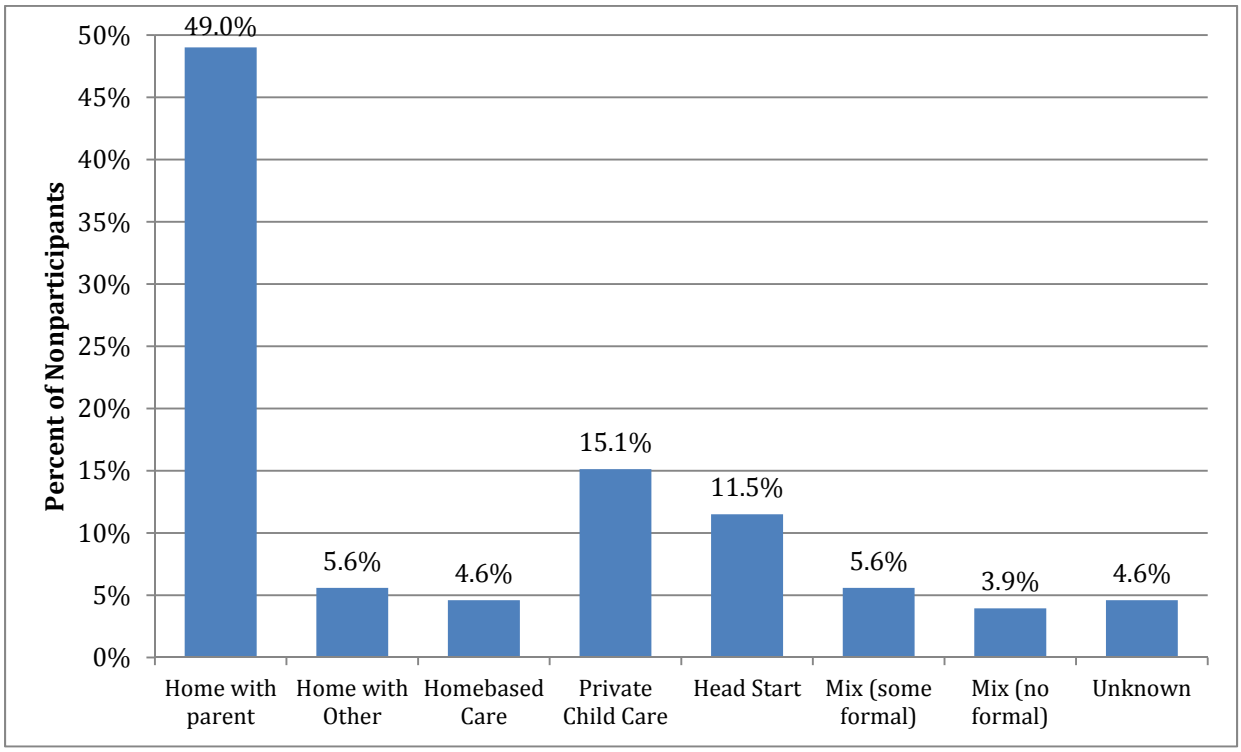
Table 3 shows standard scores on the Woodcock Johnson subscales at pretest for the entire analysis sample after missing pretest scores were imputed (a procedure described in the data analysis section below). The children in the ISS analysis sample are below the national norm of 100 on each subtest, and almost a full standard deviation below the norm on Spelling and Quantitative Concepts. These scores are broken out by Cohort in Appendix C3. Standard scores are presented for descriptive purposes only; as noted, analyses were done with the W scores, a more sensitive measure of change.

From the Parent Questionnaire we were also able to obtain information about the childcare arrangements for children who were not admitted to TN-VPK. Responses are displayed in Figure 1 below, and show that the majority of children who applied for, but did not attend TN-VPK stayed home with a parent or some other relative or caretaker (more than half of the nonparticipant group). Only 11.5% were able to attend a Head Start program as an alternative to the TN-VPK program to which they had applied while 15.1% were enrolled in private childcare.

Table 3. WJ Standard Score Pretest Means for the ISS Analysis Sample

Pretest Subscale	N	Min	Max	Mean	SD
WJ Letter-Word Identification	1077	60	164	93.0	13.5
WJ Spelling	1077	39	139	87.0	14.8
WJ Oral Comprehension	1077	46	130	92.9	13.4
WJ Picture Vocabulary	1077	26	130	94.4	18.8
WJ Applied Problems	1077	56	134	96.8	13.6
WJ Quantitative Concepts	1077	56	133	88.8	12.6
WJ Composite Score	1077	59.7	129.3	92.2	11.3

Figure 1. Childcare Arrangements for Children Who Did Not Participate in TN-VPK



Data Analysis

Identification of TN-VPK Participants and Nonparticipants for Purposes of Comparison. Determining whether a child participated in TN-VPK was not as straightforward as examining his/her place on the randomized applicant list used by the school to sequence offers of admission. Some children who were not admitted from that list ended up enrolling in a TN-VPK classroom in a different school. The parents of other children who were offered admission from the randomized list declined for various reasons and, in some cases, their children did not then enroll in TN-VPK. For other children designated for offers

of admission, the parents could not be reached and were skipped over as the school personnel worked down the randomized applicant list. In addition, some children who enrolled in a TN-VPK program were withdrawn by their parents after such a short time that they had little actual exposure to TN-VPK. These circumstances produced a number of crossovers of children designated for TN-VPK on the randomized applicant list who did not then participate in TN-VPK and children who were not designated for participation who did end up participating.

For the Intensive Substudy that is the focus of this report, we also had to take the consent rates into account (42% overall for Cohort 1, 71% for Cohort 2). With far fewer than 100% consenting and differential consent rates for children admitted vs. not admitted to TN-VPK, much of the intended randomization was undermined by the attrition of non-consented children from the randomized conditions (see Appendices A1 and A2 for details about the consent rates).

For analysis of the ISS sample we have elected to compare outcomes for the children who participated in TN-VPK with those who did not participate irrespective of the conditions designated for them on the original randomized applicant lists. Though the influence of the randomized applicant lists still provides a substantial chance component to the determination of which children participated, the crossover and non-consent rates allow other factors besides chance to also influence the initial similarity of participants and nonparticipants. We have, therefore, undertaken the analysis on the assumption that the TN-VPK participants and nonparticipants whose outcomes we are comparing may be different at baseline in ways that could bias the estimates of the effects of TN-VPK. To account for any such differences as much as possible, we have used the available baseline data to examine initial differences and to generate variables for use as statistical controls in the analysis. The details of these procedures are described later in this report.

To identify the children in the ISS analysis sample who participated in TN-VPK and those who did not participate for purposes of comparison, we made use of all the information available to us about each child's pre-k experience. This included the records in the State Education Information System showing enrollment status and information gathered during our data collection provided by parents, teachers, and school personnel. We then defined TN-VPK participants as children for whom the available information indicated that they attended any TN-VPK program for at least 20 days during the school year, whether at the school that included them on a randomized applicant list or not (mean attendance for this group was 149 days). This 20-day attendance requirement was identified by the administrator of TN-VPK at the State Department of Education as the minimum number of days they required to consider a child as having participated in TN-VPK; it constitutes one attendance reporting period. TN-VPK nonparticipants, conversely, were defined as children for whom the available information indicated that they did not attend any TN-VPK program or, if they attended, it was for fewer than 20 days.

Missing Data for the Achievement Measures. The achievement outcomes on the six Woodcock-Johnson measures (Letter-Word ID, Spelling, Oral Comprehension, Picture Vocabulary, Applied Problems, and Quantitative Concepts) and the key independent variable (TN-VPK participation) did not have any missing data in the ISS analysis sample.

There were missing values on some of the other variables used in the analyses and these were imputed so that all the cases in the ISS sample could be used. To preserve consistency between the imputation models and the analysis models used to estimate the effects of TN-VPK, the variables in the imputation function included all those that were also to be used in the analysis models (Allison, 2002). In particular, these included the pretests and posttests for the six Woodcock Johnson achievement measures, child's age at pretest, test lag from school start date, test interval between pretest and posttest, ethnicity, gender, speaking English at home, mother's education, library card use, newspaper subscriptions, magazine subscriptions, number of working parents, and cohort. The average missing value rate across all the variables in the imputation was 2.5% with a range of 0 to 5.5%. Given so few missing values, we chose to use a single imputation with SAS PROC MI and carry the results forward to all the analyses of TN-VPK effects on the outcome variables.

Missing Data for Teacher-Reported Measures. After removing children from the analysis sample of 1077 who did not have teacher reports and/or were on randomized lists that did not include at least one TN-VPK participant and one Nonparticipant, 914 eligible children had complete teacher rating outcomes on three measures (Cooper-Farran Social Behavior, Cooper-Farran Work-Related Skills, and ACBR Behavior Problems), and 913 children had complete teacher rating outcomes on another three measures (ACBR Preparedness for Kindergarten, ACBR Peer Relations, and ACBR Feelings About School). The missing data on these teacher rating measures were not imputed but missing values on the applicable covariates identified above for the Woodcock Johnson achievement outcome measures that were imputed in that procedure were also used in the analysis of the teacher rating outcomes. In addition, there were two covariates that were applicable only to the analysis of the teacher rating outcomes—the interval between the school start date and the time the rating was made and the age of each child at the time of the rating. Only 3.6% of the values on these variables were missing and here also we used a single imputation to estimate those values with the teacher rating outcome measures included in the imputation model.[1]

Baseline Equivalence for the TN-VPK Participants and Nonparticipants. As described earlier, our analysis of TN-VPK effects is based on a comparison of outcomes for children participating in TN-VPK with those for children who did not participate. Because assignment to those conditions was not determined entirely by the randomization procedure, we cannot assume that the children who participated in TN-VPK were equivalent at baseline on all relevant characteristics to the children who did not participate. The analysis of TN-VPK effects, therefore, was conducted using propensity scores to adjust for baseline differences between these groups. The initial step in constructing those propensity scores was to compare the TN-VPK participant and nonparticipant groups of children on the available baseline variables to assess their similarity and to identify any variables on which they differed.

Baseline differences were therefore examined on the children's demographic variables, pretests of the Woodcock Johnson measures, and selected family background items from the parent questionnaire. Omitted from this selection were items for which the parents'

[1] A later analysis used multiple imputation both to handle missing values and as a basis for the propensity scores. The results were similar in all relevant aspects to those reported here.

responses might be affected by the participation of their children in TN-VPK. Because children in TN-VPK had already begun attending, some parent behaviors typically encouraged by pre-k teachers, such as reading to their child, might not reflect true baseline differences. In addition, because of the difficulty locating and scheduling assessments for some children, and variation across schools in the dates for the beginning and end of the school year, some children had later pretests than others and some had more time between pretest and posttest than others. To examine the influence of those differences, the lag in days between the average start date for TN-VPK and the date of the pretest was computed. Similarly the interval in days between the pretest and posttest was recorded and examined. The baseline differences on all these variables were tested in multilevel statistical models

Table 4. Comparison of the TN-VPK Participants and Nonparticipants on Baseline Variables and Correlations of Those Variables with the WJ Composite Posttest Scale

	TN-VPK Participants			Nonparticipants				
Baseline Variable	Mean[f]	SD	N	Mean[f]	SD	N	***p*-value for Difference**	**Correlation with WJ Composite Posttest**
Age at Pretest (months)	54.0	3.51	759	54.6	3.72	285	.018**	0.20
Gender (0=F, 1=M)	0.5	0.50	773	0.5	0.50	304	.988	-0.08
Black	0.2	0.42	773	0.2	0.43	303	.516	0.12
Hispanic	0.2	0.37	773	0.2	0.44	303	.790	-0.35
WJ Letter-Word ID	319.7	27.3	759	315.6	27.5	285	.041**	0.69
WJ Spelling	350.9	28.6	759	349.7	28.8	285	.579	0.58
WJ Oral Comprehension	444.3	15.6	759	442.8	17.7	285	.191	0.65
WJ Picture Vocabulary	456.6	20.9	759	454.0	27.9	285	.109	0.68
WJ Applied Problems	391.7	26.9	759	391.1	30.2	285	.734	0.72
WJ Quantitative Concepts	407.6	14.0	759	407.6	14.2	285	.997	0.72
WJ Composite Score	395.3	17.6	759	393.7	19.3	285	.221	0.84
Pretest Lag (days)[a]	72.1	23.0	759	88.0	31.2	285	.000**	-0.07
Pre-post Interval[b]	193.2	24.6	759	189.1	38.8	285	.019**	0.03
Mother's Education[c]	2.2	0.7	736	2.0	0.7	295	.018**	0.25
Native English Speaker (1=Yes)	0.8	0.4	773	0.8	0.5	304	.568	0.39
Library Card Use [d]	1.0	0.8	728	0.9	0.9	290	.093*	0.20
Newspaper Subscriptions[e]	0.4	0.8	734	0.3	0.8	293	.489	0.08
Magazine Subscriptions[e]	0.3	0.5	736	0.3	0.5	293	.526	0.14
Number of Working Parents	1.3	0.6	739	1.2	0.6	295	.668	0.07

*[a] Days to pretest from average schools start date [b] Days between the pretest and posttest. [c] 4-point scale from less than high school to more than associate's degree. [d] 3-point scale from never/almost never used to used more than once/month. [e] 3-point scale from 0 to 4 or more. [f] Marginal mean from random effects multilevel analysis with condition blocked within schools, schools nested within districts, and no covariates. Condition was treated as a fixed effect. * p<.10 ** p<.05*

with children nested within the schools that provided the randomized applicant lists to the design, and with those schools nested within districts. Full descriptive information on pre- and posttest scores for direct assessments and teacher ratings for the total ISS analysis sample and broken down by cohort can be found in Appendices C4 and C5.

Table 4 shows the comparisons between the TN-VPK participant and nonparticipant groups for these different variables using the imputed dataset described above. This table also shows the correlation of each of these variables with the WJ Composite Scale measured at posttest. The variables that differ at baseline between the conditions and are also correlated with the outcome measures are the ones most important to control in the analysis. As Table 4 shows, the children in the TN-VPK participant and nonparticipant groups were similar on many of these variables, though there were some statistically significant differences. The direction of those significant differences overall generally favored the TN-VPK participant group.

Though standard scores on the Woodcock Johnson subscales were not used in our main effects analysis, the baseline differences are presented in Table 5 below for descriptive purposes.

Table 5. Comparison of the TN-VPK Participants and Nonparticipants on Baseline WJ Standard Scores

	TN-VPK Participants			Nonparticipants		
Baseline Variable	Mean	SD	N	Mean	SD	N
WJ Letter-Word Identification	93.8	13.5	773	91.0	13.5	304
WJ Spelling	87.1	14.7	773	86.6	15.0	304
WJ Oral Comprehension	93.8	12.7	773	90.4	14.6	304
WJ Picture Vocabulary	96.3	17.0	773	89.7	22.0	304
WJ Applied Problems	97.6	13.3	773	94.7	14.2	304
WJ Quantitative Concepts	89.3	12.6	773	87.6	12.6	304
WJ Composite Score	93.0	10.9	773	90.0	11.9	304

Statistical Analysis Approach. The effects of the TN-VPK program on the outcome variables were estimated using multilevel regression models with children *(i)* nested within the schools to which their parents initially applied for the TN-VPK program *(j)*, nested within school district *(k)*.[2] The treatment effects were estimated as constants (fixed effects) across districts and across schools. The slopes for all other variables were fixed and a random intercept at each level was included.

[2] If a school or district was represented in both cohorts, that school or district was treated as two unique entities in the analysis, which resulted in a multilevel model with 1077 children in 76 separately specified schools and 34 separately specified districts when the actual number of schools and districts was 58 and 21 respectively. Alternative nesting structures were also examined, but produced virtually identical results.

To adjust for the baseline differences summarized in Table 4 above, propensity score techniques were used. To ensure that no baseline variable that might contribute to controlling for such differences was omitted, each of the variables in Table 4 was included in the propensity scores. Specifically, the variables that went into the propensity scores were the following:

- Child age at pretest, gender (dummy coded for male yes/no), ethnicity (dummy coded for Black yes/no and Hispanic yes/no), and native language (dummy coded for English yes/no);
- Parent reports of their home library card use, newspaper subscriptions, magazine subscriptions, maternal education, and number of working parents;
- The pretest W-scores for all six Woodcock Johnson achievement scales;
- The time lag from the beginning of the school year to pretest and the interval between pretest and posttest.

To account for any differences between cohorts in the variables that differentiated the TN-VPK participant and nonparticipant groups at baseline, separate propensity scores were created for each cohort. To create the propensity scores, the variables listed above were used in three-level logistic regressions (students nested in schools, and schools nested in districts) to predict TN-VPK vs. no TN-VPK group membership, and the predicted values for the probability[3] of being in the TN-VPK group were saved as the initial propensity scores. For consistency with the final analysis models, sample weights (described later) were also used in this logistic regression. The distributions of these propensity scores for the TN-VPK participant and nonparticipant groups were then examined for overlap. Cases in either distribution with a propensity score more than .25 standard deviation below the minimum or above the maximum of the other distribution were dropped from the analysis. This reduced the sample size by only two children in the nonparticipant group in Cohort 1, but helped improve the match on the variables in the propensity score. As noted earlier, with these dropped cases, the total number of children in the analysis sample was 1077.

After computing the propensity scores, the baseline differences on the variables used to create the propensity scores were examined with the propensity scores incorporated in the analytic model to assess how well they adjusted for those baseline differences. There are various ways that propensity scores can be used in an analysis of intervention effects: as a covariate, as the basis for a weighting function, or for matching with various caliper widths for what counts as a match (Guo & Fraser, 2010). A selection of these approaches was tried in analyses with the WJ Composite outcome variable and each gave somewhat different, though substantially similar effect estimates. The approach that best balanced the baseline differences was use of the propensity score simply as a covariate in multilevel regression analysis, so that approach was used in all the analyses reported below. (Appendix C6 provides details about the performance of the different propensity score variants for balancing baseline differences.)

The proportion of children participating in TN-VPK relative to the proportion of nonparticipants varied considerably across the randomized applicant lists at the

[3] The propensity scores in logit form were also examined for their ability to balance the covariates, but were found to be less effective for that purpose than the probability estimates. The variants examined are described more fully in Appendix C6.

contributing schools. The uneven contributions of the associated schools to each of the comparison conditions was itself a possible source of differences between those conditions, including possible differences on variables that had not been measured at baseline. We therefore created sample weights to balance the representation of the respective schools in the two comparison groups. This was done by weighting the data from the TN-VPK participants and nonparticipants within each applicant list so that their proportionate contributions matched the proportions in the total sample. The proportion of children participating in TN-VPK in the full analysis sample was 69.3% with 30.7% thus in the nonparticipant group. We therefore weighted the cases from each condition contributed by each applicant list upward or downward as needed to match those proportions. That is, if there were fewer than 69.3% of the cases from an applicant list in the TN-VPK condition, and thus more than 30.7% in the no TN-VPK condition, each TN-VPK case was weighted upward and each nonparticipant case was weighted downward in a way that made their contributions to the total sample match the target proportions, and vice versa if the discrepancies went the other way.

The sample weights (W_{jt}) for the children on school applicant list *j*, and in condition *t*, used to accomplish this were created by the formula below:

$W_{jt} = \frac{n_j P_t}{n_{jt}}$, where *j* indicates the applicant list, *t* represent the condition (TN-VPK or no TN-VPK), n_j is the total number of children in the analysis sample from that applicant list, n_{jt} is the total number of children from that applicant list in condition t, and P_t is the proportion of the total analysis sample in condition *t* (.693 for TN-VPK participants, .307 for nonparticipants). These sampling weights were then used in all the multilevel analyses reported below unless otherwise indicated.

The sample weights adjusted for the varyingly different proportions of TN-VPK participants and nonparticipants that each randomized list contributed to the analysis sample. A further source of variation across the applicant lists and the TN-VPK participants and nonparticipants within each list was the participation rate of each group in the analysis sample. Mainly as a result of the different parental consent rates discussed earlier (detailed information about consent rates can be found in Appendices A1 and A2), different proportions of the total number of TN-VPK participants and nonparticipants on the randomized applicant lists were represented in the overall ISS analysis sample. If the children of consenting and nonconsenting parents differ in ways related to the outcomes, as is likely, the varying rates of inclusion in the analysis sample for TN-VPK participants and nonparticipants across the applicant lists might inappropriately influence the estimates of the effects of TN-VPK on those outcomes. We therefore created two *inclusion rate* variables for each applicant list that were used as Level 2 covariates in the analysis. These were defined as (a) the number of TN-VPK participants included in the analysis sample from a given applicant list divided by the total number of TN-VPK participants on that list, and (b) the number of TN-VPK nonparticipants included in the analysis sample from that list divided by the total number of nonparticipants on the list. Because the TN-VPK participant and nonparticipant inclusion rates typically differed for each list, we also included the interaction of these two inclusion rates at Level 2 in the analysis. In addition,

because the overall inclusion rates differed between the cohorts, we included a dummy code for cohort as a covariate.

For the analysis sample, the inclusion rate averaged 43% for the nonparticipants and 63% for the TN-VPK participants. Cohort 1 inclusion rates were lower than those for Cohort 2. The average participant and nonparticipant inclusion rates across randomized applicant lists in Cohort 1 were 44% and 27%, respectively, and were 72% and 49% for Cohort 2. Additional details for these inclusion rates can be found in Appendix C7.

The individual variables used to create the propensity scores were also included in each analysis as covariates along with the propensity score. Because of the correlations of some of these variables with the outcome variables (especially the pretest), these covariates help improve statistical power. Used as individual covariates, they are also able to adjust for any remaining baseline imbalance between the TN-VPK participant and nonparticipant groups on any of these variables that was not fully adjusted by the propensity score. Our goal is to be as conservative as possible to make sure that the group of TN-VPK Participants and Nonparticipants are equivalent at the outset prior to any analysis of outcomes.

Despite the seemingly large sample size (1077) available for analysis, statistical power was a concern. The unconditional intraclass correlation coefficients (ICCs) that index the proportion of variance within each nesting level relative to the total variance are reported in Appendix C8. For some of the WJ achievement measures, these ICCs are large enough to produce a reduced effective sample size in multilevel analysis with the associated reduction in statistical power. With this in mind, we have reported statistical significance in the analyses that follow at the alpha=.10 level as well as at the conventional .05 level.

To summarize, the analysis models that produced the results reported below were multilevel regressions with children nested within schools and schools nested within districts. The school level was indexed so that it corresponded to applicant lists and constituted a blocking factor with TN-VPK participants and nonparticipants appearing in each school block (as well as within each district within which the school blocks were nested). Treatment effects were not allowed to vary randomly across school blocks or district blocks. A propensity score created separately for each cohort from baseline variables in the form of a predicted probability of each child being a TN-VPK participant was included as a covariate in the analyses. The baseline variables used to create the propensity scores were also included individually as covariates as were the school level TN-VPK participant and nonparticipant inclusion rates and their interaction. Finally, a weighting function was included in all analyses that weighted the cases within each school/applicant list so that the contributions of the TN-VPK participants and nonparticipants were proportionate to their respective percentages in the total analysis sample. More specifically, the multilevel regression models predicting posttests on the outcome variables took the following form:

Child Level:

$$Posttest_{ijk} = \pi_{0jk} + \pi_{1jk}Black_{ijk} + \pi_{2jk}Hispanic_{ijk} + \pi_{3jk}Male_{ijk} + \pi_{4jk}Native\ English\ Speaker_{ijk} + \pi_{5jk}Library\ Card\ Use_{ijk} + \pi_{6jk}Newspaper\ Subscriptions_{ijk} + \pi_{7jk}Magazine\ Subscriptions_{ijk} + \pi_{8jk}Mother's\ Education_{ijk} + \pi_{9jk}Number\ of\ Working\ Parents_{ijk} + \pi_{10jk}Age\ at\ Pretest_{ijk} + \pi_{11jk}Test\ Lag_{ijk} + \pi_{12jk}Test\ Interval_{ijk} + \pi_{13jk}Pretest_{ijk} + \pi_{14jk}Propensity\ Score_{ijk} + \pi_{15jk}\boldsymbol{TN\text{-}VPK\ Participation}_{ijk} + e_{ijk}$$

Randomized List/School Level:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}TN\text{-}VPK\ Participant\ Inclusion\ Rate_{jk} + \beta_{02k}TN\text{-}VPK\ Nonparticipant\ Inclusion\ Rate_{jk} + \beta_{03k}Inclusion\ Rate\ Interaction_{jk} + \beta_{04k}Cohort_{jk} + r_{0jk}$$

District Level:

$$\beta_{00k} = \gamma_{000} + \mu_{00k}$$

Reasonable alternative decisions could have been made about many of the specific features of the statistical analysis model we chose for the main analysis of TN-VPK effects. We have conducted extensive sensitivity analysis to explore the influence of different choices for the key features of these models on the estimates of those effects. The results showed substantial similarity across these variants with the analysis approach we adopted being among the more conservative. Summaries of these sensitivity analyses can be found in Appendix D.

We should note also that the analysis models we adopted for the results reported below are somewhat different from those used in the reports released earlier summarizing the preliminary findings of this study. These revised results, however, are substantially similar to the earlier ones and do not change any of the substantive conclusions about the nature and general magnitude of the TN-VPK effects found.

Results of the Analyses of TN-VPK Effects

TN-VPK Effects on the Achievement Measures. To determine the impact of the TN-VPK program on the end of pre-k Woodcock Johnson achievement measures, multilevel regressions as described above were conducted for the composite variable and each of the individual scales. Table 6 shows the full model results for the WJ Composite measure and Table 7 summarizes the results and reports the effect sizes for the six individual scales as well as the WJ Composite. The full model results for the analysis of each of the individual scales can be found in Appendix D1.

As can be seen in these tables, the children who participated in TN-VPK gained much more during the pre-k year on all the Woodcock Johnson measures examined than the children who did not participate. These effects were statistically significant on the composite measure as well as on each of the WJ literacy, language, and math measures with standardized mean difference effect sizes ranging from .12 to .46. The largest effects were

on Letter-Word Identification and Quantitative Concepts while the smallest were on Applied Problems, Picture Vocabulary, and Oral Comprehension. (Detailed results for each of the analyses summarized in Table 7 are presented in Appendix D1).

In further analysis, we investigated whether the TN-VPK effects were different for the two cohorts for which data were combined in the ISS analysis sample. The cohort by TN-VPK Participation treatment variable interaction was added to the model analyzing effects on the WJ Composite (Table 6 above). The *p*-value for the statistical significance of this interaction term (p=.100) was very close to falling under the alpha=.10 standard we set for reporting effects. Estimated separately, the effect size on the WJ Composite for TN-VPK participation was .41 for Cohort 1 and .26 for Cohort 2.

Table 6. Full Model Results for the Analysis of the Effect of TN-VPK on the End of Pre-K WJ Composite Measure W-Scores

Variable	b	SE	*p*-value
Intercept	74.82**	9.87	0.000
Inclusion Rate: Nonparticipants	-0.80	1.67	0.631
Inclusion Rate: TN-VPK Participants	-0.58	1.73	0.739
Inclusion Rate Interaction	-8.77	7.35	0.237
Cohort (Cohort 2 = reference)	0.70	0.92	0.445
Black	1.40*	0.69	0.044
Hispanic	1.46†	0.77	0.059
Male	0.47	0.57	0.404
Native English Speaker	-1.09	1.33	0.412
Library Card Use	-0.22	0.28	0.424
Newspaper Subscriptions	-0.05	0.25	0.829
Magazine Subscriptions	0.43	0.49	0.374
Mother's Education	-0.17	0.35	0.627
Number of Working Parents	-0.02	0.46	0.969
Age at Pretest	-0.09	0.10	0.405
Test Lag	-0.09**	0.03	0.009
Test Interval	-0.01	0.03	0.815
Pretest	0.88**	0.02	0.000
Propensity Score	-0.87	2.32	0.708
TN-VPK Participation	**4.69****	**0.71**	**0.000**

*p<.05, **p<.01, †p<.10

To provide some context for the nature and magnitude of the TN-VPK effects on the achievement outcomes, we examined them in relation to the gains made by the children who did not participate in TN-VPK over the period when the TN-VPK participants were attending pre-k. Children not in TN-VPK did, of course, make learning gains during that period as well. We can thus better assess the nature of the TN-VPK effects by considering how much they add to the gains found without TN-VPK. To do that, we first calculated an effect size for the pre-post gain made by the nonparticipants on each measure by dividing the difference between their sample-weighted mean pretest and covariate-adjusted and

sample-weighted mean posttest by the respective pooled sample-weighted standard deviation shown in Table 7 above. This converted both the gain between the beginning and end of the pre-k year and the difference between the TN-VPK participants and nonparticipants at the end of the pre-k year into comparable standard deviation units that allowed their proportional relationship to be examined. To represent that proportional relationship, we divided the effect sizes for the TN-VPK effects on each measure by the pre-post gain effect size for the nonparticipants. This allowed the TN-VPK effects shown in Table 7 to be characterized in terms of the percentage increase in pre-post gain shown by the TN-VPK participants relative to the nonparticipants.

Table 7. TN-VPK Effect Estimates for the Woodcock Johnson Achievement Measures

Outcome Measure	N in TN-VPK Group	N in No TN-VPK Group	TN-VPK Effect Estimate in W-Score Units	*p*-value	Pooled Posttest Standard Deviation[a]	Effect Size
WJ Composite Score	773	304	4.69**	.000	15.06	.31
Literacy Measures						
Letter-Word Identification	773	304	12.10**	.000	26.35	.46
Spelling	773	304	6.29**	.000	24.58	.26
Language Measures						
Oral Comprehension	773	304	2.01*	.025	16.89	.12
Picture Vocabulary	773	304	3.32**	.002	17.94	.19
Math Measures						
Applied Problems	773	304	3.00**	.005	22.54	.13
Quantitative Concepts	773	304	5.08**	.000	15.75	.32

Notes: [a] The standard deviations pooled for use in the calculation of the effect sizes are sample-weighted values that use the same sample weights as used in the analysis that produces the effect estimates.
*$p<.05$, **$p<.01$.

Figure 2 below illustrates the results of these calculations and shows that the effect of TN-VPK was to boost the learning gain on the WJ Composite measure by 38% for the TN-VPK participants relative to the nonparticipants during the pre-k year.

Table 8 shows these relative gain estimates for all six of the Woodcock Johnson scales. The percentage improvement in learning gains for the TN-VPK participants relative to nonparticipants ranged from 21% to 92%. As noted above, the WJ Composite measure, which provides the single best summary of overall academic skills, shows gains during the pre-k year by the children who participated in TN-VPK that were 38% greater than those made over the same period by the children who did not participate in TN-VPK. Their relative gains were even larger in some of the individual skill areas, notably Letter-Word Identification, Picture Vocabulary, and Quantitative Concepts.

Figure 2. Effects of TN-VPK on WJ Composite Score Gain during the Pre-K School Year

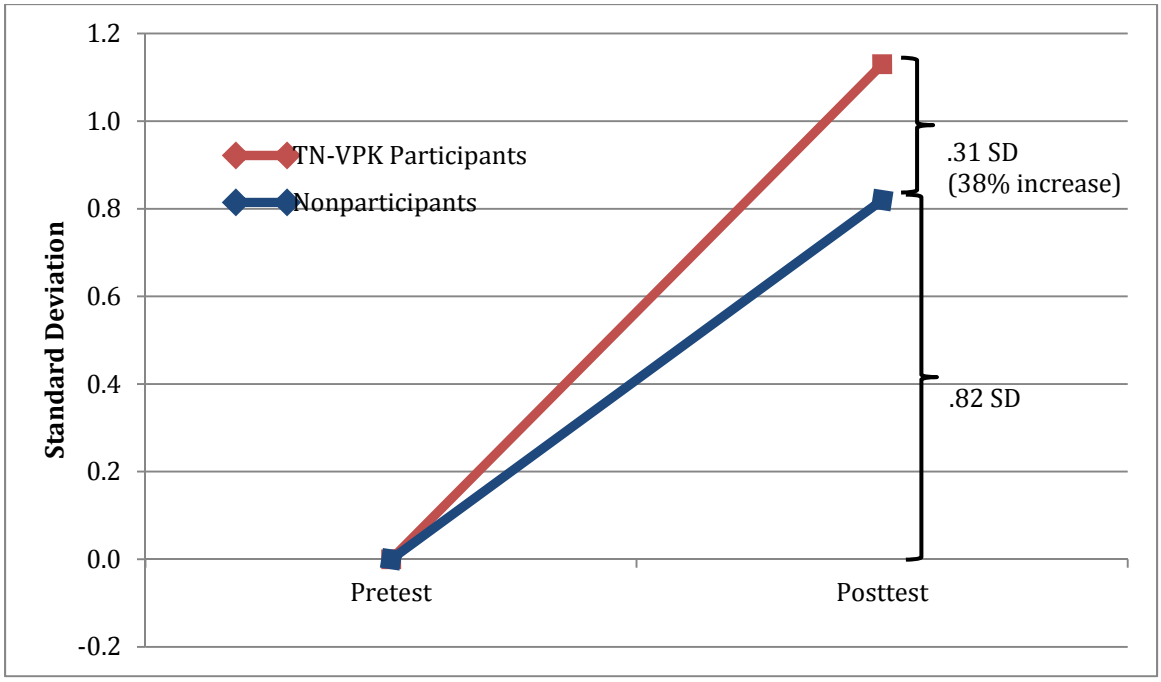


Table 8. TN-VPK Effect Sizes for the WJ Scales Relative to the Pre-Post Gains of the Nonparticipating Children over the Same Period

Outcome Measure	Pre-Post Gain without TN-VPK in Effect Size Units	TN-VPK Effect Size	Improvement with TN-VPK
WJ Composite Score	.82	.31	38%
Literacy Measures			
Letter-Word Identification	.50	.46	92%
Spelling	.80	.26	32%
Language Measures			
Oral Comprehension	.38	.12	31%
Picture Vocabulary	.25	.19	74%
Math Measures			
Applied Problems	.64	.13	21%
Quantitative Concepts	.64	.32	50%

Another perspective is provided by examining the effects of TN-VPK on the standard scores for the Woodcock Johnson achievement measures in contrast to the results on the W-scores that are reported above. The standard scores are calibrated to have a normative value of 100 at each age, based on the norming sample used by the test developers. A mean

of 100 on any of these measures, therefore, indicates that the tested children have a score that is average for their age.

The full results of the analyses with the standard scores on the Woodcock Johnson measures are presented in Appendix D2 and D3. The TN-VPK effects on these measures were all statistically significant and the corresponding effect sizes are quite similar to those shown in Tables 7 and 8 for effects on the W-scores. Figures 3 and 4 below display the unadjusted, unweighted standard scores for each WJ subscale at pretest and posttest, compared to the normative mean of 100, first for the TN-VPK nonparticipants, then for the participants. These figures show how each group gained across the year relative to the normative average score for these tests.

What Figures 3 and 4 reveal is that the children who did not participate in TN-VPK performed below the normative average at the beginning of the pre-k year and, after taking the difference in age between the beginning and end of that year into account, they made very little progress closing those gaps. In contrast, the children who participated in TN-VPK also began the pre-k year with scores below the normative averages but, by the end of the year, they had made significant progress in closing those gaps. Indeed, on the Letter-Word Identification and Applied Problems scales, they had entirely closed the gaps and reached the respective normative averages.

Figure 3. Standard Score Pretest and Posttest Means on the WJ Achievement Measures for TN-VPK Nonparticipants

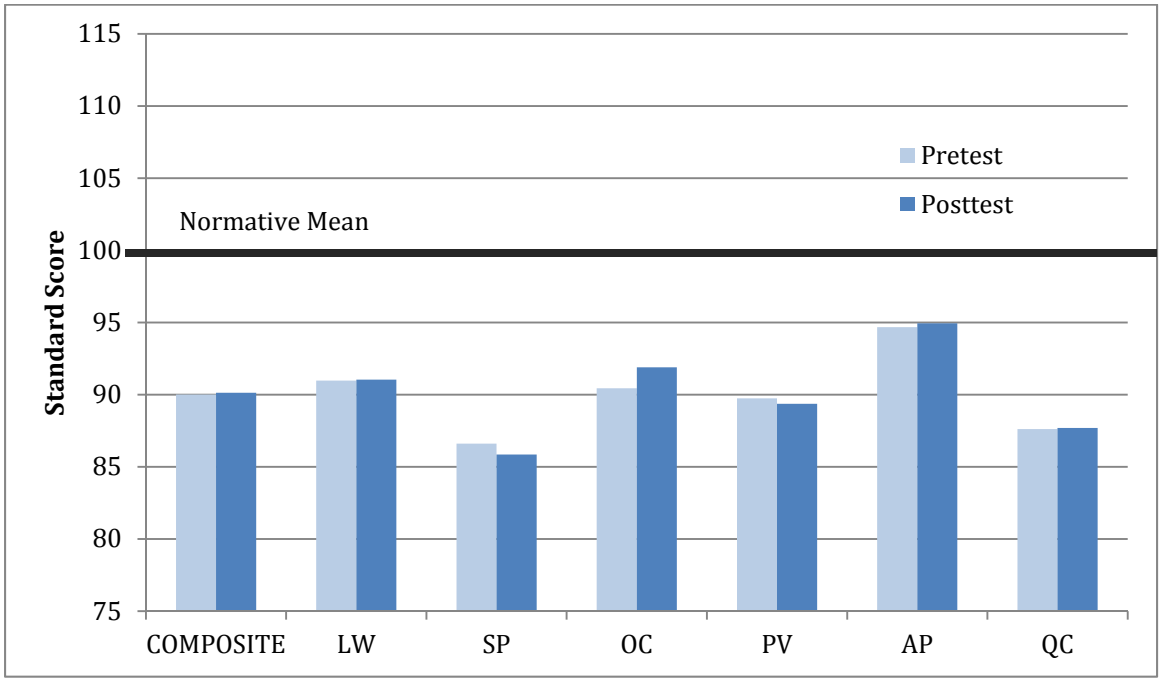
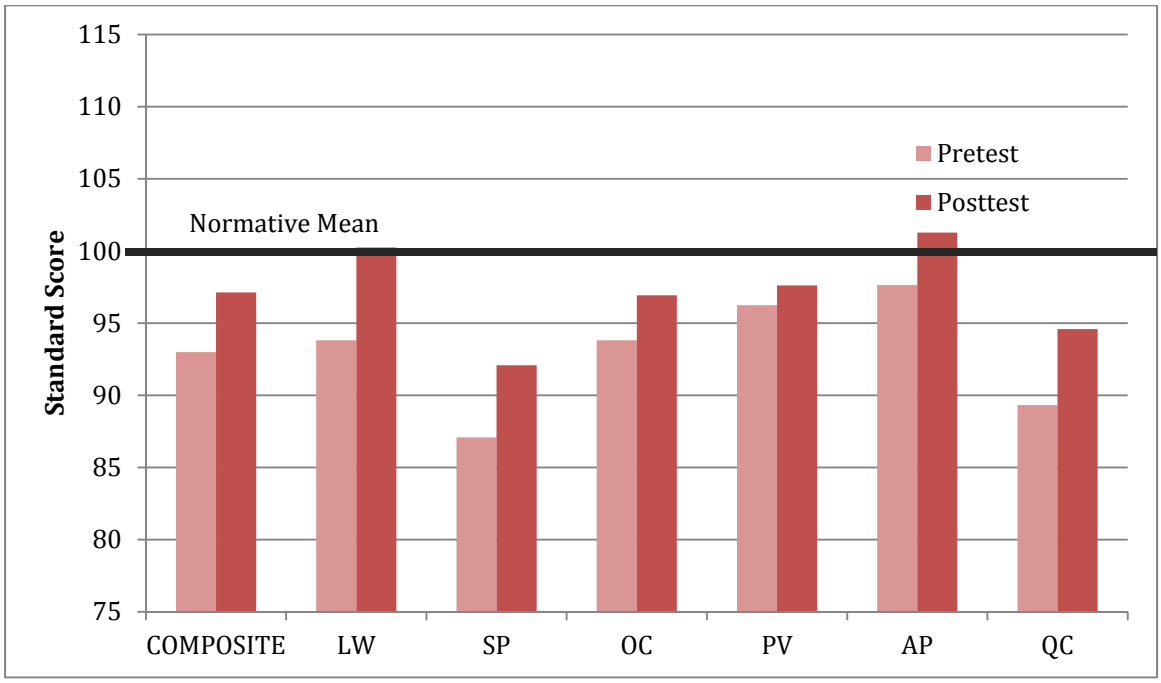


Figure 4. Standard Score Pretest and Posttest Means on the WJ Achievement Measures for TN-VPK Participants



Robustness of the Estimates of the TN-VPK Effects on Achievement. As mentioned earlier, a series of sensitivity analyses was conducted to explore the influence on the TN-VPK effect estimates of alternative choices on key features of the statistical analysis models. These analyses are reported in Appendix E and generally showed that the estimates of the effects of TN-VPK participation on the WJ achievement measures were substantially similar across all the alternative analyses explored.

One of these alternative analyses is especially revealing and warrants a summary here. The erosion of the initial randomization on the randomized applicant lists that occurred in the creation of the Intensive Substudy analysis sample stemmed mainly from parental consent rates that were not high enough to sustain the original randomization. To examine the robustness of the estimates of the TN-VPK effects under more optimal conditions, one of our sensitivity analyses was conducted using only children from the subset of randomized applicant lists that maintained the randomized admission sequence well and had relatively good parental consent rates for both the TN-VPK participant and nonparticipant groups.

We first looked at the TN-VPK participant and nonparticipant consent rates for each of the 76 randomized applicant lists that contributed children to the ISS analysis sample and selected those for which at least 75% of the parents in each group consented. This left us with 23 randomized lists, one from Cohort 1 and 22 from Cohort 2. We then made an additional distinction based on the number of crossovers on each list—children that should have been randomly assigned to TN-VPK or not, but instead ended up in the other group. From the 23 randomized applicant lists with good consent rates, we selected the lists with no more than 20% of the children as crossovers. This left us with 11 lists, all from Cohort 2, that included 136 children who participated in TN-VPK and 88 who did not.

The randomization to TN-VPK participation or not for the children on these 11 randomized applicant lists was clearly not perfect, nor can we assume that these children are fully representative of those in the full ISS analysis sample. Nonetheless, examining the TN-VPK effect estimates for these more completely randomized children provides a check on the plausibility of the estimates that resulted from the analyses reported above that relied much more heavily on statistical control techniques. We therefore estimated the TN-VPK effects on the WJ Composite scale for the children on these 11 lists using several different techniques and compared the resulting effect estimates with that from the analysis approach reported above.

The details for these analyses and their results are provided in Appendix E. Despite the much smaller sample size provided by these 11 randomized applicant lists, all these effect estimates were statistically significant and quite consistent with each other and with the estimate from the full ISS analysis sample. This consistency gives us confidence that our statistical models, despite their complexity, are providing credible estimates of TN-VPK effects.

Differential TN-VPK Effects on the WJ Composite for Different Children. In addition to investigating the overall effects of participation in TN-VPK on the Woodcock Johnson achievement measures at the end of the pre-k year, we were also interested in whether the effects were different for children with certain different characteristics. In particular, we examined whether the TN-VPK effects differed depending on child gender, age at the start of the Pre-K year, baseline skills, or native language. Because ethnicity and native language were confounded (see the crosstabulation in Table 9), we did not examine ethnicity as a moderator but focused only on whether English was the child's native language.

Table 9. Ethnicity by Native Language (Frequency and Percentage for Each Ethnic Group)

Ethnic Group	Native English Speaker	Non-Native English Speaker
Black	233 (95%)	13 (5%)
Hispanic	33 (16%)	168 (84%)
Other	595 (94%)	35 (6%)

For each of these analyses, the statistical model was identical to the ones described above for the main effects of TN-VPK on the Woodcock Johnson measures except for the addition of a term in the multilevel regressions representing the interaction between the TN-VPK participation variable and the respective child characteristic. These analyses, however, were only conducted with the WJ Composite score as the outcome and not repeated for each interaction for each of the six individual WJ scales. Because we found a high degree of overlap between Non-native English speakers and children who began the pre-k year scoring lower than their peers, we analyzed the interactions of TN-VPK participation and native language and baseline pretest score together. In particular, in that analysis we included both the interaction of native English and TN-VPK participation and the interaction of pretest score and TN-VPK participation. (We also tested the three-way

interaction among these but it was not statistically significant and was dropped from the analysis.)

Table 10 displays a summary of the results of these moderator analyses; detailed results can be found in Appendix D4. The interactions of gender and age with TN-VPK participation was not statistically significant, indicating that the effects of TN-VPK on children's WJ Composite score were similar for girls and boys and for younger and older children. There was a statistically significant interaction between TN-VPK participation and native English that showed greater gains for the children who were not native English speakers. The interaction of TN-VPK participation and the pretest scores on the WJ Composite, tested with the native English interaction also in the analysis, was not statistically significant. Thus once the lower pretest scores of the non-native English speakers were taken into account, there were no other differential effects associated with low baseline performance.

Table 10. Coefficients for the Interactions between Selected Moderator Variables and TN-VPK Participation on the WJ Composite Outcome

Moderator	TN-VPK x Moderator Effect Estimate in W-Score Units	Standard Error	*p*-value
Gender (M=0, F=1)	-0.41	1.19	.731
WJ Composite Score pretest	-0.02	0.05	.635
Native English Speaker (No=0, Yes=1)			
Including Hispanic covariate	-5.60*	2.34	.017
Not including Hispanic covariate	-5.65*	2.39	.018

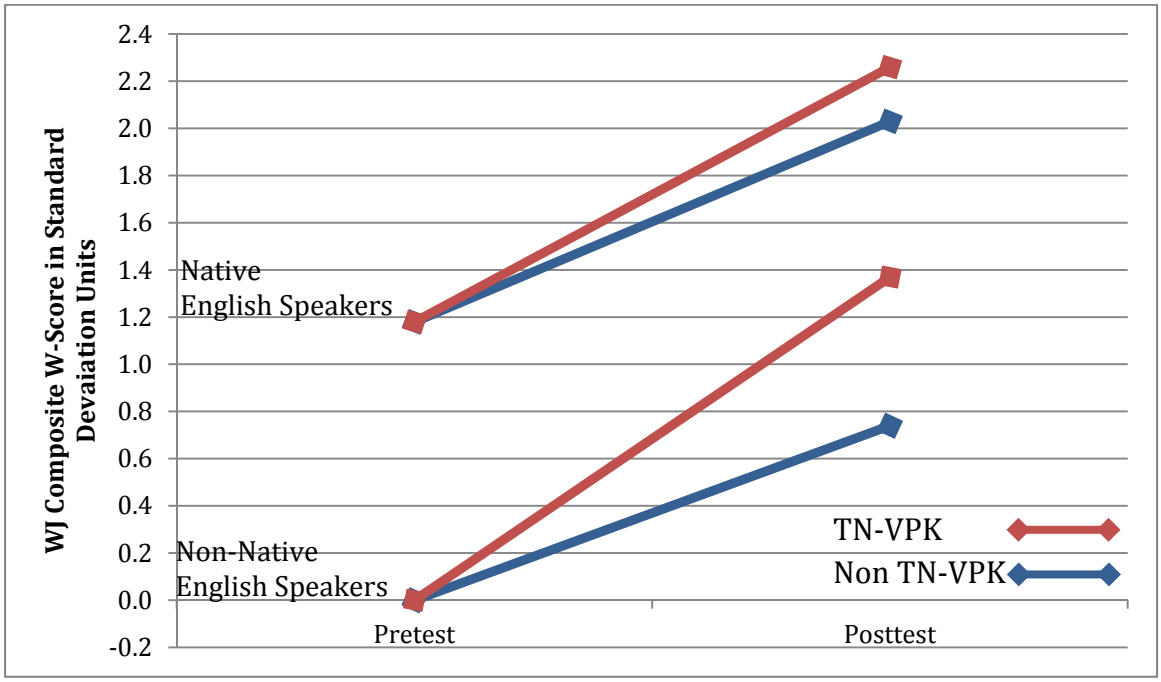
*p<.05, **p<.01

Because of the confounding of native language and ethnicity (Table 9 above), we examined the interaction of native language and TN-VPK participation with Hispanic ethnicity included as a covariate in the analysis and then again with it excluded. As Table 10 shows, the results were virtually identical. This differential effect of TN-VPK for Non-native English speakers is displayed in Figure 5, depicted as differential pre-post gain over the pre-k year represented in standard deviation units that allow differences to be interpreted as standardized mean difference effect sizes.[4]

[4] We determined the values for Figure 5 by first calculating the sample-weighted unadjusted pretest means for the four TN-VPK x native language groups and then averaging the TN-VPK participant and nonparticipant means for each language group. The posttest means for each of the four TN-VPK x native language groups were estimated as the predicted means from the regression model using the respective TN-VPK and native language values and the sample-weighted grand-mean values for all other covariates. Finally, the pre-post gains and posttest TN-VPK effects were converted into standard deviation units (effect sizes) using the pooled TN-VPK participant and nonparticipant sample-weighted posttest standard deviation value, i.e., the same value as used for the results shown in Figure 2 earlier.

When represented in standard deviation units, as in Figure 5, the difference between the TN-VPK participants and nonparticipants on the WJ Composite posttest for each language group is the effect size for that group. We can assess the relative magnitude of these effect sizes, as we did earlier for the overall effect of TN-VPK (Figure 2), by considering how much of an increase in the pre-post gain it represents for the TN-VPK participants in each language group relative to the gain of the respective nonparticipants. Table 11 shows the pre-post gain for the TN-VPK nonparticipants in each language group in standard deviation units, that is, the effect sizes for the pre-post gain, and the effect size for the TN-VPK effect for each language group. As such, it provides some specific values for the gains and differences shown in Figure 5.

Figure 5. Illustration of Native Language as a Moderator of the Effect of TN-VPK Participation on the WJ Composite Gain during the Pre-K School Year



In particular, Table 11 shows that both non-native English speakers and native speakers showed notable gains during the pre-k year whether they were in TN-VPK or not. The effects of TN-VPK for the non-native English speakers, however, were much larger than for the native speakers (effect sizes of .63 vs. .23). The proportionate increase for the non-native English speakers that resulted from participation in TN-VPK, therefore, was much larger than that for native speakers. The non-native English speakers in TN-VPK had 85% greater gains over the pre-k year than their counterparts who were not in TN-VPK. For the native English speakers, those in TN-VPK had gains that were 27% greater than the comparable children who were not in TN-VPK.

Table 11. TN-VPK Effect Sizes for the WJ Composite Score Relative to the Pre-Post Gains of the Nonparticipating Children over the Same Period: Native English Speakers Compared with Non-Native English Speakers

Moderator Group	Pre-Post Gain without TN-VPK in Effect Size Units	TN-VPK Effect Size[b]	Improvement with TN-VPK
Non-Native English speakers [a]	0.74	0.63	85%
Native English speakers [a]	0.85	0.23	27%

[a] Results from the analysis model that included the Hispanic covariate. [b] These effect sizes were estimated without controlling for the pretest by TN-VPK participation interaction, which was not statistically significant when tested with the language interaction in the analysis because of the confounding of low pretest and non-native English. See Appendix D4 for more detail.

Another perspective on the effects of TN-VPK for the non-native English speakers is to consider the performance gap between them and their native English speaking peers. As Figure 5 above makes clear, there is a large performance gap on the WJ Composite at the beginning of the pre-k year between the lower scoring non-native English speakers and the higher scoring native speakers. This gap is 1.18 in standard deviation units, which can be read as an effect size. Figure 6 below illustrates how that gap was considerably reduced by the end of the pre-k year for the children who participated in TN-VPK, but not for those children who did not participate.

Figure 6. The Gap on the WJ Composite between Non-native and Native English Speakers at the Beginning and End of the Pre-K Year for Children who Participated in TN-VPK and Those who Did Not

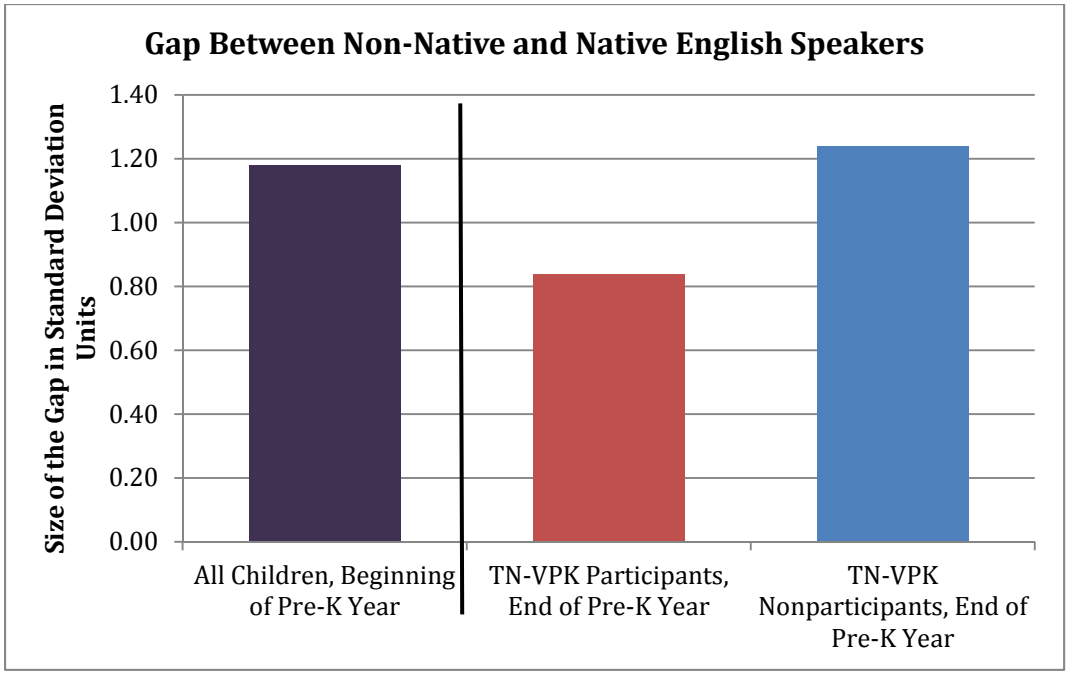


Figure 6 shows the influence on the gap between native and non-native English speakers of the much larger effect TN-VPK had for the non-native speakers. For the children who did not participate in TN-VPK, the performance gap between the non-native and native English speakers actually increased somewhat by the end of the pre-k year.

TN-VPK Effects on the Teacher Reported Measures. Analyses similar to those used for the achievement outcomes were conducted to examine the differences between the kindergarten teachers' ratings of the classroom behavior of the children who had participated in TN-VPK during the prior school year and those who had not participated. There were 914 children in the ISS analysis sample with teacher ratings obtained early in the kindergarten year, 650 who had attended TN-VPK and 264 who had not. To ensure that this reduced sample was not different from the full analysis sample (N=1077), we compared the means and variances on all the baseline variables used in the creation of the propensity scores. The two samples were virtually identical. (See Appendix D5 for the comparisons of baseline variables between these two samples.)

New sampling weights were created for this reduced sample but the same covariates and nesting structure were used in the multilevel models as were used in the analyses reported above for the achievement outcomes with only two differences. Pretest measures of the kindergarten teachers' ratings were not available and we substituted the baseline score on the WJ Composite Scale. Though that is not a pretest for the teacher ratings, it did provide some statistical control for initial differences between the TN-VPK participants and nonparticipants on their cognitive skills and correlated well enough with the teacher ratings to help with statistical power. In addition, new propensity scores were created for this reduced sample using the same variables that went into the propensity scores for the full analysis sample. These new propensity scores showed good covariate balance between the TN-VPK participants and nonparticipants at baseline and did not require any children to be trimmed from the sample because of non-overlapping propensity score values.

Table 12 reports the results of these analyses. There was a statistically significant difference favoring the children who had attended TN-VPK on the kindergarten teachers' ratings of Readiness for Kindergarten ($p<.05$). In addition, there were differences at the $p<.10$ level of significance that also favored the children who had attended TN-VPK on the teachers' ratings of Work-Related Skills and Social Skills. There were no statistically significant differences on the other teacher rating measures. (Details of these analyses are provided in Appendix D6).

The results reported in Table 12 demonstrate that participation in TN-VPK the previous year made at least some difference in school readiness that was apparent to the kindergarten teachers. In addition, it indicates that the kindergarten teachers were relatively discriminating in their ratings. If teachers knew about the prior TN-VPK experience of some of these children and had positive opinions about the program, we would expect them to give better ratings to the children who had attended TN-VPK on all these measures. The fact that they did so only on the measures most closely related to the aims of the TN-VPK program adds credibility to these results.

Table 12. TN-VPK Effect Estimates for Kindergarten Teachers' Ratings

Measure	N in TN-VPK Group	N in No TN-VPK Group	TN-VPK Effect Estimate	*p*-value	Pooled Posttest Standard Deviation[a]	Effect Size
Cooper-Farran Social Behavior	650	264	0.33†	.099	0.96	.34
Cooper-Farran Work-Related Skills	650	264	0.35†	.056	1.16	.30
ACBR Preparedness for Kindergarten	649	264	0.45*	.013	1.46	.31
ACBR Peer Relations	649	264	0.12	.203	0.98	.12
ACBR Behavior Problems	650	264	-0.34	.115	1.29	.26
ACBR Feelings About School	649	264	0.04	.510	0.32	.13

[a] *These pooled values were calculated using the sample-weighted standard deviations for each TN-VPK participation condition.* **p*<.05, ***p*<.01, †*p*<.10

Conclusions

This report presents the results of one component (the Intensive Substudy) of the evaluation of the Tennessee Voluntary Prekindergarten program that has been undertaken by the Peabody Research Institute at Vanderbilt University. It reports on the effectiveness of TN-VPK for increasing the school readiness of two cohorts of participating children during the respective pre-k years on a number of relevant academic and behavioral outcome measures. Those outcome measures are available for a sample of children from a larger randomized study whose parents gave consent for them to be individually assessed at the end of the pre-k year.

Because the consent rates were modest and uneven across participating TN-VPK programs and there was not perfect adherence to the original random assignment, the outcome data from the Intensive Substudy were analyzed with statistical controls to adjust for any initial differences between TN-VPK participants and nonparticipants on a broad set of baseline measures. As a further check on the validity of the results, extensive sensitivity analyses were conducted to ensure that the estimates of the TN-VPK effects were not unduly influenced by different decisions about the analysis approach applied.

The results of these analyses show strong positive and statistically significant improvements on nearly all the key school readiness outcome measures examined for children who attended TN-VPK relative to comparable children who did not attend. The children who participated in TN-VPK showed gains during the pre-k year on early literacy, language, and math achievement measures that ranged from 21% to 92% greater than those of the children who were not in TN-VPK. Moreover, children in TN-VPK who were not native English speakers showed even greater relative gains on the achievement measures.

Ratings by kindergarten teachers early in the fall after the pre-k year also showed positive and statistically significant effects for TN-VPK. These teachers rated the children who had participated in TN-VPK the year before higher on their preparedness for kindergarten than the children who had not attended TN-VPK. With a more modest degree of statistical

significance, they also reported positive effects for TN-VPK on the participating children's work-related skills and social behavior in the classroom.

These results make a convincing case that the Tennessee Voluntary Prekindergarten program is achieving its objective of boosting the school readiness skills of the economically disadvantaged children it serves.

Next Steps

The children in the Intensive Substudy are being followed with annual individual assessments at the end of each grade year that are planned to continue through the end of third grade. At the time this report was written, the children in Cohort 1 of the ISS have completed first grade and the children in Cohort 2 have completed kindergarten. Outcome data from the individual assessments of these children are being combined with data from the State Education Information System on such other outcomes as grade retention, special education placements, disciplinary infractions, and attendance. Analysis of these data will address the question of the effects of TN-VPK in the years after pre-k. Reports on those findings will be forthcoming soon.

The data from the State Education Information System mentioned above are being collected for all of the more than 3000 children in the original randomizations of the two cohorts of children, not only those in the Intensive Substudy. The effects of TN-VPK on the outcomes available from that source for this larger sample will also be analyzed and reported once sufficient data have been collected.

Finally, this evaluation project includes another parallel study that uses a regression-discontinuity design to assess TN-VPK effects at the end of the pre-k year for a representative sample of TN-VPK programs statewide. Data are being collected for that study on rolling basis through four regions across the state. Those data are complete for two of these regions and will soon be complete for a third. When sufficient data have accumulated, analyses will begin and those results will also be reported.

References

Allison, P. (2002). *Missing data.* Thousand Oaks, CA: Sage.

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why and how. *Journal of Experimental Criminology, 2*, 23-44.

Bloom, H. S., Zhu, P., & Unlu, F. (2006). *Finite Sample Bias from Instrumental Variables Analysis in Randomized Trials.* MDRC working paper. Available at http://www.mdrc.org/sites/default/files/Finite%20Sample%20Bias%20from%20Instumental%20Variables%20full%20report.pdf

Cooper, D. H., & Farran, D. C. (1991). *The Cooper-Farran Behavioral Rating Scales*. Brandon, VT: Clinical Psychology Publishing.

Farran, D.C., Bilbrey, C. & Lipsey, M. (2003). *Academic and Classroom Behavior Record.* Unpublished scale available from D.C. Farran, Peabody Research Institute, Vanderbilt University, Nashville, TN.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis. Statistical methods and applications*. Thousand Oaks: Sage.

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

Zhu, P., Jacob, R., Bloom, H. & Xu, Z. (2011). *Designing and Analyzing Studies That Randomize Schools to Estimate Intervention Effects on Student Academic Outcomes Without Classroom-Level Information.* Working Paper. MDRC.

Appendices

Appendix A

Sample Selection & Timeline

A1: Children Contributed to the ISS Analysis Sample by Individual Schools

District	School	Cohort	TN-VPK Participants	Non-participants	Participant Consent Rate	Non-participant Consent Rate
District A	School 1	1	14	2	0.71	0.33
District A	School 1	2	8	7	0.89	0.89
District A	School 2	1	17	1	0.86	0.22
District A	School 2	2	10	6	1.00	0.88
District A	School 3	2	4	10	0.80	1.00
District B	School 4	2	33	7	0.82	0.93
District C	School 5	1	2	3	0.07	0.14
District C	School 5	2	4	4	0.23	0.18
District C	School 6	1	8	17	0.38	0.33
District C	School 6	2	15	20	0.73	0.77
District C	School 7	1	13	8	0.38	0.34
District C	School 7	2	4	1	0.33	0.33
District C	School 8	1	8	2	0.36	0.20
District C	School 9	1	6	3	0.25	0.63
District C	School 10	1	12	1	0.50	0.20
District C	School 10	2	3	1	1.00	0.17
District C	School 11	2	5	6	0.29	0.56
District C	School 12	1	10	6	0.40	0.20
District C	School 12	2	5	12	0.33	0.39
District C	School 13	2	6	9	0.43	0.42
District D	School 14	1	18	1	0.86	1.00
District D	School 14*	2	17	1	0.88	1.00
District E	School 15	2	4	1	1.00	1.00
District F	School 16	2	3	2	1.00	1.00
District G	School 17	2	10	2	1.00	0.50
District G	School 18	2	8	3	0.69	0.59
District G	School 19	2	4	4	0.40	0.78
District G	School 20	2	12	1	0.57	0.13
District G	School 21	2	4	6	1.00	0.58
District G	School 22	2	5	1	1.00	0.33
District G	School 23	1	15	2	0.58	0.25
District G	School 24	2	8	2	1.00	1.00
District H	School 25	2	9	1	0.60	0.13
District H	School 26	1	2	3	1.00	0.60
District H	School 26	2	4	2	0.50	0.25
District I	School 27	2	7	3	0.54	0.43
District I	School 28*	2	9	1	0.83	0.50
District J	School 29	2	2	2	0.50	1.00
District J	School 30	1	1	1	0.50	0.67

District	School	Cohort	TN-VPK Participants	Non-participants	Participant Consent Rate	Non-participant Consent Rate
District K	School 31	1	4	1	0.25	0.50
District K	School 31	2	1	1	0.29	0.33
District K	School 32	2	3	2	0.50	0.83
District K	School 33*	2	1	1	0.40	0.33
District K	School 34	2	10	3	0.71	0.86
District K	School 35	2	2	1	0.40	0.50
District K	School 36	2	2	1	0.50	0.33
District L	School 37*	1	33	1	0.69	0.60
District L	School 37	2	5	3	0.71	0.63
District M	School 38	2	26	1	0.96	1.00
District N	School 39	2	21	6	0.92	1.00
District O	School 40	2	26	5	0.66	0.42
District P	School 41	2	4	3	1.00	1.00
District P	School 42	2	6	4	1.00	1.00
District P	School 43	1	6	3	0.46	0.27
District P	School 43	2	11	6	1.00	0.78
District Q	School 44	1	7	4	0.24	0.24
District R	School 45	1	1	2	0.10	0.29
District R	School 45	2	8	2	0.62	1.00
District S	School 46	1	17	4	0.43	0.17
District S	School 46	2	12	4	0.80	0.80
District S	School 47	2	13	37	0.93	0.87
District S	School 48	2	6	1	0.67	0.50
District S	School 49	2	5	3	0.83	0.60
District S	School 50	2	6	4	1.00	0.86
District T	School 51	1	4	1	0.16	0.25
District T	School 51	2	14	3	0.60	0.60
District T	School 52	2	22	3	1.00	1.00
District T	School 53	1	11	1	0.55	0.17
District T	School 53	2	22	2	0.92	1.00
District T	School 54	1	17	1	0.40	0.13
District T	School 54	2	40	1	1.00	0.25
District U	School 55	2	21	6	0.88	0.88
District U	School 55	1	8	4	0.39	0.33
District U	School 56	2	18	8	0.82	0.89
District U	School 57	2	14	3	0.73	1.00
District U	School 58	2	17	3	0.81	1.00

**Indicates that the randomized school applicant list had insufficient cases with outcome data in either the TN-VPK participant or nonparticipant group to be used in the analysis of the teacher child report data.*

A2: Consent Rates for Children on Randomized Applicant Lists in the ISS Sample

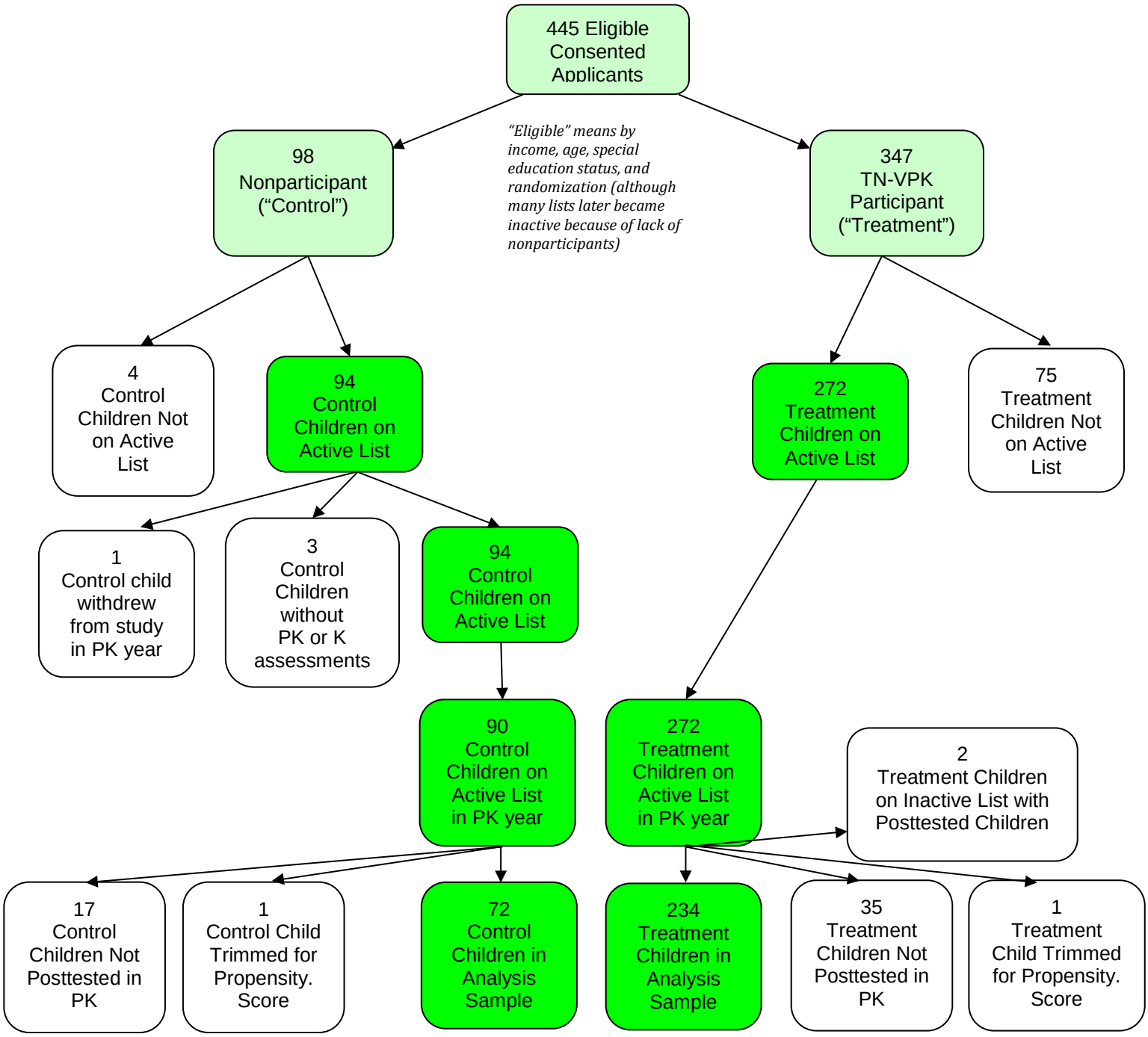
	TN-VPK Nonparticipant Consent Rate					TN-VPK Participant Consent Rate				
	N	Min	Max	M	SD	N	Min	Max	M	SD
Lists Eligible for the ISS Sample	80	0.13	1.00	0.57	0.31	80	0.05	1.00	0.64	0.28
Cohort 1	27	0.13	1.00	0.34	0.20	27	0.05	1.00	0.47	0.26
Cohort 2	53	0.13	1.00	0.68	0.29	53	0.23	1.00	0.74	0.24
Lists in the Analysis Sample	76	0.13	1.00	0.58	0.31	76	0.07	1.00	0.65	0.27
Cohort 1	23	0.13	1.00	0.35	0.22	23	0.07	1.00	0.46	0.24
Cohort 2	53	0.13	1.00	0.68	0.29	53	0.23	1.00	0.74	0.24

A3: Children Consented in the Intensive Substudy and the Selection of Cases for Inclusion in the ISS Analytic Sample

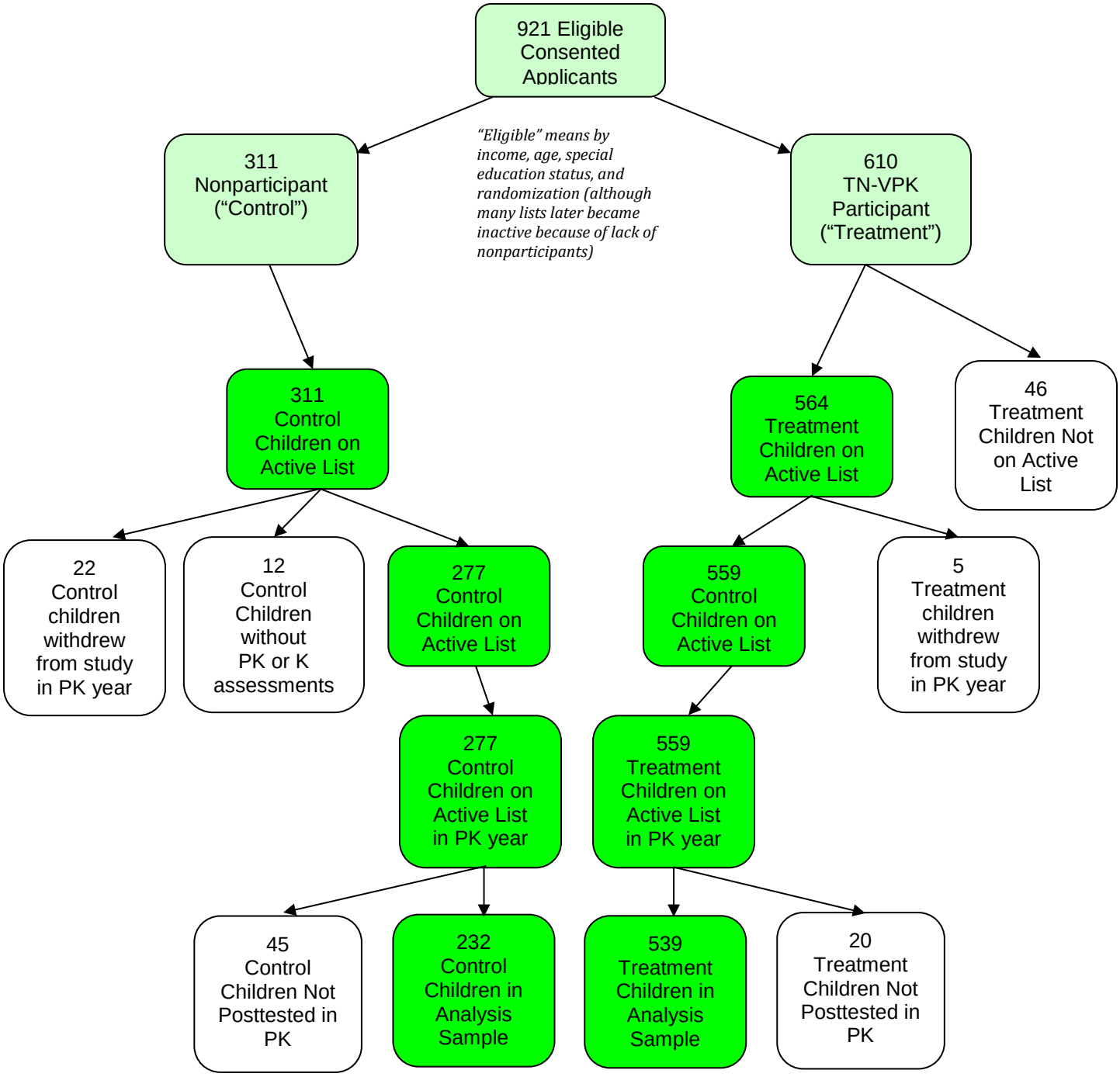
Status of Cases Selected and Not Selected for the ISS Analysis Sample

Source	Cohort 1	Cohort 2	Total
Full ISS Sample			
Not Eligible	83	85	168
Withdrew from study in PK	1	27	28
Never assessed in PK or K	3	12	15
ISS List Ineligible	79	46	125
Sample Eligible	362	836	1198
Active TN-VPK Participant	272	559	831
Active Nonparticipant	90	277	367
Analysis Sample (from Sample Eligible cases above)			
Not Eligible	56	65	121
Not Posttested in K	52	65	117
ISS Analysis Sample List Ineligible	2	0	2
Trimmed b/c of Propensity Score	2	0	2
Sample Eligible	306	771	1077
Active TN-VPK Participant	234	539	773
Active Nonparticipant	72	232	304

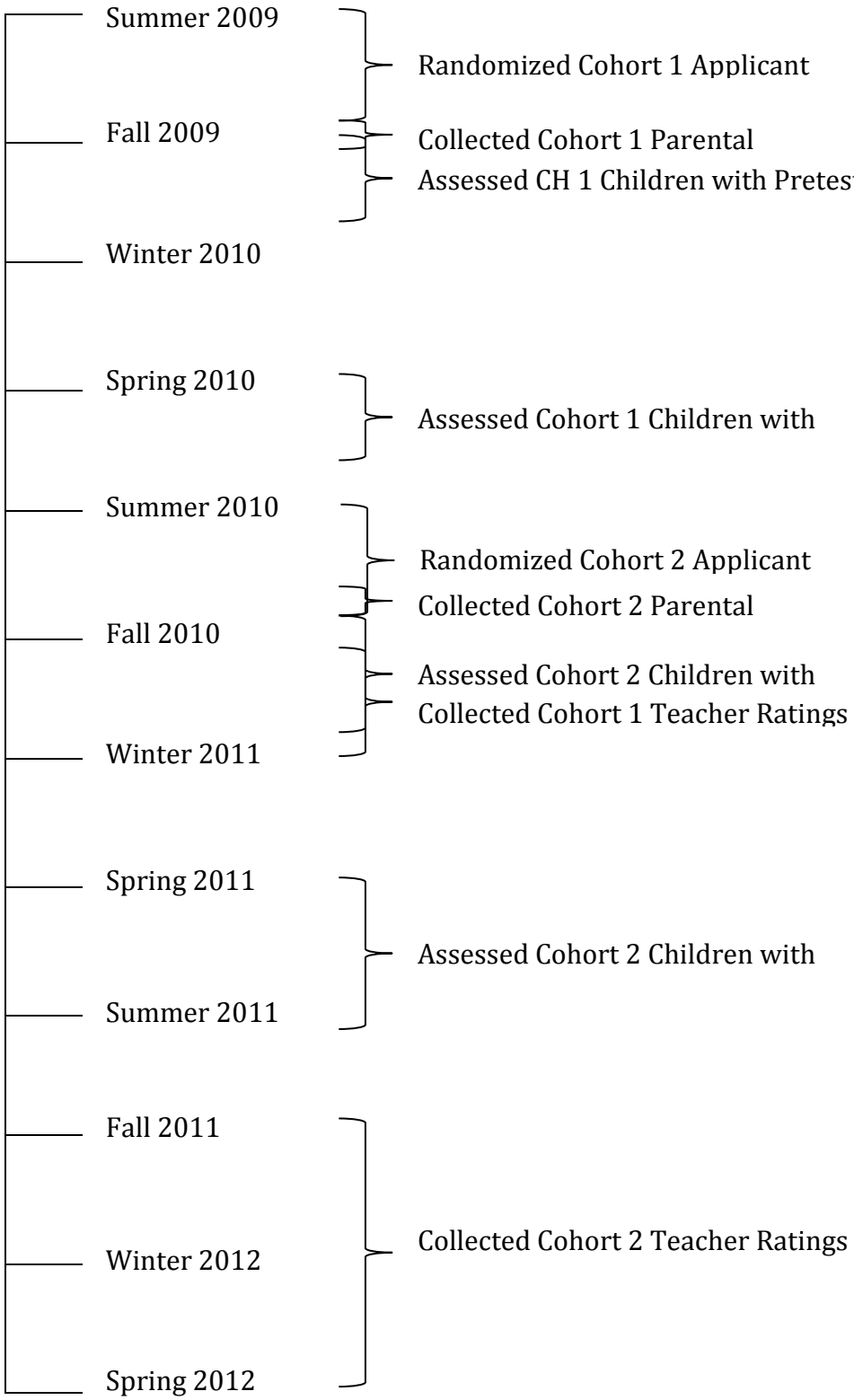
A4: RCT Cohort 1 Intensive Substudy Analysis Sample



A5: RCT Cohort 2 Intensive Substudy Analysis Sample



A6: Timeline of Study Events Related to the Creation of the ISS Analysis Sample



Appendix B

Data Collection Instruments

B1: WJIII Achievement Battery Subtests and Sample Items

Letter-Word Identification

Letter-Word Identification assesses children's letter and word identification ability. Items include identifying and pronouncing letters and words presented to the child.

- Example: *Point to the S*

H C A S B W

- Example: *Point to the word "cat"*

cat my on red

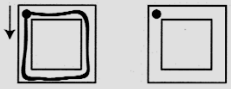
- Example: *What is this letter?*

R N k

Spelling

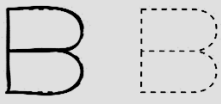
Spelling assesses children's prewriting skills, such as drawing lines and tracing, writing letters, and spelling orally presented words.

- Example:



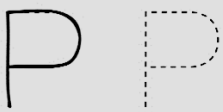
Say: **Watch me.** Start at dot in left square and draw line counterclockwise around square, ending at dot. Hand pencil to subject, point to right square, and say: **Now you draw a line just like I did. Stay on the road.** Collect pencil when subject has finished.

▲ ***Correct:*** line going completely around square, drawn either clockwise or counterclockwise, and inside boundaries at least half of time



Say: **Watch me.** Trace "B" on left. Hand pencil to subject, point to "B" on right, and say: **Now you make one just like I did. Stay on the line.** Collect pencil when subject has finished.

▲ ***Correct:*** "B" that is recognizable if standing alone, has been drawn on broken line at least 75% of time, and has corners that are sharp, not curved



Say: **Watch me.** Trace "P" on left. Hand pencil to subject, point to "P" on right, and say: **Now you do it just like I did. Stay on the line.** Collect pencil when subject has finished.

▲ ***Correct:*** "P" that is recognizable if standing alone and has been drawn on broken line at least 75% of time

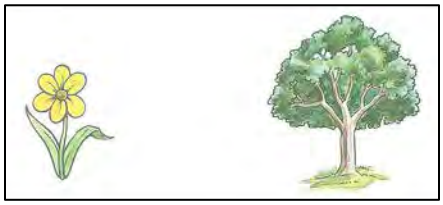
Hand pencil to subject, point to "S," and say: **Now you make one just like this. Make it here** (point to space to right of "S").

▲ ***Correct:*** "S" that is recognizable if standing alone

Picture Vocabulary

Picture Vocabulary assesses children's ability to name objects presented in pictures and point to the picture that goes with a word; it measures early language development and lexical knowledge.

- Example: *Put your finger on the flower.*



- Example: *What is this?*

Oral Comprehension

Oral Comprehension assesses children's ability to complete analogies and provide words with similar or different meanings from key words; it measures their listening ability and understanding.

- Example: *Listen carefully and finish what I say.*

"A bird flies, a fish _____."

"We ride in _____."

"Water looks blue and grass looks _____".

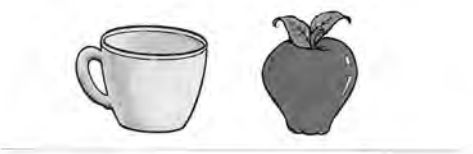
"Houses are for people, garages are for _____."

Applied Problems

Applied Problems assesses children's ability to solve numerical and spatial problems presented verbally with accompanying pictures of objects.

- Example:

How many apples are there in this picture?



How many boats are there?

How many birds are there?

Quantitative Concepts

Quantitative Concepts assesses children's knowledge of mathematical concepts, including vocabulary, numbers, shapes, sequences, and symbols; it measures aspects of quantitative reasoning and math knowledge.

- Example:

What is this called?

Point to the largest star.
Now point to the smallest star.

What number is this?

This is the first house (points).
*Point to the **last** house.*
*Point to the **middle** house.*

B2: Teacher Child Report Forms and Sample Items

Cooper-Farran Behavioral Rating Scales

Work Related Skills

The Work Related Skills score is a factor score derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about the child's ability to work independently, listen to the teacher, remember and comply with instructions, complete games and activities, function within designated time periods, and otherwise behave appropriately in relation to classroom work and prescribed activities.

- Example: *PLEASE READ THE DESCRIPTORS AND THEN CIRCLE THE NUMBER THAT BEST DESCRIBES THE CHILD ON THAT ITEM. YOU MAY USE THE EVEN-NUMBERED POINTS IF THE CHILD'S SKILL FALLS BETWEEN THE BEHAVIORAL DESCRIPTORS.*

  1. *Performance of Daily Nonacademic Tasks*

1	2	3	4	5	6	7
Cheerfully does own chores, then takes on extra duties		Independently attends to routines		Will do chores, but only with prodding		Often refuses to perform daily chores

Social Behavior

The Social Behavior score is a factor score also derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about the child's social interactions with peers including appropriate behavior while participating in group activities, play, and outdoor games; expression of feelings and ideas during group discussions; and response to others mistakes or misfortunes. The same instructions as given to teachers in the previous example on this page apply to these questions, as well.

- Example: *Response to Helpful Criticism from Teacher*

1	2	3	4	5	6	7
Accepts criticism easily, uses it to improve performance		Attends to criticism, but does not apply it		Ignores or pretends not to hear criticism		Angrily rejects teacher's attempts to help

Academic and Classroom Behavior Record (ACBR)

Readiness for Kindergarten

The Readiness for Kindergarten score is a factor score derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about how well prepared the child is for kindergarten in literacy and language skills, math skills, and social behavior.

- Example: *How well prepared for kindergarten work is this child in **math**?*

1	2	3	4	5	6	7
Extremely well prepared; understands numbers, shapes, and patterns		Above average preparation for this grade level		Slightly below average preparation for this grade level		Very unprepared for this grade level; cannot count to ten or identify shapes

Liking for School

The Liking for School score is a factor score derived from a principal components factor analysis that summarizes the teachers' responses to a set of interrelated items about the child liking or disliking school, having fun at school, enjoying and engaging in classroom activities, and seeming happy at school.

- Example: *How do you think this child feels about school? Circle response to each.*

a.	Likes to Come to School	Always	Sometimes	Never
b.	Dislikes School	Always	Sometimes	Never

Behavior Problems

The Behavior Problems score relates to whether the child has shown any behavior problems from a list including explosive and overactive behaviors, attention problems, physical or relational aggression, social withdrawal or anxiety, motor difficulties, and the like. Teachers' responses were coded into two outcome variables: (1) a dichotomous variable indicating whether any of the behavior problems on the list have been observed (0= no, 1=yes), and (2) the number of behavior problems on the list that were observed.

- Example: *This year, has the child had any of the following behavior problems that you feel caused concern? Please check all that apply.*

    - ○ Explosive Behaviors (e.g., temper outbursts, easily provoked, unpredictable behavior)
    - ○ Attention Problems (e.g., had difficulty concentrating or staying on task)
    - ○ Overactive Behaviors (e.g., acts impulsively without thinking, disrupts ongoing activities)

B3: Parent Questionnaire

Your Child's Educational and Care Arrangements

1. Which of the following best describes your child's educational care arrangements?

___ Public pre-K classroom Name of school: _____

What is the teacher's name: _____

___ Head Start Center Name of center: _____

___ Private Child Center Name of Center: _____

___ Home-based Child Care in other than parent's/guardian's home

___ Stay at home with someone other than the parent

___ Stay at home with the parent? Which parent _____

2. Is your child now attending an additional center or school, such as a "before" or "after" school program at a school or center on a regular basis? □ Yes □ No

If yes, explain: _____

3. How satisfied are you with the educational and care arrangements for your child?

a. very dissatisfied
b. somewhat dissatisfied
c. somewhat satisfied
d. very satisfied

4. Do you feel the activities your child engages in during the day will help prepare her/him for kindergarten next year?

a. no b. somewhat c. yes

5. Are you at all concerned about how well s/he will do in kindergarten?

a. very concerned
b. a little concerned
c. not sure
d. not very concerned
e. not at all concerned

What Your Child Does at Home

6. How many hours per day does your child watch TV or videos on:

a. Monday-Friday _____ b. Saturday _____ c. Sunday _____

7. About how many children's books do you have in your home now, including library books? Please only include books that are for children. _____

8. Does your family subscribe to newspapers/magazines? □ Yes □ No

If yes, a. # Newspapers___ b. # Magazines___ c. # Child Magazines___

9. Does anyone in your home have a library card? □ Yes □ No

If Yes, how often is it used?

Almost never	Once or twice a year	Every few months	More than once a month	At least once a week
1	*2*	*3*	*4*	*5*

You and your Child

The next questions are about your child's activities with family members.

10. In a typical **week**, how often do you or any other family member do the following things with your child? *Use the following descriptors to rate how often.*

Almost Never	Not every week but sometimes	1 - 3 times	4 - 6 times	Daily
1	*2*	*3*	*4*	*5*

a)	Sing songs	*1*	*2*	*3*	*4*	*5*
b)	Do chores, like setting the table or caring for pets	*1*	*2*	*3*	*4*	*5*
c)	Play games or do puzzles	*1*	*2*	*3*	*4*	*5*
d)	Play sports or exercise	*1*	*2*	*3*	*4*	*5*
e)	Practice the names of letters	*1*	*2*	*3*	*4*	*5*
f)	Practice reading words	*1*	*2*	*3*	*4*	*5*
g)	Practice letter sounds	*1*	*2*	*3*	*4*	*5*
h)	Read to your child	*1*	*2*	*3*	*4*	*5*
i)	Talk about what happened at school	*1*	*2*	*3*	*4*	*5*
j)	Play board or card games	*1*	*2*	*3*	*4*	*5*
k)	Play with blocks	*1*	*2*	*3*	*4*	*5*
l)	Play counting games or sing counting songs	*1*	*2*	*3*	*4*	*5*
m)	Count different things	*1*	*2*	*3*	*4*	*5*
n)	Play with electronic toys (LeapFrog)	*1*	*2*	*3*	*4*	*5*
o)	Play on computer	*1*	*2*	*3*	*4*	*5*
p)	Read to yourself	*1*	*2*	*3*	*4*	*5*
q)	Your spouse or partner read to him/herself?	*1*	*2*	*3*	*4*	*5*

11. Has your child ever been referred for special education services? ☐ Yes ☐ No

a. *If "yes,"* did your child qualify for services? ☐ Yes ☐ No

If "yes,"

b. What was your child's diagnosis? _____

c. How old was your child when s/he was diagnosed? _____

d. What services did/does your child receive? _____

You and Your Family

12. Which of the following educational levels have you completed? (Circle all that apply)

a. 8th grade
b. some high school
c. graduated high school or GED
d. vocational/technical school
e. some college
f. associate's degree
g. bachelor's degree
h. master's degree

13. Tell us about your work (if applicable).

Job/Position _____

Employer _____

14. Which of the following educational levels has your spouse/partner completed? (Circle all that apply)

a. 8th grade
b. some high school
c. graduated high school or GED
d. vocational/technical school
e. some college
f. associate's degree
g. bachelor's degree
h. master's degree
i. don't know

15. Tell us about your spouse/partner's work (if applicable).

Job/Position _____

Employer _____

16. Is any language other than English regularly spoken in your home? ☐ Yes ☐ No

a. If **"yes"**, what is the **primary** language spoken in your home? (Circle one)

English Spanish Other_____

17. Was your child born in the United States? ☐ Yes ☐ No

a. ***If "no,"*** In what year did your child come to the United States to stay?

_____

Appendix C

Data Description for the Analytic Sample

C1. Demographic Characteristics by Cohort of the Children in the ISS Analysis Sample

CHARACTERISTIC	**Cohort 1** (N=306) Mean	**Cohort 2** (N=771) Mean
Age at start of pre-k year (months)	51.8	51.8
White	.48	.61
Black	.28	.21
Hispanic	.23	.17
Asian and others	.01	.02
Male	.50	.47
Native English speaker	.76	.82

C2. Family Context Information from the Parent Questionnaire by Cohort for the ISS Analysis Sample

	Cohort 1					Cohort 2				
VARIABLE	N	Min	Max	Mean	SD	N	Min	Max	Mean	SD
Mother's Education (1-4)	304	1	4	2.06	0.70	727	1	4	2.11	0.75
Library Card Use (0-2)	304	0	2	0.86	0.85	714	0	2	0.93	0.83
Newspaper Subscriptions (0-3)	302	0	2	0.38	0.78	725	0	3	0.35	0.77
Magazine Subscriptions (0-2)	304	0	2	0.29	0.53	725	0	2	0.27	0.51
Number of Working Parents	304	0	2	1.10	0.67	730	0	2	1.31	0.61
Weekday TV hours watched (1-4)	304	1	4	2.17	0.64	727	1	4	1.94	0.80
Saturday TV hours watched (1-4)	304	1	4	2.37	0.94	726	1	4	2.07	0.91
Sunday TV hours watched (1-4)	297	1	4	2.17	0.90	725	1	4	1.92	0.87
Number children's books (0-5)	303	0	5	3.93	1.19	728	0	5	4.13	1.26
How often parents read to child (1-5)	304	1	5	4.26	1.00	727	1	5	4.28	1.00
How often parents count (1-5)	304	1	5	4.17	1.12	727	1	5	4.17	1.13

Mother's Education (1=less than high school, 2=high school diploma/GED, 3=associate's degree, 4=more than associate's degree); Library Card Use (0=no card/used almost never, 1=used once or twice a year or every few months, 2=used more than once a year or at least weekly); Newspaper Subscriptions (0=0, 1=1, 2=2-3, 3=>3); Magazine Subscriptions (0=0, 1=1-3, 2=>3); TV hours watched (1=< 1, 2=1-3, 3=3-5, 4=>4); Number children's books (0=0, 1=1-5, 2=6-10, 3=11-19, 4=20-50, 5=>50); How often read/count with child (1=Almost never, 2=Not weekly but sometimes, 3=1-3 times/week, 4=4-6 times/week, 5=Daily).

C3. WJ Standard Score Pretest and Posttest Means for Analysis Sample by Cohort (Unadjusted Unweighted)

	Cohort 1					Cohort 2				
WJ SUBSCALE	N	Min	Max	Mean	SD	N	Min	Max	Mean	SD
Pretest										
WJ Composite Score	306	62.5	118.0	92.0	11.2	771	59.7	129.3	92.2	11.4
Letter-Word Identification	306	61	132	93.1	12.9	771	60	164	93.0	13.8
Spelling	306	51	126	88.0	13.3	771	39	139	86.5	15.3
Oral Comprehension	306	59	130	92.3	13.4	771	46	130	93.1	13.3
Picture Vocabulary	306	26	129	93.2	20.1	771	26	130	94.9	18.2
Applied Problems	306	56	127	96.8	13.7	771	58	134	96.8	13.6
Quantitative Concepts	306	60	125	88.7	12.3	771	56	133	88.9	12.7
Posttest										
WJ Composite Score	306	56.3	122.5	95.7	11.7	771	47.3	127.0	94.9	11.2
Letter-Word Identification	306	60	150	98.0	13.0	771	56	165	97.5	13.7
Spelling	306	37	141	91.9	14.2	771	32	143	89.7	14.0
Oral Comprehension	306	64	132	95.2	14.8	771	63	138	95.6	14.5
Picture Vocabulary	306	21	130	94.6	17.7	771	16	131	95.6	15.7
Applied Problems	306	54	129	100.9	12.5	771	44	132	98.9	13.4
Quantitative Concepts	306	50	135	93.7	14.5	771	53	136	92.2	14.2

C4. Pre- and Post- WJ W-Scores for the Total Analytic Sample and Each Cohort (unadjusted and unweighted)

TOTAL ANALYTIC SAMPLE	N	Min	Max	Mean	SD
Pretest					
Age at Pretest (months)	1044	45	71	54.1	3.6
Letter-Word ID	1044	264	467	318.9	27.4
Spelling	1044	277	432	351.1	28.6
Oral Comprehension	1044	418	485	443.5	16.3
Picture Vocabulary	1044	374	491	454.9	23.3
Applied Problems	1044	318	453	390.8	27.9
Quantitative Concepts	1044	386	461	407.8	14.0
WJ Composite Score	1044	339.5	454.0	394.5	18.1
Posttest					
Age at Posttest (months)	1077	52	75	61.1	3.5
Letter-Word ID	1077	264	491	341.2	27.5
Spelling	1077	277	468	375.0	25.3
Oral Comprehension	1077	418	496	451.5	17.0
Picture Vocabulary	1077	374	501	461.7	18.7
Applied Problems	1077	318	458	408.1	24.1
Quantitative Concepts	1077	386	470	421.0	16.1
WJ Composite Score	1077	341.2	463.7	409.8	17.1
Test Lag	1044	27	183	71.7	26.4
Test Interval	1044	80	295	196.8	29.2

BY COHORT	Cohort 1					Cohort 2				
	N	Min	Max	Mean	SD	N	Min	Max	Mean	SD
Pretest										
Age at Pretest (months)	303	47	71	54.5	3.7	741	45	63	54.0	3.5
Letter-Word ID	303	264	408	319.1	26.2	741	264	467	318.7	27.9
Spelling	303	287	432	353.8	26.3	741	277	432	349.9	29.4
Oral Comprehension	303	418	485	442.9	16.4	741	418	485	443.7	16.3
Picture Vocabulary	303	374	491	453.5	25.1	741	374	491	455.4	22.5
Applied Problems	303	318	436	391.3	27.7	741	318	453	390.7	28.0
Quantitative Concepts	303	386	451	407.8	13.6	741	386	461	407.7	14.2
WJ Composite Score	303	343.2	441.	394.7	17.8	741	339.0	454.0	394.4	18.3
Posttest										
Age at Posttest (months)	306	55	75	60.9	3.6	771	52	70	61.2	3.5
Letter-Word ID	306	264	470	341.6	26.6	771	264	491	341.1	27.9
Spelling	306	287	468	377.3	25.6	771	277	464	374.0	25.1
Oral Comprehension	306	418	489	451.1	17.2	771	418	496	451.7	16.9
Picture Vocabulary	306	374	494	460.7	20.4	771	374	501	462.1	17.9
Applied Problems	306	318	453	410.1	21.8	771	318	458	407.3	24.9
Quantitative Concepts	306	386	465	421.9	16.0	771	386	470	420.6	16.2
WJ Composite Score	306	354.2	458.	410.5	17.4	771	341.2	463.7	409.5	17.0
Test Lag (days)	303	42	139	82.4	27.4	741	27	183	67.3	24.6
Test Interval	303	115	238	181.7	29.9	741	80	295	202.9	26.5

C5. Kindergarten Teacher Ratings for the Total Analytic Sample and Each Cohort (unadjusted and unweighted)

TOTAL ANALYTIC SAMPLE					
	N	**Min**	**Max**	**Mean**	**SD**
Cooper-Farran					
Social Skills	964	1.95	7.00	5.74	0.92
Work-Related Skills	964	1.44	7.00	4.99	1.14
ACBR					
Preparedness for K	963	1.00	7.00	4.50	1.45
Peer Relations	963	1.00	7.00	5.21	0.99
Behavior Problems	964	0.00	9.00	0.80	1.25
Feelings About School	963	0.33	2.00	1.76	0.31
Test Lag (days)	931	406.0	856.0	444.8	41.2
Age at Rating	931	58.0	84.0	66.9	3.7

	Cohort 1					**Cohort 2**				
BY COHORT	**N**	**Min**	**Max**	**Mean**	**SD**	**N**	**Min**	**Max**	**Mean**	**SD**
Cooper-Farran										
Social Skills	277	3.24	7.00	5.82	0.85	687	1.95	7.00	5.71	0.95
Work-Related Skills	277	1.56	7.00	5.12	0.99	687	1.44	7.00	4.94	1.19
ACBR										
Preparedness for K	276	1.00	7.00	4.65	1.40	687	1.00	7.00	4.45	1.46
Peer Relations	276	1.00	7.00	5.29	0.92	687	1.00	7.00	5.18	1.01
Behavior Problems	277	0.00	9.00	0.63	1.11	687	0.00	7.00	0.87	1.30
Feelings About School	276	0.67	2.00	1.78	0.29	687	0.33	2.00	1.75	0.32
Test Lag (days)	244	430.0	616.0	457.0	26.7	687	406.0	856.0	440.5	44.5
Age at Rating (months)	244	61.0	81.0	67.5	3.6	687	58.0	84.0	66.7	3.8

C6. Covariate Balance with Different Propensity Score Variants Represented as Effect Sizes for Baseline Differences between the TN-VPK Participant and Nonparticipant Groups

COVARIATE	Without propensity score	Logit as covariate	Probability as covariate	Inverse Probability as weight (IPWT)	Sample weights + Logit as covariate	Sample weights + Probability as covariate
Native English Speaker	0.03	-0.05	-0.06	0.04	-0.02	-0.01
Black	0.05	0.02	0.00	-0.04	0.01	0.00
Hispanic	-0.01	0.04	0.06	0.01	0.04	0.05
Male	0.00	-0.03	0.00	-0.02	0.05	0.06
Library Card Use	0.13	-0.02	-0.02	0.07	0.00	0.00
Newspaper Subscriptions	0.07	0.03	0.01	-0.03	0.02	0.02
Magazine Subscriptions	0.11	0.05	0.01	-0.03	-0.06	-0.06
Mother's Education	0.15	-0.03	-0.03	0.03	-0.06	-0.06
N of Working Parents	0.02	-0.01	0.01	0.03	-0.04	-0.04
Age at Pretest	-0.19	-0.03	0.00	0.00	-0.02	0.01
Test Lag	-0.49	0.05	0.09	-0.09	0.02	0.04
Test Interval	0.08	-0.05	-0.06	0.03	-0.06	-0.08
WJ Letter-Word ID	0.14	-0.03	-0.04	0.05	0.00	0.01
WJ Spelling	0.05	-0.03	-0.01	0.02	-0.02	0.00
WJ Oral Comprehension	0.08	-0.07	-0.07	-0.01	-0.10	-0.09
WJ Picture Vocabulary	0.10	-0.04	-0.08	-0.01	-0.04	-0.05
WJ Applied Problems	0.02	-0.05	-0.06	0.02	-0.06	-0.06
WJ Quantitative Concepts	0.02	-0.03	-0.02	0.04	0.00	0.02
WJ_Composite_Score	0.09	-0.05	-0.06	0.03	-0.04	-0.03

C7. Summary for Inclusion Rates across the Randomized Applicant Lists from the Schools Contributing to the ISS Analytic Sample

Inclusion Rates for the Total Analytic Sample					
	N of Lists	**Min**	**Max**	**Mean**	**SD**
TN-VPK Nonparticipants	76	0.08	1.00	0.43	0.26
TN-VPK Participants	76	0.07	1.00	0.63	0.27

INCLUSION RATES BY COHORT	**Cohort 1**					**Cohort 2**				
	N of Lists	**Min**	**Max**	**Mean**	**SD**	**N of Lists**	**Min**	**Max**	**Mean**	**SD**
TN-VPK Nonparticipants	23	0.08	1.00	0.27	0.19	53	0.09	1.00	0.49	0.26
TN-VPK Participants	23	0.07	1.00	0.44	0.24	53	0.14	1.00	0.72	0.25

C8. Within Proportion of Variance and Unconditional ICC's for Randomized Applicant Lists (Schools) and School Districts

	Level 1 (Child)	**Level 2 (R-List/School)**		**Level 3 (School District)**	
OUTCOME	% Variance	ICC	p-value	ICC	*p*-value
WJ Composite Score	0.87	0.13**	.000	0.00	.307
WJ Letter-Word ID	0.91	0.09**	.000	0.00	.264
WJ Spelling	0.91	0.05**	.001	0.04**	.004
WJ Oral Comprehension	0.87	0.11**	.000	0.02	.198
WJ Picture Vocabulary	0.85	0.12**	.000	0.03†	.078
WJ Applied Problems	0.87	0.10**	.000	0.04*	.018
WJ Quantitative Concepts	0.93	0.06**	.000	0.02	.198
Cooper-Farran Social Behavior	0.99	0.00	>.50	0.01*	.036
Cooper-Farran Work-Related Skills	0.85	0.03*	.015	0.11	.190
ACBR Preparedness for Kindergarten	0.94	0.03*	.014	0.03*	.040
ACBR Peer Relations	0.96	0.03†	.073	0.01†	.093
ACBR Behavior Problems	0.99	0.01	.169	0.00	.361
ACBR Feelings About School	1.00	0.00	.400	0.00	.121

* $p<.05$ ** $p<.01$, †$p<.10$

Appendix D

Detailed Analysis Results for TN-VPK Effects

D1. Full Model Results for the Analysis of the Effects of TN-VPK on the End of Pre-K WJ W-Scores

	WJ Composite		Letter-Word Identification		Spelling		Oral Comprehension		Picture Vocabulary		Applied Problems		Quantitative. Concepts	
VARIABLE	b	*p*	b	*p*	b	*p*	b	*p*	b	*p*	b	*p*	b	*p*
Intercept	74.8	0.000	109.6	0.000	174.2	0.000	131.0	0.000	131.	0.000	136.9	0.000	69.7	0.000
Inclusion Rate: Nonpartic.	-0.80	0.631	-4.41	0.288	3.45	0.422	0.14	0.927	-0.03	0.987	-0.58	0.876	-0.96	0.698
Inclusion Rate: Participant	-0.58	0.739	-0.41	0.927	-4.80	0.270	-2.57	0.095	-0.49	0.792	-1.44	0.708	-4.06	0.076
Inclusion Rate Interaction	-8.77	0.237	-10.7	0.611	-28.0	0.140	0.48	0.937	-2.87	0.708	8.93	0.391	-1.78	0.854
Cohort (CH 2 = reference)	0.70	0.445	-2.75	0.225	-0.64	0.797	0.29	0.703	0.28	0.752	1.68	0.502	0.22	0.858
Black	1.40	0.044	3.09	0.052	6.57	0.000	-1.81	0.041	0.46	0.500	0.14	0.908	0.46	0.734
Hispanic	1.46	0.059	6.16	0.001	4.21	0.083	-4.59	0.002	-1.14	0.350	0.96	0.605	0.25	0.867
Male	0.47	0.404	-0.28	0.752	-2.95	0.008	1.08	0.106	2.08	0.014	-1.82	0.118	0.30	0.613
Native English Speaker	-1.09	0.412	3.68	0.071	-2.78	0.315	4.62	0.000	0.78	0.692	0.38	0.868	-2.17	0.235
Library Card Use	-0.22	0.424	-1.25	0.261	0.37	0.641	0.74	0.024	0.21	0.657	-0.39	0.324	0.45	0.281
Newspaper Subscriptions	-0.05	0.829	-0.07	0.942	-1.10	0.080	-0.16	0.760	-0.27	0.498	0.09	0.848	0.55	0.075
Magazine Subscriptions	0.43	0.374	1.10	0.418	0.68	0.577	-0.17	0.844	1.08	0.022	1.66	0.089	0.18	0.795
Mother's Education	-0.17	0.627	0.26	0.820	1.13	0.230	0.31	0.556	-0.83	0.168	1.20	0.104	0.19	0.598
Number of Wrkg. Parents	-0.02	0.969	0.25	0.842	0.11	0.909	0.20	0.714	0.46	0.265	0.54	0.496	-0.65	0.290
Age at Pretest	-0.09	0.405	-0.09	0.686	0.14	0.486	0.09	0.301	0.17	0.115	0.35	0.034	0.21	0.070
Test Lag	-0.09	0.009	-0.06	0.414	-0.06	0.482	-0.04	0.248	0.10	0.073	0.02	0.798	-0.07	0.043
Test Interval	-0.01	0.815	0.01	0.917	-0.02	0.751	0.00	0.922	0.12	0.005	0.07	0.111	0.02	0.499
Pretest	0.88	0.000	0.70	0.000	0.54	0.000	0.71	0.000	0.63	0.000	0.57	0.000	0.83	0.000
Propensity Score	-0.87	0.708	4.57	0.325	9.72	0.104	-1.39	0.491	3.43	0.195	11.72	0.026	0.93	0.700
TN-VPK Participation	**4.69**	**0.000**	**12.10**	**0.000**	**6.29**	**0.000**	**2.01**	**0.025**	**3.32**	**0.002**	**3.00**	**0.005**	**5.08**	**0.000**

Level 1 and Level 2 Variance Components

	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p
Intercept	2.26	0.011	7.23	0.050	5.29	0.005	0.57	0.278	0.96	0.006	0.19	>.500	3.25	0.006
Level 1 Residual	49.68		307.17		314.82		92.64		85.6		220.10		107.06	

Level 3 Variance Components

	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p
Intercept	0.05	0.213	4.29	0.025	10.52	0.004	0.13	0.377	2.66	0.002	17.89	0.000	0.04	>.500

D2. Full Model Results for the Analysis of the Effects of TN-VPK on the End of Pre-K Standard Scores

	WJ Composite		Letter-Word Identification		Spelling		Oral Comp.		Picture Vocabulary		Applied Problems		Quantitative. Concepts	
VARIABLE	b	*p*	b	*p*	b	*p*	b	*p*	b	*p*	b	*p*	b	*p*
Intercept	40.2	0.000	63.8	0.000	72.5	0.000	38.7	0.000	5.28	0.575	39.0	0.000	44.5	0.001
Inclusion Rate: Nonparticipants	-0.12	0.926	-1.78	0.384	2.30	0.347	0.48	0.709	0.29	0.862	0.43	0.839	-0.58	0.801
Inclusion Rate: Participants	-0.86	0.538	-0.64	0.772	-3.33	0.188	-2.78	0.048	-0.47	0.787	-1.25	0.538	-3.91	0.077
Inclusion Rate Interaction	-6.33	0.241	-4.55	0.663	-13.2	0.209	-0.10	0.985	-3.52	0.605	2.03	0.745	-0.36	0.967
Cohort (Cohort 2 = reference)	0.60	0.394	-1.35	0.240	-0.38	0.802	0.26	0.705	0.40	0.585	1.30	0.312	0.30	0.796
Black	0.96	0.073	1.51	0.053	3.46	0.001	-1.56	0.054	0.30	0.615	0.02	0.979	0.40	0.761
Hispanic	1.14	0.110	3.08	0.001	2.41	0.107	-3.77	0.003	-0.82	0.478	0.75	0.533	0.35	0.823
Male	0.27	0.516	-0.26	0.560	-1.90	0.005	0.88	0.151	1.75	0.020	-1.18	0.038	0.09	0.871
Native English Speaker	-0.57	0.576	2.02	0.025	-1.32	0.419	4.10	0.000	1.27	0.471	0.61	0.642	-1.80	0.314
Library Card Use	-0.18	0.394	-0.66	0.204	0.11	0.792	0.63	0.028	0.15	0.729	-0.17	0.425	0.34	0.369
Newspaper Subscriptions	-0.03	0.882	0.02	0.965	-0.56	0.150	-0.08	0.866	-0.28	0.438	0.02	0.941	0.53	0.082
Magazine Subscriptions	0.28	0.428	0.45	0.492	0.35	0.619	-0.24	0.752	0.94	0.037	0.81	0.140	0.15	0.808
Mother's Education	-0.07	0.787	0.16	0.762	0.72	0.155	0.33	0.485	-0.56	0.301	0.62	0.111	0.39	0.221
Number of Working Parents	-0.04	0.914	0.05	0.929	0.06	0.907	0.15	0.755	0.46	0.209	0.17	0.731	-0.68	0.234
Age at Pretest	-0.28	0.001	-0.53	0.000	-0.55	0.000	-0.16	0.018	0.00	0.964	-0.24	0.014	-0.30	0.006
Test Lag	-0.05	0.066	-0.02	0.526	0.00	0.985	-0.03	0.326	0.08	0.048	0.02	0.548	-0.06	0.087
Test Interval	-0.03	0.127	-0.03	0.414	-0.04	0.311	-0.02	0.281	0.08	0.015	0.02	0.513	-0.02	0.497
Pretest	0.85	0.000	0.66	0.000	0.55	0.000	0.73	0.000	0.67	0.000	0.62	0.000	0.79	0.000
Propensity Score	-0.62	0.729	2.31	0.297	6.66	0.051	-1.34	0.473	2.67	0.251	6.85	0.009	0.69	0.754
TN-VPK Participation	**3.52**	**0.000**	**6.07**	**0.000**	**3.71**	**0.000**	**2.02**	**0.014**	**3.03**	**0.001**	**1.83**	**0.002**	**4.62**	**0.000**
Level 1 and Level 2 Variance Components														
	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p
Intercept	1.27	0.003	1.96	0.026	2.01	0.003	0.42	0.255	0.37	0.022	0.12	0.298	2.89	0.004
Level 1 Residual	26.22		73.58		99.18		72.01		69.98		68.79		89.53	
Level 3 Variance Components														
	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p
Intercept	0.18	0.203	0.96	0.034	4.48	0.000	0.20	0.275	1.84	0.003	4.68	0.000	0.04	>.500

D3. Effect Sizes for WJ Standard Scores

OUTCOME MEASURE	Pre-Post Gain without Pre-K in ES Units	Pre-K Effect Size
WJ Composite Score	.01	.33
Literacy Measures		
Letter-Word Identification	.01	.47
Spelling	-.05	.27
Language Measures		
Oral Comprehension	.10	.14
Picture Vocabulary	-.02	.19
Math Measures		
Applied Problems	.02	.15
Quantitative Concepts	.01	.33

D4. Full Model Results for the Analysis of the Moderator Effects on the WJ Composite Score

	Moderator: Gender		**Moderator: Age at Start of Pre-K Year**		**Moderator: Native English Speaker (with Hisp. covariate)**		**Moderator: Native English Speaker (no Hisp. covariate)**	
VARIABLE	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	74.9	0.000	70.0	0.000	414.1	0.000	414.3	0.000
Inclusion Rate: Nonparticipants	-0.85	0.613	-0.89	0.598	-0.34	0.844	-0.34	0.837
Inclusion Rate: Participants	-0.54	0.756	-0.51	0.768	-0.54	0.759	-0.51	0.774
Inclusion Rate Interaction	-8.66	0.247	-8.69	0.237	-9.29	0.189	-9.11	0.201
Cohort (Cohort 2=Reference)	0.70	0.446	0.67	0.460	0.96	0.352	0.95	0.352
Black	1.39	0.046	1.40	0.043	1.48	0.034	1.34	0.044
Hispanic	1.48	0.052	1.50	0.040	1.42	0.101	NA	NA
Male	0.77	0.410	0.46	0.427	0.57	0.311	0.53	0.347
Native English Speaker	-1.06	0.434	-1.09	0.424	2.83	0.140	1.98	0.317
Library Card Use	-0.22	0.431	-0.21	0.442	-0.25	0.344	-0.26	0.337
Newspaper Subscriptions	-0.06	0.807	-0.05	0.856	-0.09	0.703	-0.09	0.713
Magazine Subscriptions	0.44	0.372	0.45	0.368	0.28	0.556	0.29	0.546
Mother's Education	-0.16	0.640	-0.17	0.643	-0.06	0.851	-0.11	0.757
Number of Working Parents	-0.01	0.976	-0.03	0.945	-0.03	0.941	-0.01	0.978
Age at Pretest	-0.09	0.406	0.01	0.960	-0.08	0.447	-0.07	0.466
Test Lag	-0.09	0.009	-0.09	0.006	-0.08	0.022	-0.08	0.023
Test Interval	-0.01	0.812	-0.01	0.803	0.00	0.998	0.00	0.965
Pretest	0.87	0.000	0.88	0.000	0.89	0.000	0.89	0.000
Propensity Score	-0.82	0.719	-0.84	0.713	-0.68	0.777	-0.46	0.845
TN-VPK Participation	4.85	0.000	4.66	0.000	9.10	0.000	9.16	0.000
Pretest * TN-VPK Participation	NA	NA	NA	NA	-0.02	0.635	-0.02	0.641
TN-VPK Partic. * Moderator	**-0.41**	**0.731**	**-0.16**	**0.500**	**-5.60**	**0.017**	**-5.65**	**0.018**
Level 1 and Level 2 Variance Components								
	VC	p	VC	p	VC	p	VC	p
Intercept	2.25	0.011	2.18	0.013	2.32	0.006	2.02	0.012
Level 1 Residual	49.68		49.63		48.40		48.66	
Level 3 Variance Components								
	VC	p			VC	p	VC	p
Intercept	0.05	0.210	0.05	0.215	0.09	0.227	0.16	0.194

D5: Comparison of Baseline Variables between the ISS Analysis Sample with WJ Outcome Measures and the Reduced Analysis Sample with Teacher Reported Outcome Measures

	ISS Analysis Sample (N=1077)		Sample with Teacher Measures (N=914)	
VARIABLE	Mean	SD	Mean	SD
Proportion TN-VPK participants	0.72	0.45	0.71	0.45
Black	0.23	0.42	0.22	0.42
Hispanic	0.19	0.39	0.20	0.40
Male	0.48	0.50	0.47	0.50
Native English Speaker	0.80	0.40	0.79	0.41
Library Card Use	0.91	0.83	0.91	0.83
Newspaper subscription	0.37	0.76	0.38	0.77
Magazine subscription	0.28	0.51	0.28	0.52
Mother's Education	2.09	0.73	2.10	0.74
Number of Working Parents	1.25	0.63	1.27	0.62
Age at Pretest (months)	54.6	3.59	54.7	3.65
WJ Letter-Word Identification Pretest	318.5	27.4	318.4	27.4
WJ Spelling Pretest	350.8	28.7	351.1	28.6
WJ Applied Problems Pretest	390.6	27.8	390.4	28.2
WJ Picture Vocabulary Pretest	454.5	23.3	454.2	24.0
WJ Oral Comprehension Pretest	443.3	16.3	443.5	16.6
WJ Quantitative Concepts Pretest	407.5	14.1	407.6	14.1
WJ Composite Score Pretest	394.2	15.9	394.2	16.2

D6. Full Model Results for the Analysis of the Effects of TN-VPK on the Kindergarten Teacher Reported Outcomes

	CF Behavioral Skills		CF Work-Related Skills		ACBR Preparedness for K		ACBR Peer Relations		ACBR Behavior Problems		ACBR Feelings About School	
VARIABLE	b	p	b	p	b	p	b	p	b	p	b	p
Intercept	0.52	0.470	-10.83	0.000	-18.85	0.000	-4.09	0.011	10.40	0.000	-1.05	0.026
Inclusion Rate: Nonpart.	-0.17	0.449	0.08	0.742	-0.10	0.730	-0.03	0.889	-0.36	0.147	0.03	0.717
Inclusion Rate: Participants	0.12	0.595	-0.12	0.673	-0.23	0.462	0.10	0.643	0.01	0.961	0.01	0.910
Inclusion Rate Interaction	1.18	0.160	0.96	0.158	0.83	0.447	-0.28	0.726	-1.70	0.151	0.15	0.531
Cohort (Cohort 2 = reference)	0.02	0.864	0.14	0.194	0.03	0.809	0.12	0.137	-0.24	0.022	0.03	0.143
Black	0.00	0.984	0.09	0.242	0.11	0.182	-0.10	0.278	-0.02	0.826	0.03	0.106
Hispanic	0.41	0.003	0.44	0.000	0.53	0.000	0.19	0.001	-0.60	0.000	0.14	0.000
Male	-0.25	0.018	-0.39	0.005	-0.16	0.189	-0.08	0.300	0.43	0.001	-0.07	0.036
Native English Speaker	-0.23	0.033	-0.46	0.001	-0.30	0.110	-0.25	0.004	0.47	0.000	-0.08	0.010
Library Card Use	-0.07	0.081	-0.08	0.004	-0.12	0.005	-0.08	0.021	0.10	0.005	-0.02	0.056
Newspaper Subscription	-0.02	0.645	0.06	0.074	0.08	0.233	0.03	0.171	-0.08	0.066	0.02	0.040
Magazine Subscription	0.07	0.215	-0.01	0.877	-0.05	0.436	0.00	0.956	0.03	0.729	0.00	0.866
Mother's Education	0.02	0.529	0.01	0.778	0.06	0.209	0.02	0.654	-0.11	0.005	0.03	0.003
Number of Working Parents	-0.03	0.449	0.01	0.921	-0.09	0.225	0.13	0.008	-0.02	0.773	0.00	0.983
Age at Pretest	0.01	0.463	0.01	0.077	0.00	0.693	-0.01	0.128	-0.01	0.239	0.00	0.047
Test Lag	0.00	0.027	0.00	0.170	0.00	0.128	0.00	0.981	0.00	0.018	0.00	0.057
Pretest	0.02	0.000	0.04	0.000	0.06	0.000	0.02	0.000	-0.03	0.000	0.01	0.000
Propensity Score	-0.22	0.448	0.01	0.979	0.29	0.216	0.20	0.238	-0.05	0.897	0.08	0.422
TN-VPK Participation	**0.33**	**0.099**	**0.35**	**0.056**	**0.45**	**0.013**	**0.12**	**0.203**	**-0.34**	**0.115**	**0.04**	**0.510**
Level 1 and Level 2 Variance Components												
	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p
Intercept	0.02	0.113	0.06	0.000	0.09	0.000	0.01	0.101	0.02	0.007	0.00	0.027
Level 1 Residual	0.77		0.84		1.28		0.82		1.33		0.09	
Level 3 Variance Components												
	VC	p	VC	p	VC	p	VC	p	VC	p	VC	p
Intercept	0.01	0.130	0.00	>.500	0.00	>.500	0.00	0.278	0.00	>.500	0.00	0.434

Appendix E

Sensitivity Analyses

E1. Description of the Sensitivity Analyses

In order to examine the robustness of the findings presented in the main body of this report, we ran several different sensitivity analyses that addressed different issues. Detailed results for each of these analyses are provided in the remainder of this appendix, but here we provide a brief description of the analytic approach and results of each of these analyses.

1. ***Unbalanced Design***: Because each of the randomized applicant lists from the participating schools had different proportions of TN-VPK participants and nonparticipants, we **used sample weights** in our main analyses that weighted the data from the participants and nonparticipants on each list equally. However, we also ran those same analyses without the sampling weights and the results were similar. The table presented as Appendix E2 below reports the results from our main analytic model without sampling weights and without propensity scores as well as models with various combinations of sampling weights and ways of incorporating the propensity scores.
2. ***Baseline Non-Equivalence***: In the main text of this report we described how the propensity scores were created and used to deal with differences in baseline characteristics between the TN-VPK participant and nonparticipant groups. In all of the analyses presented in the main text of the report, the propensity score probability was used as a covariate because it had the best performance in balancing the baseline covariates (see Appendix C6 above). However, we also investigated **using the propensity score in the analysis as a covariate in log-odds (logit) form**. The result was almost identical with the model with the propensity score probability as a covariate (Appendix E2 below).
3. ***Inclusion Rate Covariates and their Interaction***: In the main analysis, we incorporated inclusion rates for the TN-VPK participant and nonparticipant groups, and their interaction, to take account of the fact that different proportions of the full randomization sample were represented for each of these groups in the Intensive Substudy sample. We ran our main analysis model **without these inclusion rate variables** and the result was nearly identical (Appendix E2 below).
4. ***Nesting Structure:*** In our main analysis, we used a three-level HLM with students nested within randomized applicant lists (R-lists), and R-lists nested within cohort-specific school districts with cohort included as a covariate. In that analysis, we coded each R-List as being associated with a distinct unique school, though this was not always actually the case—some schools provided more than one R-List, raising a question as to whether schools and R-Lists should be separated as nesting levels in the analysis. We therefore **explored analysis models with different nesting structures**:
    - If no schools provided multiple R-Lists (i.e., contributed to both cohorts), the data structure is clear: children (L1) are nested within R-lists (L2, same as schools), R-lists are nested within districts (L3), and districts are nested within cohorts (L4). Because there were only two cohorts, level-4 can be treated as a fixed effect with cohort coded as a dummy variable, and a 3-level HLM can be

used to estimate the impact effects. This is the nesting structure we assumed in our main analysis model even though it is not exactly aligned with the data because about one-third of the schools contributing to the ISS analysis sample did, in fact, contribute R-Lists to both cohorts.

- With the partial nesting of R-Lists from each cohort within schools, the actual data structure in the ISS analysis sample was complicated. Complete specification of the cross-classified model that would best represent this structure involves spatial and temporal nesting:
  - Spatial nesting: children (L1) nested within R-lists (L2), (some) R-lists nested within schools (L3), and schools nested within districts (L4).
  - Temporal nesting: (some) R-lists (L2) nested within cohort (L3).

Our attempts to analyze the data using this partially cross-classified data structure in a 4-level model were not successful. In HLM, SAS, and SPSS, attempts to specific this data structure resulted in analyses that would not converge. We therefore explored a reduced model in which (a) cohort was treated as a fixed effect (dummy-coded covariate) and, (b) because there are at most two R-lists (L2) available per school (L3), we collapsed L3 to L2, i.e., nested children (L1) within R-lists (L2), and R-lists within districts (L3). In this reduced model, the original dependences between schools and R-Lists are moved to the district level (Zhu, Jacob, Bloom, & Xu, 2011). There are two options for coding district in this reduced model: (1) using the natural district as level 3 (Model 1), and (2) using the cohort-specific district as level 3, i.e., giving each district a different ID code for each cohort to which it contributed (Model 2). The results for Models 1 and 2 are in Appendix E3 below. Note that Model 2 is what we used in the main analyses presented in this report. The parameter estimates for the TN-VPK effect were nearly identical for these two models.

5. ***Inconsistent Randomization and Consent Rates***: As described in the main body of the report, the Intensive Substudy sample included only children with parental consent. The modest consent rates for many of the R-Lists undermined the initial randomization that produced the R-List. In addition, some consented children did not participate in the TN-VPK condition to which the R-List assigned them, most often when parents of a child not admitted to the school providing the R-List were able to get their child admitted at another TN-VPK school. In order to estimate the robustness of the TN-VPK effects under more optimal conditions, we ran the same analyses we used with the full ISS analytic sample **using only those school lists that maintained their randomization relatively well and had good consent rates** for both TN-VPK participant and nonparticipant groups. We determined which randomized lists were best through a multi-stage process that took account of the consent rates for each individual randomized list that contributed children to the analysis sample and the number of crossovers on each list (children who ended up in a different condition than the one to which the randomized list assigned them). This selection identified 11 lists that implemented the best randomization. We then applied various analysis options to estimate the TN-VPK effects for the children assigned to conditions from these randomized lists.

The first of these analysis options was the same procedure we used in the main analyses presented in this report. The effect size estimate resulting from that analysis was then compared with the estimates from a set of alternative analysis approaches to assess their robustness in the face of that variation. Details about this procedure and the alternative analyses used are described in Appendix E4 below. The results are summarized in Appendix E5 and more fully reported in Appendices E6-E11. Overall, this exercise revealed little variation in the effect estimates, suggesting that our results are robust to the different analytic methods investigated.

E2. Full Results of Sensitivity Analyses Investigating Propensity Score and Sampling Weight Combinations

	(1) Without sample weights, without propensity score		(2) With sample weights, without propensity score		(3) Without sample weights, with propensity score probability		(4) With sample weights, with propensity score probability		(5) With sample weights, with propensity score logit		(6) With sample weights, with propensity score probability, without inclusion rates	
VARIABLE	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	68.1	<0.001	72.8	<0.001	70.4	<0.001	75.0	<0.001	74.4	<0.001	72.0	<0.001
Inclusion Rate: Nonpart.	-0.34	0.803	-0.71	0.654	-0.45	0.759	-0.80	0.631	0.72	0.447	NA	
Inclusion Rate: Participant	-0.91	0.648	-0.68	0.679	-0.81	0.689	-0.58	0.737	1.39	0.040	NA	
Inclusion Rate Interaction	-10.07	0.133	-8.43	0.255	-10.48	0.113	-8.89	0.227	1.47	0.054	NA	
Cohort (Cohort 2 = reference)	0.99	0.372	0.63	0.464	1.07	0.344	0.72	0.430	0.48	0.396	0.75	0.363
Black	1.23	0.010	1.41	0.045	1.22	0.010	1.40	0.044	-1.09	0.393	1.45	0.038
Hispanic	1.26	0.124	1.44	0.065	1.30	0.113	1.47	0.059	-0.22	0.428	1.47	0.055
Male	0.15	0.748	0.47	0.402	0.15	0.753	0.47	0.403	-0.05	0.828	0.43	0.446
Native English Speaker	-0.65	0.459	-1.13	0.383	-0.60	0.525	-1.10	0.411	0.43	0.407	-1.08	0.420
Library Card Use	-0.21	0.470	-0.23	0.422	-0.20	0.474	-0.22	0.425	-0.16	0.620	-0.25	0.377
Newspaper Subscription	0.03	0.911	-0.07	0.796	0.04	0.867	-0.05	0.833	-0.02	0.963	-0.07	0.782
Magazine Subscription	0.32	0.512	0.45	0.350	0.29	0.541	0.43	0.375	-0.09	0.415	0.47	0.331
Mother's Education	0.35	0.330	-0.18	0.586	0.36	0.331	-0.17	0.630	-0.09	0.005	-0.17	0.628
Number of Working Parents	-0.09	0.812	-0.01	0.980	-0.10	0.794	-0.02	0.967	-0.01	0.766	-0.01	0.977
Age at Pretest	-0.05	0.604	-0.09	0.411	-0.05	0.595	-0.09	0.404	0.88	<0.001	-0.09	0.391
Test Lag	-0.07	<0.001	-0.08	<0.001	-0.08	0.008	-0.09	0.008	-0.14	0.733	-0.07	0.046
Test Interval	0.01	0.624	0.00	0.979	0.00	0.950	-0.01	0.798	-0.80	0.625	0.00	0.922
Pretest	0.87	<0.001	0.87	<0.001	0.87	<0.001	0.88	<0.001	-0.59	0.733	0.87	<0.001
Propensity Score	NA		NA		-1.00	0.661	-0.93	0.688	-8.94	0.238	-0.11	0.961
TN-VPK Participation	**4.93**	**<0.001**	**4.62**	**<0.001**	**5.08**	**<0.001**	**4.69**	**<0.001**	**4.67**	**<0.001**	**4.67**	**<0.001**
Level 1 and Level 2 Variance Components												
Random Effect	VC	p-value	VC	p-value	VC	p-value	VC	p-value	VC	p-value	VC	p-value
Intercept	2.37	0.012	2.18	0.037	2.48	0.010	2.32	0.030	2.25	0.031	2.16	0.042
Level 1 Residual	51.81		49.72		51.75		49.67		49.70		49.74	
Level 3 Variance Components												
Random Effect	VC	p-value	VC	p-value	VC	p-value	VC	p-value	VC	p-value	VC	p-value
Intercept	0.04	0.431	0.09	0.190	0.01	>.500	0.01	0.226	0.02	0.186	0.35	0.083

E3: Sensitivity Analysis for Investigating the Nesting Structure

	(1) Students nested within R-lists (Cohort-specific schools), R-list nested within District (natural district); Cohort is at level 2 (R-lists).			**(2) Students nested within R-lists (Cohort-specific schools), R-list nested within Cohort-specific-District; Cohort is at level 3 (district).**		
FIXED EFFECT	b	SE	*p*-value	b	SE	*p*-value
Intercept	73.4	9.42	<0.001	75.0	9.87	<0.001
Inclusion Rate: Nonparticipant	-0.74	1.96	0.704	-0.80	1.66	0.631
Inclusion Rate: Participant	-0.49	1.62	0.761	-0.58	1.72	0.737
Inclusion Rate Interaction	-7.90	6.40	0.217	-8.89	7.35	0.227
Cohort (Cohort 2 = reference)	0.52	0.93	0.579	0.72	0.91	0.435
Black	1.37	0.62	0.027	1.40	0.69	0.044
Hispanic	1.39	0.49	0.005	1.47	0.78	0.059
Male	0.44	0.52	0.397	0.47	0.57	0.403
Native English Speaker	-1.05	1.44	0.466	-1.10	1.33	0.411
Library Card Use	-0.24	0.25	0.330	-0.22	0.28	0.425
Newspaper Subscription	-0.08	0.21	0.709	-0.05	0.25	0.833
Magazine Subscription	0.45	0.47	0.343	0.43	0.49	0.375
Mother's Education	-0.19	0.35	0.591	-0.17	0.34	0.630
Number of Working Parents	-0.01	0.40	0.985	-0.02	0.46	0.967
Age at Pretest	-0.09	0.11	0.429	-0.09	0.10	0.404
Test Lag	-0.08	0.04	0.053	-0.09	0.03	0.008
Test Interval	0.00	0.03	0.965	-0.01	0.03	0.798
Pretest	0.87	0.02	<0.001	0.88	0.02	<0.001
Propensity Score	-0.04	2.74	0.987	-0.93	2.31	0.688
TN-VPK Participation	**4.65**	**0.69**	**<0.001**	**4.69**	**0.71**	**<0.001**
Level 1 and Level 2 Variance Components						
Random Effect	SD	Variance	*p*-value	SD	Variance	*p*-value
Intercept	1.28	1.63	0.026	1.52	2.32	0.03
Level 1 Residual	7.05	49.73		7.05	49.67	
Level 3 Variance Components						
Random Effect	SD	Variance	*p*-value	SD	Variance	*p*-value
Intercept	0.84	0.71	0.048	0.07	0.01	0.192

E4. Analyses Using Only the Children from the Most Fully Implemented Randomized Applicant Lists

The erosion of the initial randomization on the randomized applicant lists that occurred in the creation of the Intensive Substudy analysis sample stemmed mainly from parental consent rates that were not high enough to sustain the original randomization. To examine the robustness of the estimates of the TN-VPK effects under more optimal conditions, one of our sensitivity analyses was conducted using only children from those randomized applicant lists that maintained the randomized admission sequence well and had relatively good parental consent rates for both the TN-VPK participant and nonparticipant groups.

We first looked at the TN-VPK participant and nonparticipant consent rates for each of the 76 randomized applicant lists that contributed children to the ISS analysis sample and selected those for which at least 75% of the parents in each group consented. This left us with 23 randomized lists, one from Cohort 1 and 22 from Cohort 2. We then made an additional distinction based on the number of crossovers on each list—children that should have been randomly assigned to TN-VPK or not, but instead ended up in the other group. From the 23 randomized applicant lists with good consent rates, we selected the lists with no more than 20% of the children as crossovers. This left us with 11 lists, all from Cohort 2, that included 136 children who participated in TN-VPK and 88 who did not.

The randomization to TN-VPK participation or not for the children on these 11 randomized applicant lists was clearly not perfect, nor can we assume that these children are fully representative of those in the full ISS analysis sample. Nonetheless, examining the TN-VPK effect estimates for these more completely randomized children provides a check on the plausibility of the estimates that resulted from the analyses reported above that relied more heavily on statistical control techniques. We therefore estimated the TN-VPK effects on the WJ Composite scale for the children on these 11 lists using several different techniques and compared the resulting effect estimates with that from the analysis approach reported above. Because all 11 lists came from Cohort 2, we used the analysis model for the full ISS analysis sample to estimate only the Cohort 2 effect size, which was a statistically significant .24, as our basis for comparison.

The statistical analyses we applied to generate effect size estimates from the data for the children on the 11 selected lists included:

1. A quasi intent-to-treat (ITT) analysis with crossovers included in the intended TN-VPK condition rather than in the condition they actually experienced. We refer to this as a quasi ITT analysis because the less than 100% consent rate meant that not all the randomized cases were accounted for. This estimate used only the covariates without a propensity score and without sample weights.
2. An estimate comparing actual TN-VPK participants and nonparticipants using covariates that included a propensity score created for this sample from those covariates and without sample weights.
3. An estimate from the same model as number 2 above, but including sampling weights; that is, an estimate comparing actual TN-VPK participants and nonparticipants using covariates that included the propensity score created for this sample from those covariates and using sample weights.

4. An estimate of treatment-on-the treated (TOT) effects comparing actual TN-VPK participants and nonparticipants using instrumental variable techniques to produce a complier average causal effect (CACE) estimate (Angrist, 2006; Bloom, Zhu, & Unlu, 2010).

The effect size estimates resulting from each of these analyses are reported in Appendix E5 below (all based on the full Cohort 2 sample weighted pooled standard deviations for comparability). The full results from each analysis are provided after that. Despite the much smaller sample size provided by these 11 randomized applicant lists, all these effect estimates were statistically significant and consistent with each other and with the estimate from the full ISS analysis sample. This consistency gives us confidence that our statistical models, despite their complexity, are providing credible estimates of TN-VPK effects.

E5. Effect Estimates for the WJ Composite Measure from Sensitivity Analysis Using Cases from the 11 Most Fully Implemented Randomized Applicant Lists

ANALYSIS APPROACH	N in TN-VPK Group	N in No TN-VPK Group	TN-VPK Effect Estimate in W-Score Units	*p*-value	Effect Size
Full Cohort 2 Sample Estimate – Propensity Score with weights	539	232	3.87**	.000	.24
Quasi Intent-to-Treat Identification	136	86	3.67**	.000	.23
Propensity Score w/o weights	136	86	4.04**	.000	.25
Propensity Score with weights	136	86	3.40**	.000	.21
Instrumental variable CACE	136	86	3.94**	.000	.25

**p<.01. Effect sizes were calculated based on the weighted pooled standard deviation in the cohort 2 sample.

E6: Baseline Covariate Balance for the Data from the 11 Most Fully Randomized Lists (Effect Sizes)

	TOT					ITT
COVARIATE	**Without propensity score**	**Logit as covariate**	**Probability as covariate**	**Proportional sample weights + Logit as covariate**	**Proportional sample weights + Probability as covariate**	**Without propensity score**
Native English Speaker	0.04	0.03	-0.05	0.04	0.03	0.01
Black	-0.02	0.08	0.04	0.10	0.09	-0.23
Hispanic	-0.05	0.07	0.04	0.06	-0.02	0.04
Male	0.00	0.02	0.00	0.11	0.11	-0.08
Library Card Use	0.06	0.06	-0.01	-0.03	-0.07	0.10
Newspaper Subscriptions	0.01	0.12	0.00	-0.10	-0.15	-0.03
Magazine Subscriptions	0.00	0.07	0.00	-0.07	-0.09	-0.02
Mother's Education	0.15	0.03	-0.02	0.07	0.01	-0.03
N of Working Parents	0.07	0.07	-0.01	0.23	0.22	-0.03
Age at Pretest	0.08	0.11	-0.01	0.02	-0.03	0.10
Test Lag	-0.47	-0.21	0.01	-0.20	-0.06	-0.51
Test Interval	-0.25	0.22	0.09	0.26	0.16	-0.09
Letter-Word ID	0.14	0.07	-0.01	0.02	-0.05	0.18
Spelling	-0.11	0.07	0.02	0.02	-0.02	-0.04
Oral Comprehension	0.21	-0.03	-0.05	0.12	0.14	0.32
Picture Vocabulary	0.15	0.04	-0.04	0.05	0.02	0.10
Applied Problems	-0.02	0.00	0.01	0.03	0.05	-0.09
Quantitative Concepts	-0.12	0.08	0.06	-0.06	-0.10	-0.05
WJ_Composite Score	0.03	0.05	0.01	0.05	0.02	0.07

E7: Results from TOT Analyses of TN-VPK Effects on WJ Composite W-Scores from the 11 More Fully Randomized Lists

	TOT									
	Without propensity score		**Propensity score probability as covariate**		**Propensity score logit as covariate**		**Sample weighted, propensity score probability as covariate**		**Sample weighted, propensity score logit as covariate**	
EFFECT	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	89.8	<.0001	53.4	0.047	98.4	0.059	48.2	0.073	43.5	0.355
Inclusion Rate: Nonparticipant	-22.2	0.242	5.80	0.820	-28.0	0.430	14.8	0.576	12.6	0.728
Inclusion Rate: Participant	-29.7	0.195	-5.87	0.834	-34.9	0.329	3.32	0.909	2.61	0.944
Inclusion Rate Interaction	95.4	0.257	-16.7	0.879	118.5	0.423	-53.6	0.635	-42.5	0.778
Black	2.10	0.093	2.94	0.026	1.84	0.306	3.88	0.003	4.34	0.015
Hispanic	3.05	0.115	2.86	0.135	2.96	0.156	3.56	0.054	4.75	0.017
Male	0.57	0.536	0.53	0.560	0.55	0.550	0.82	0.358	0.95	0.292
Native English Speaker	0.90	0.669	0.58	0.780	0.81	0.701	-1.42	0.487	-0.51	0.807
Library Card Use	0.41	0.499	0.19	0.748	0.44	0.481	0.13	0.819	0.20	0.737
Newspaper Subscriptions	1.06	0.104	0.82	0.215	1.10	0.105	0.10	0.881	0.17	0.799
Magazine Subscriptions	1.03	0.342	1.37	0.207	0.94	0.415	2.71	0.016	2.69	0.022
Mother's Education	-1.56	0.039	-1.96	0.012	-1.46	0.114	-2.12	0.005	-2.05	0.014
Number of Working Parents	0.03	0.968	-0.23	0.769	0.08	0.925	-0.10	0.898	-0.13	0.873
Age at Pretest	-0.23	0.127	-0.33	0.036	-0.21	0.217	-0.45	0.004	-0.42	0.014
Test Lag	-0.11	0.008	0.02	0.772	-0.15	0.425	0.08	0.368	0.10	0.555
Test Interval	-0.06	0.125	0.06	0.423	-0.08	0.599	0.09	0.228	0.12	0.430
Pretest	0.90	<.0001	0.90	<.0001	0.90	<.0001	0.90	<.0001	0.90	<.0001
Propensity Score	NA		8.72	0.065	-0.24	0.851	12.79	0.005	1.77	0.155
TN-VPK Participation	**5.08**	**<.0001**	**4.04**	**0.003**	**5.09**	**<.0001**	**3.40**	**0.002**	**4.28**	**<.0001**
Effect Size	**0.32**		**0.26**		**0.32**		**0.21**		**0.27**	

Note: The effect sizes were calculated using the sample weighted pooled standard deviation from the Cohort 2 sample.

E8: Results from Instrumental Variable and ITT Analyses of TN-VPK Effects on WJ Composite W-Scores from the 11 More Fully Randomized Lists

	Complier Average Causal Effect				ITT		Cohort 2 Sample	
	IV Method without sample weights		IV Method with sample weights		Without propensity Score		Sample weighted, propensity score probability as covariate	
EFFECT	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	95.9	<.0001	109.5	<.0001	99.4	<.0001	85.6	<.0001
Inclusion Rate: Nonparticipant	-30.0	0.034	-31.4	0.019	-27.7	0.154	-0.97	0.632
Inclusion Rate: Participant	-34.8	0.074	-40.5	0.034	-32.5	0.163	-1.35	0.516
Inclusion Rate Interaction	123.3	0.082	133.2	0.050	115.2	0.180	-10.28	0.238
Black	2.11	0.114	2.43	0.061	2.27	0.077	1.90	0.009
Hispanic	2.94	0.138	3.34	0.085	3.04	0.123	2.26	0.025
Male	0.60	0.535	0.80	0.406	0.65	0.488	0.81	0.123
Native English Speaker	-0.03	0.990	-2.71	0.189	0.93	0.663	-1.31	0.207
Library Card Use	0.51	0.428	0.48	0.439	0.43	0.489	-0.01	0.980
Newspaper Subscriptions	1.13	0.105	0.26	0.700	1.12	0.093	0.28	0.440
Magazine Subscriptions	0.84	0.463	1.93	0.105	1.02	0.357	0.39	0.487
Mother's Education	-1.19	0.126	-0.85	0.247	-1.30	0.088	-0.14	0.706
Number of Working Parents	0.24	0.770	0.57	0.476	0.24	0.763	0.30	0.493
Age at Pretest	-0.18	0.251	-0.24	0.125	-0.19	0.194	-0.08	0.324
Test Lag	-0.15	0.000	-0.17	<.0001	-0.15	0.001	-0.10	0.001
Test Interval	-0.08	0.016	-0.11	0.001	-0.09	0.013	-0.02	0.453
Pretest	0.90	<.0001	0.89	<.0001	0.89	<.0001	0.86	<.0001
Propensity Score	NA		NA		NA		-1.00	0.613
TN-VPK Participation	**3.75**	**0.011**	**3.94**	**0.003**	**3.67**	**0.003**	**3.87**	**<.0001**
Effect Size	**0.24**		**0.25**		**0.23**		**0.24**	

Note: The effect sizes were calculated using the sampling weighted pooled standard deviation from Cohort 2 sample.

E9: TOT, IV, and ITT Analysis of the Cooper-Farran Social Behavior Teacher Ratings from the 11 Most Fully Randomized Lists

	TOT		**Complier Average Causal Effect**				**ITT**	
	Sample weighted, propensity score probability as covariate		**IV Method without sample weights**		**IV Method with sample weights**		**Without propensity Score**	
EFFECT	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	2.85	0.212	1.75	0.486	1.94	0.429	1.77	0.462
Inclusion Rate: Nonparticipants	-7.10	0.002	-0.60	0.746	-1.61	0.374	-0.60	0.735
Inclusion Rate: Participants	-5.03	0.040	-1.68	0.496	-2.40	0.339	-1.68	0.476
Inclusion Rate Interaction	23.6	0.022	-0.36	0.969	2.64	0.768	-0.36	0.967
Black	0.00	0.982	-0.05	0.810	-0.01	0.964	-0.05	0.802
Hispanic	0.71	0.013	0.44	0.146	0.68	0.028	0.44	0.128
Male	0.00	0.980	-0.23	0.128	-0.01	0.952	-0.23	0.111
Native English Speaker	0.00	0.996	-0.37	0.251	-0.11	0.745	-0.37	0.230
Library Card Use	0.00	0.969	0.04	0.690	-0.06	0.563	0.04	0.676
Newspaper Subscriptions	-0.20	0.052	-0.04	0.732	-0.22	0.051	-0.04	0.720
Magazine Subscriptions	0.21	0.256	0.33	0.079	0.28	0.162	0.33	0.066
Mother's Education	0.08	0.504	0.00	0.974	0.04	0.722	0.00	0.973
Number of Working Parents	-0.12	0.273	-0.06	0.624	-0.15	0.216	-0.06	0.608
Age at Pretest	-0.02	0.288	0.00	0.995	-0.03	0.227	0.00	0.994
Test Lag	0.00	0.145	0.00	0.393	0.00	0.227	0.00	0.372
Pretest	0.01	0.025	0.01	0.115	0.01	0.043	0.01	0.099
Propensity Score	-1.64	0.001	NA		NA		NA	
TN-VPK Participation	**0.32**	**0.027**	**0.17**	**0.390**	**0.41**	**0.036**	**0.14**	**0.368**
Effect Size	**0.30**		**0.16**		**0.39**		**0.13**	

Note: The effect sizes were calculated using the sampling weighted pooled standard deviation from the 11 Lists.

E10. TOT, IV, and ITT Analysis of the Cooper-Farran Work-Related Skills Teacher Ratings from the 11 Most Fully Randomized Lists

	TOT		**Complier Average Causal Effect**				**ITT**	
	Sample weighted, propensity score probability as covariate		**IV Method without sample weights**		**IV Method with sample weights**		**Without propensity Score**	
Effect	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	-10.2	<.0001	-12.8	<.0001	-11.1	<.0001	-12.8	<.0001
Interaction Rate: Nonparticipants	-8.14	0.001	-2.01	0.274	-3.37	0.060	-1.96	0.335
Interaction Rate: Participants	-5.14	0.035	-1.23	0.614	-2.83	0.254	-1.15	0.645
Inclusion Rate Interaction	28.7	0.005	4.95	0.582	10.4	0.241	4.73	0.606
Black	0.00	0.991	-0.02	0.904	0.01	0.955	-0.02	0.905
Hispanic	0.32	0.256	0.13	0.673	0.29	0.344	0.13	0.652
Male	-0.09	0.555	-0.24	0.108	-0.10	0.536	-0.24	0.092
Native English Speaker	-0.69	0.023	-0.98	0.003	-0.78	0.015	-0.96	0.006
Library Card Use	0.00	0.975	0.04	0.670	-0.05	0.596	0.04	0.692
Newspaper Subscriptions	-0.01	0.952	0.05	0.632	-0.02	0.826	0.05	0.623
Magazine Subscriptions	-0.04	0.820	0.12	0.531	0.03	0.893	0.12	0.502
Mother's Education	0.09	0.463	-0.10	0.400	0.07	0.592	-0.10	0.384
Number of Working Parents	0.01	0.952	0.04	0.753	-0.01	0.930	0.04	0.750
Age at Pretest	-0.02	0.470	0.00	0.994	-0.02	0.388	0.00	0.986
Test Lag	0.01	0.009	0.01	0.038	0.01	0.019	0.00	0.041
Pretest	0.04	<.0001	0.04	<.0001	0.04	<.0001	0.04	<.0001
Propensity Score	-1.40	0.005	NA		NA		NA	
TN-VPK Participation	**0.36**	**0.011**	**0.30**	**0.120**	**0.50**	**0.010**	**0.26**	**0.106**
Effect Size	**0.32**		**0.26**		**0.44**		**0.22**	

Note: The effect sizes were calculated using the sampling weighted pooled standard deviation from the 11 Lists.

E11: TOT, IV, and ITT Analysis of the ACBR Preparedness Scale Teacher Ratings from the 11 Most Fully Randomized Lists

	TOT		Complier Average Causal Effect				ITT	
	Sample weighted, propensity score probability as covariate		IV Method without sample weights		IV Method with sample weights		Without propensity Score	
EFFECT	b	*p*-value	b	*p*-value	b	*p*-value	b	*p*-value
Intercept	-17.3	<.0001	-18.8	<.0001	-17.9	<.0001	-18.7	<.0001
Inclusion Rate: Nonparticipant	-6.93	0.010	-3.93	0.067	-3.63	0.078	-3.93	0.114
Inclusion Rate: Participant	-3.87	0.172	-2.25	0.430	-2.23	0.433	-2.21	0.446
Inclusion Rate Interaction	23.7	0.047	14.3	0.174	10.9	0.286	14.3	0.189
Black	-0.12	0.605	0.16	0.510	-0.08	0.729	0.16	0.498
Hispanic	0.55	0.099	0.32	0.362	0.51	0.144	0.32	0.347
Male	0.11	0.501	-0.03	0.859	0.11	0.535	-0.03	0.852
Native English Speaker	-0.27	0.439	-0.51	0.171	-0.35	0.336	-0.49	0.208
Library Card Use	0.01	0.938	-0.03	0.816	-0.03	0.777	-0.03	0.804
Newspaper Subscriptions	0.08	0.502	-0.04	0.746	0.06	0.651	-0.04	0.734
Magazine Subscriptions	-0.03	0.902	0.20	0.342	0.03	0.890	0.21	0.315
Mother's Education	0.07	0.613	-0.01	0.971	0.08	0.588	-0.01	0.965
Number of Working Parents	-0.08	0.569	-0.11	0.428	-0.07	0.589	-0.11	0.402
Age at Pretest	-0.04	0.187	-0.02	0.511	-0.04	0.178	-0.02	0.495
Test Lag	0.01	0.004	0.01	0.008	0.01	0.005	0.01	0.008
Pretest	0.06	<.0001	0.06	<.0001	0.06	<.0001	0.06	<.0001
Propensity Score	-0.93	0.108	NA		NA		NA	
TN-VPK Participation	**0.46**	**0.006**	**0.42**	**0.062**	**0.60**	**0.008**	**0.35**	**0.055**
Effect Size	**0.34**		**0.30**		**0.44**		**0.26**	

Note: The effect sizes were calculated using the sample weighted pooled standard deviation from the 11 Lists.