

Effects of a State Prekindergarten Program on Children's Achievement and Behavior through Third Grade

Working Paper

Peabody Research Institute

January 2016

**Effects of a State Prekindergarten Program on Children's Achievement
and Behavior through Third Grade**

Mark W. Lipsey and Dale C. Farran

Vanderbilt University

Kerry G. Hofer

Abt Associates

Abstract

Many states have implemented or expanded state-funded prekindergarten programs in the last decade, encouraged by claims about the benefits that can be expected and the importance of early experiences for children's development, especially for economically disadvantaged children. However, there is remarkably little methodologically adequate evidence about the effects of such programs. Using a subsample of children with parental consent from a larger sample of children randomly assigned to attend the Tennessee pre-k program or not, this study examined effects on cognitive and noncognitive outcomes through third grade. At the end of the pre-k year, program participants showed better outcomes than comparable nonparticipants on achievement measures and ratings of school readiness by kindergarten teachers. But those effects were not sustained in subsequent years and, indeed, by the end of third grade the pre-k participants scored lower on the achievement measures than nonparticipants. These results raise questions about the way state pre-k programs have been designed and implemented.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

INTRODUCTION

As state policymakers consider instituting, expanding, or altering their prekindergarten programs, the advice available to them from various professional and advocacy groups is heavily infused with claims about the substantial benefits that can be expected from such programs (e.g., Barnett, 2013; ReadyNation, n.d.; SREB, 2015). However, that advice has not typically relied on current research evidence about the effects of pre-k programs implemented at statewide scale. Instead, the rationale for state expansion of pre-k programs is typically based on the “widely advertised success of a few model programs” (Fitzpatrick, 2008, p. 1) combined with recognition of the importance of early experiences for children, especially children growing up in poverty.

Importance of Early Experiences for Children from Low Income Families

Poverty in the United States creates pernicious environments for the development of young children, beginning in utero. Experiences of poverty before age 5, especially, have both immediate and long lasting consequences for children’s academic achievement and behavior (Currie & Rossin-Slater, 2014; Duncan, Ziol-Guest, & Kalil, 2010). Summarizing longitudinal studies, Almond and Currie (2010) concluded that characteristics of children at age 7 explain much of the variation in their later educational achievement and even subsequent earnings and employment. These realizations have fueled the push for intervening with poor children before school entry in an attempt to remediate the adverse effects and alter the lifelong trajectories of children from low-income families.

Recognition that poverty produces an early educational disadvantage that persists throughout the school years is not a new insight. The link between educational achievement and poverty has been acknowledged at least since the 1960s when President

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Johnson began the war on poverty (Farran, 2007). That recognition motivated the creation of Head Start in 1964 with its focus on school readiness skills. Thus the United States has had fifty years' experience creating interventions prior to formal school entry for children whose families live in poverty. Despite these efforts, the achievement gap between children in poverty and higher income children has grown in recent years (Reardon, 2011).

Model Programs

The early childhood intervention programs that provide the models and rationale for expansion of public pre-k began in the 1960s and were set up as experiments that focused on IQ as the target outcome. They demonstrated immediate and significant effects on IQ measures (Lazar et al., 1982), but those positive effects dissipated by the end of 6th grade and sometimes earlier. The effects of these programs on academic achievement also persisted for some years, but then generally faded as well (Campbell et al., 2001; Schweinhart et al., 2005). The two experimental programs whose participants have been followed the longest are the Perry Preschool and Abecedarian programs. It is their long-term effects on school completion, employment, marriage stability, criminal behavior, and the like that are most often cited as the justification for further public investments in pre-k and as the basis for the claim that the value of the benefits will outweigh the costs. The process by which these preschool programs influenced such long-term life outcomes despite their lack of sustained effects on cognitive measures is somewhat ambiguous and does not have a strong empirical base. The most fully developed theory is that these interventions enhanced children's noncognitive skills (e.g., inhibition of externalizing behavior, self-regulation, and academic motivation) in ways that had positive cumulative

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

effects over subsequent years (Heckman, Pinto, & Savelyev, 2013).

With the publication in *Science* of Heckman's 2006 call for investments in early childhood education for disadvantaged children, momentum increased dramatically within states for policymakers to create or expand publicly funded pre-k programs. Heckman (2008) based his conclusions about the benefits of such investments on analyses of the Perry Preschool program and more recent studies of the Chicago Child Parent Center (CPC) program (Reynolds et al., 2011). However, those programs included elements difficult to duplicate at statewide scale, and no state pre-k program has actually replicated the model programs on which Heckman's analyses were based. The CPC program, for example, extended through several years in elementary school and required substantial parent involvement. Abecedarian began when children were 6 weeks old, continued until kindergarten, and provided full day care for 50 weeks of the year. The Perry Preschool program targeted African American children with low IQ, enrolled children for two years, and provided weekly home visits. Moreover, the cost of those programs would be more than any public program currently allocates for pre-k. In today's dollars, the cost per child per year to implement the Perry Preschool program has been estimated at \$20,000, and at \$16,000- \$40,000 for Abecedarian (Minervino & Pianta, 2014). A critical question for state pre-k policy, therefore, is whether programs with weaker components and constrained budgets implemented at scale can deliver the benefits expected of them (Baker, 2011).

Evaluations of State Pre-K Programs

Reliance by state policymakers on generalizations from the longitudinal findings for the widely cited model programs is to some extent understandable given the inadequate evidence available about the effectiveness of current statewide pre-k programs (Duncan &

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Magnuson, 2013). Prior to the study of the Tennessee program presented here, there has been only one well-controlled longitudinal study of a scaled-up publicly-funded pre-k program—the Head Start Impact Study—and that was for a national, not a state program.

The Head Start Impact Study began in 2002 and involved 84 grantee programs and 5,000 children randomly assigned to receive an offer of admission or not (Puma et al., 2012). The children in the 4-year old cohort admitted to Head Start made greater gains across the pre-k year than nonparticipating children on many of the cognitive measures of language and literacy achievement but none of the math measures. However, by the end of kindergarten the control children had caught up, erasing the differences between the two groups. Subsequent positive effects were found on only one achievement measure at the end of 1st grade and another at the end of 3rd grade. On the noncognitive social-emotional measures of the sort hypothesized by Heckman and colleagues to be mediators for the long-term effects of the model programs (Heckman, Pinto, & Savelyev, 2013), there were no statistically significant effects at the end of pre-k or kindergarten. A few positive effects appeared in parent reports at the end of the 1st and 3rd grade years, but teacher and child reports in those years showed either null or negative effects.

Research specifically on the effects of state pre-k programs has been far less rigorous (Farran & Lipsey, in press; Gilliam & Zigler, 2001). The strongest design used for assessing effects at the end of the pre-k year is the age-cutoff regression discontinuity design (RDD). The Gormley et al. (2005) evaluation of the Tulsa pre-k program was the first to use this RDD. Since then, a number of studies have applied the age-cutoff RDD to statewide pre-k programs (e.g., Wong et al., 2008). There are potentially problematic methodological issues inherent in this design (Lipsey et al., 2015), but it is nonetheless less

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

vulnerable to bias than many other non-experimental alternatives.

The age-cutoff RDD studies have attended almost exclusively to language, literacy, and math achievement outcomes and, with very few exceptions, have reported positive effects. Only the Peisner-Feinberg et al. (2014) RDD study of the Georgia universal pre-k program examined noncognitive outcomes in addition to achievement, finding a significant effect on social awareness but not on social skills or problem behavior. Though not involving a statewide program, an age-cutoff RDD of the publicly funded Boston pre-k program (Weiland & Yoshikawa, 2013) also found positive effects on noncognitive outcomes (e.g., executive functioning) at the end of pre-k along with achievement effects.

A limitation of the age-cutoff RDD is that it does not allow for longitudinal follow up—the control group completes pre-k within a year and no longer provides an informative comparison. Studies of the extent to which the effects of state pre-k programs are sustained past the end of the pre-k year have relied on notably weak designs. The largest group of these studies use data obtained after the pre-k year has ended to construct crudely matched samples of children who did and did not previously attend pre-k and then compare their outcomes in later grades. In a typical example, Barnett et al. (2013) identified children from kindergarten classrooms who had and had not attended the New Jersey pre-k program the year before, matched only on age, gender, race, and eligibility for free or reduced price lunch, and followed them through 5th grade. No baseline performance or family variables are available in these designs to assess initial group equivalence or to use as statistical controls. In particular, the groups are inherently different on whatever motivation, value for education, aspirations for their children, and other such characteristics that led one group of parents, but not the other, to enroll their

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

children in the pre-k program. It is not surprising that the results of these comparisons overwhelmingly favor the children who attended the state pre-k programs.

The only other studies that investigate longer-term effects of state pre-k programs are a few difference in difference (DD) studies that examine before and after differences in state or county level student outcomes as a pre-k program is rolled out compared to differences over a comparable period for another area in which there was no analogous pre-k expansion. The challenge for these studies is to isolate the difference made in the target outcomes by pre-k implementation from the other influential factors occurring over the same time in the same locations. Fitzpatrick (2008), for example, used a DD design to investigate the effects of the Georgia universal pre-k program that grew from 14% participation in 1995 to 55% in 2008. Initial analyses indicated positive pre-k effects on 4th grade NAEP reading and math scores, but further analyses exploring control group variants and different inference models did not yield completely robust conclusions. Similar effects that were generally positive but sensitive to the selection of comparison states were found in the Cascio and Schanzenbach (2013) DD study of the Georgia and Oklahoma pre-k programs. By contrast, however, DD analyses of the More at Four pre-k program in North Carolina showed effects on 3rd grade state achievement scores that were robust to a range of model variations (Ladd, Muschkin, & Dodge, 2014).

The difficulty of drawing firm conclusions from DD analyses in the dynamic context of state pre-k expansion is illustrated by an ambitious study conducted by Rosinsky (2014). She compared the 2007, 2009, and 2011 4th grade NAEP math scores across multiple states to program enrollment six years previously in Head Start, state-funded pre-k, and special education preschools. Surprisingly she found a negative association between

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

enrollment in public pre-k and NAEP math scores, with the largest negative effect associated with state-funded programs.

Summary

The widespread advocacy for expansion of state pre-k programs and the high expectations for their benefits for disadvantaged children are based chiefly on the dramatic long term effects found for a few small, intensive model programs implemented long ago. Research on the actual effects of contemporary state funded pre-k programs, by contrast, is most notable for the prevalence of weak designs that do not support confident causal inference. The most convincing results from those studies show positive effects at the end of the pre-k year on cognitive outcomes, mainly language, literacy, and math achievement measures. Noncognitive social-emotional outcomes have only rarely been examined in these studies, but some positive effects on those have been reported as well.

The quality of the evidence about whether those effects are sustained past the end of the pre-k year, however, is especially poor. A number of studies using post hoc matched designs report finding pre-k effects on achievement and related academic outcomes well into elementary and even middle school grades. But these designs are so vulnerable to rather obvious sources of selection bias that their findings are not credible. More promising are the few difference in difference studies of the aggregate effects on achievement test scores as state pre-k programs have expanded. The assumptions on which their analytic models rest are impossible to verify, however, and their results are not impressively robust. Especially notable in these longer-term studies of state pre-k programs is the near total absence of evidence about effects on the kinds of noncognitive outcomes that have been hypothesized to be the key mediators between pre-k and the

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

positive life outcomes found for the widely cited model programs.

The Head Start Impact Study looms large in this context as the only randomized study of a publicly-funded scaled-up pre-k program. The findings of positive effects on cognitive outcomes for the 4-year old cohort at the end of the pre-k year are consistent with those from the analogous but less well controlled studies of state programs. And the null effects on noncognitive outcomes at the end of the pre-k year are not inconsistent given how seldom such outcomes have been examined in state pre-k studies. The rapid fade out of the achievement effects and the emergence of negative effects on many of the noncognitive measures by 3rd grade, however, do not offer encouragement for the prospects of sustained effects from scaled-up state-funded programs.

The research study presented here investigates the effects of the statewide and state-funded Tennessee pre-k program on cognitive and noncognitive outcomes through 3rd grade. It uses a subgroup of children with parental consent who participated in a random assignment design analogous to that used in the Head Start Impact Study. Because it was necessary to seek consent after randomization, and consent rates were modest, selection bias was a threat. However, the pre-k participants and nonparticipants were quite comparable on a wide array of baseline variables that, additionally, were used as statistical controls in the analyses via propensity scores. Also, by the nature of the design, both groups of children were from families that had attempted to enroll them in the pre-k program, creating further comparability on a range of potentially important unobserved variables. As such, this study uses a better controlled design than has appeared heretofore to investigate the short- and medium-term outcomes of a state pre-k program.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

THE TENNESSEE VOLUNTARY PREKINDERGARTEN PROGRAM

The Tennessee Voluntary Prekindergarten program (TN-VPK) is a state funded pre-k program offered to the neediest children in Tennessee. By statute, eligibility is restricted to children who qualify for the federal free or reduced price lunch program (FRPL), followed by such other at-risk children as those with disabilities or English Language Learners as space allows. TN-VPK is a full school-day program that operates on the same calendar as the rest of the public school system, requires a licensed teacher and aide in every classroom, a maximum of 20 children per class, and a curriculum chosen from a state-approved list. According to the quality standards promulgated by the National Institute for Early Education Research (NIEER), the TN-VPK program is among the top state pre-k programs, meeting 9 of the 10 NIEER benchmarks (Barnett et al., 2014).

All funds flow through Local Education Agencies, and the current annual investment of nearly \$90 million supports 935 classrooms in 135 of the 136 school districts across all 95 counties in Tennessee. All but 62 (6.6%) of these classrooms are located in public schools, though funding for the sites not in public schools is still administered by the local education agency. From its pilot year in 2004, the program has grown from serving 3,000 children to more than 18,000 as of fiscal year 2014. Despite that growth, the program enrolls fewer than half of the eligible children in the state (Grehan et al., 2011) and many school systems in the state receive more eligible applicants than they can accommodate.

In 2009 the Peabody Research Institute at Vanderbilt University launched a study of the TN-VPK program in coordination with the Division of School Readiness and Early Learning at the Tennessee Department of Education. That study has multiple components; this report describes the findings of one of those components that investigated the

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

following research questions:

1. Does participation in TN-VPK improve the school readiness at kindergarten entry of the economically disadvantaged children served?
2. Does TN-VPK have differential effects for different subgroups of children and, if so, what are the characteristics of the children who show larger or smaller effects of TN-VPK participation?
3. Are the effects of TN-VPK participation sustained through the kindergarten, 1st, 2nd, and 3rd grade years?

METHODS

This study is part of a larger TN-VPK evaluation comprised of two components: a randomized control trial (RCT) implemented in selected oversubscribed sites and a regression discontinuity design applied to a representative sample of TN-VPK classrooms across Tennessee. The RCT, in turn, consists of two overlapping parts. The *full sample* of participants in the RCT involves more than 3,000 children randomly assigned to receive an offer of admission to TN-VPK or not. These children are being followed in the state's education database with attention to such outcomes as attendance, retention in grade, special education placements, disciplinary actions, and state achievement test scores. Complete data for this sample through 3rd grade are not yet available, but results will be reported in another paper when they are. The present report describes the findings for an *intensive substudy sample* that consists of 1076 of the children in the full sample for whom parental consent was obtained for annual assessments through their 3rd grade year. Prior research reports have more fully described the components of the overall study and presented findings from earlier waves of data collection (Lipsey et al., 2011, 2013a, 2013b).

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Random Assignment

Many TN-VPK sites across the state receive more eligible applicants than they can accommodate, creating a situation in which some applicants must be denied admission out of necessity. For school year 2009-10 and again in 2010-11, the personnel in a number of those sites agreed to randomly select the applicants to whom they would offer admission rather than use their customary procedures. These programs sent their applicant lists to the research team where each list was sorted into random order and promptly returned. The school staff were asked to fill their TN-VPK seats in the order that children appeared on the randomized list by attempting to contact a child's parents at least three times on different days of the week and times of the day to offer admission. If they were unable to contact the parent after these attempts or the parent declined the offer, they could then move on to the next child on the randomized list whose parents had not yet been contacted.

Once all the slots in a given program were filled, the children remaining on the list who were not offered admission were identified as the waiting list. If a child who had been offered admission did not show up for the program when school started, the next child in order on the waiting list was offered that place. Any children not offered admission after that point became the control group of TN-VPK nonparticipants. Note that this procedure produces a randomized block design in which each applicant list is a block with its own randomly assigned treatment and control groups.

Intensive Substudy Sample

Attempts were made to contact parents of children on randomized applicant lists at the beginning of the school year to request consent for periodic individual assessments of their children. Though very few parents explicitly refused, making contact and obtaining a

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

response from the parents proved challenging. For the 2009-10 cohort, State Department of Education officials interpreted the confidentiality requirements for children eligible for free or reduced price lunch in a way that only allowed parents to be contacted through a mailing sent centrally from the Department of Education. For that cohort, the consent rate was 24%. The 2010-11 cohort was then added to the study and arrangements were negotiated to allow parents to be approached about consent as an adjunct to the TN-VPK application process in sites willing to accommodate this procedure. The consent rate for this second cohort was 68%; the overall consent rate for both cohorts combined was 42%.¹

However, not all of the 1331 consented children resulting from this procedure were eligible for the intensive substudy sample (ISS). We restricted that sample to children who were age-eligible for kindergarten the next year, were income-eligible for TN-VPK (qualified for FRPL), and who had not applied for the sole purpose of receiving out-of-classroom special education services. We further restricted the sample to children who had applied to schools for which there were consented children in both the TN-VPK participant and nonparticipant groups and for whom useable assessment data were available at the end of the pre-k year. These procedures resulted in a total of 1076 children who were represented on 76 randomized applicant lists created at 58 schools in 21 districts spread widely across the state and representing urban, suburban, and rural areas.

Identification of the children in the ISS analysis sample who participated in TN-VPK and those who did not was based on records in the State Education Information System showing enrollment status plus information provided by parents, teachers, and school

¹ These consent rates are computed as a percentage of the number of children in each cohort of the full sample and differ somewhat from those we have reported before, which were computed as the average across the randomized applicant lists.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

personnel gathered during our data collection. We then defined TN-VPK participants as children for whom the available information documented attendance at a TN-VPK program for at least 20 days during the school year, the minimum number of days required by the Department of Education to consider a child enrolled in TN-VPK. TN-VPK nonparticipants, conversely, were defined as children for whom available information indicated that they had not attended any TN-VPK program or, if they attended, it was for fewer than 20 days.² By this definition, the ISS sample included 773 TN-VPK participants and 303 nonparticipants. Within the 76 randomized applicant lists that provided this final sample, the consent rate was 63% for participants, 45% for the nonparticipants, and 56% overall.

TN-VPK participants attended pre-k classes an average of 159 days ($SD=22.5$) during the school year. For the nonparticipants, parent interviews identified the alternative arrangements parents made for their children during the pre-k year. A majority of these children did not attend any center-based preschool program after they were not admitted to the TN-VPK program. A little more than 59% were cared for at home, 11.5% attended Head Start, 15.1% were in private childcare, and the child care arrangements for the remainder were mixed or not reported.

To assess the effects of TN-VPK, outcomes were compared for the participant and nonparticipant groups described above; that is, the effects of treatment-on-the-treated were estimated. The modest consent rates for participation in the ISS produced considerable attrition from the intent-to-treat groups created on the randomized applicant lists for the full sample. Moreover, the different consent rates for the TN-VPK participants

² There were only three children included in the nonparticipant group with TN-VPK attendance of less than 20 days; dropping them has no consequential effect on the results of any of the analyses reported here.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

and nonparticipants meant there was potentially biasing differential attrition that would have to be addressed in the analyses. Given these unavoidable compromises to the randomization, we elected to treat the ISS as a quasi-experiment and compare outcomes for TN-VPK participants and nonparticipants irrespective of the conditions to which they were assigned on the original randomized applicant lists. Nonetheless, of the 1076 children in the ISS sample, 86% are in the respective participant or nonparticipant groups to which they were randomly assigned. The remaining 14% included 76 (25.1%) of the 303 children assigned to the control group of nonparticipants who were nonetheless admitted to TN-VPK, mainly in place of children who were supposed to be admitted but could not be reached by school personnel. Conversely, 76 (9.8%) of the 773 children assigned to receive offers of admission did not end up actually participating in TN-VPK.

Data Collection

Children in the ISS were individually assessed by trained research staff in the fall and spring of their pre-k year. TN-VPK participants were assessed in their schools and nonparticipants were assessed at a location convenient for the parents, e.g., Head Start centers, libraries, parks, and homes. Children in both groups were assessed in the spring of each subsequent year through the 3rd grade year whether or not they stayed in the same school or district. Early in the kindergarten year and in the spring of the 1st, 2nd, and 3rd grade years, children's classroom behaviors were also rated by their teachers. The ratings by the kindergarten teachers near the beginning of the kindergarten year were treated as pre-k outcomes reflecting the school readiness of the children upon entry into formal schooling. The retention rate for the ISS was at least 92% for each of the four years following the pre-k year, and the modest amount of attrition that did occur was similar for

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

TN-VPK participants and nonparticipants. Table 1 shows the number and proportion of children who received direct assessments each year.

Measures

Parent Questionnaire

During the pre-k year, parents of consented children were interviewed via telephone about the alternate arrangements made if their child was not in TN-VPK, their own education and employment and that of their spouse/partner, and the home language and literacy environment. When needed, these interviews were conducted by Spanish-speaking interviewers.

Direct Assessments

Children's academic achievement was assessed with a selection of scales from the Woodcock Johnson III Achievement Battery (WJ; Woodcock, McGrew, & Mather, 2001). The scales administered at the beginning and end of the pre-k year included two measures of literacy (Letter-Word Identification and Spelling), two measures of language (Oral Comprehension and Picture Vocabulary), and two measures of math skills (Applied Problems and Quantitative Concepts). At the end of the kindergarten year, and each subsequent year through the 3rd grade year, two additional scales were added: another language (Passage Comprehension) and math measure (Calculation). These scales were all administered in English, which was the language of instruction in all the pre-k classrooms.

Letter-Word Identification measures children's ability to identify and pronounce letters and words. The Spelling subtest measures children's ability to draw simple shapes and write orally presented letters and words. Oral Comprehension measures children's ability to listen to and provide a missing key word to an orally presented passage. Picture

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Vocabulary tests children's expressive vocabulary. Applied Problems measures children's ability to solve numerical and spatial problems accompanied by pictures. Quantitative Concepts measures children's understanding of number identification, sequencing, shapes, and symbols and, in a separate section, ability to manipulate the number line. Passage Comprehension assesses reading comprehension through matching picture or text representations with similar semantic properties. Calculation assesses math computation skills through the completion of visually-presented numeric problems.

These WJ scales were moderately to highly intercorrelated; to provide summary achievement indices, composite scores were created as the mean across the individual scales. One composite score combined the original six subscales administered from the beginning of pre-k (WJ Composite6). Another combined those original six subscales with the two first administered at the end of the kindergarten year (WJ Composite8).

Teacher Ratings

Two teacher rating instruments were completed by kindergarten, 1st, 2nd, and 3rd grade teachers. The Cooper-Farran Behavioral Rating Scales (Cooper & Farran, 1991) required teachers to rate each child's work-related and interpersonal skills. The Work-Related Skills scale assesses ability to work independently, listen to the teacher, remember and comply with instructions, complete tasks, and otherwise engage appropriately in classroom activities. The Interpersonal Skills scale assesses social interactions with peers including appropriate behavior in group activities, play, and outdoor games; expression of feelings and ideas; and response to others' mistakes or misfortunes.

The second measure, the Academic Classroom and Behavior Record (ACBR; Farran, Bilbrey, & Lipsey, 2003), consisted of four scales. Readiness for Grade Level Work asked

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

how well prepared the child was for grade level work in literacy, language, and math skills as well as social behavior. Liking for School included items about 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 indicated whether the child has shown explosive or overactive behaviors, attention problems, physical or relational aggression, or social withdrawal or anxiety. Peer Relations items asked whether other children like the target child and how many close friends the child has.

Analysis

Missing Data

The mean missing value rate across all variables was 6.2% (range: 0.0% to 14.5%) for TN-VPK participants and 6.4% (range: 0.0% to 17.2%) for nonparticipants. To retain the full sample in all analyses, multiple imputation of the missing values was done separately for participant and nonparticipant data using Mistler's (2013) procedure for multilevel data. Three groups of related variables within each condition were separately imputed in a 2-level structure with children nested within their school-level randomized applicant lists. Fifty imputed files were produced and stacked for analysis with the results of those analyses then pooled so as to include the uncertainty associated with the imputations in the standard error estimates. These imputations produced a small number of missing value estimates that were outliers relative to the distribution of observed values. For continuous variables, imputed values falling outside Tukey's (1977) outer fence for the observed values were recoded to the respective outer fence. For integer values (e.g., ratings on a 7-point scale), imputed values falling outside the range from one scale step below the lowest observed value to one scale step above were recoded to those values. For

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

a small number of dichotomous variables to be used as moderators in interaction terms in the analysis (e.g., gender), imputed values were rounded to the nearest observed value.

Baseline Equivalence

The baseline variables for the ISS are shown in Table 2, some of which are differentiated in ways that overlap with others (e.g., Hispanic race/ethnicity is further divided into native and nonnative English speakers). Because the consent rates were different for the first and second cohorts of children, these baseline variables were first examined for differences between the cohorts using a multilevel analysis with children nested within randomized applicant lists and lists nested within school districts. Of the 22 variables on which the cohorts were compared, significant differences were found only for number of working parents, with a mean of 1.1 for the 2009-10 cohort and 1.3 for the 2010-11 cohort. Given this substantial baseline similarity between the cohorts, their data were combined for all subsequent analyses.

The results of an analogous analysis on the combined cohorts for baseline differences between the TN-VPK participants and nonparticipants are also shown in Table 2. These results demonstrated that participants and nonparticipants were substantially similar, but there were statistically significant differences on the WJ Letter-Word Identification scale and mother's education, both favoring the participant group, and a difference on the WJ Picture Vocabulary scale at $p < .10$. The effect sizes indexing the magnitude of the various baseline differences, nonetheless, were relatively modest—none greater than .19—and all fell under the Imbens and Rubin (2015, p. 277) rule of thumb of .25 for baseline differences too large to adjust with covariates in a regression model.

A more problematic difference between the TN-VPK participants and

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

nonparticipants resulted from the practicalities of arranging individual assessments for so many children under field conditions that made it difficult to obtain every assessment within the desired tight time windows at the beginning and end of each school year. This was especially the case for nonparticipants during the initial year when they were not in TN-VPK classrooms so that ad hoc arrangements had to be made with the parents to assess them at some other location. As a result, the timing of assessments was variable and, in particular, it was not possible to obtain baseline pretest assessments as early in the school year as desired. Table 3 shows the mean days from the date on which the respective TN-VPK classes began to the date on which each wave of assessments was administered. The variability in the timing is indicated by the standard deviations, and an unfortunately long average lag is evident before it was possible to obtain pretest assessments for both groups. Most notably, there were significant timing differences between the participants and nonparticipants during the early waves. In consideration of these differences, and the few lesser ones found for the child and family characteristics shown in Table 2, we constructed propensity scores to assist with the task of statistically matching the groups and reducing any bias in the effect estimates that might be caused by these initial differences.

Propensity Scores

The propensity scores were created via a multilevel logistic regression predicting treatment condition with children nested in their randomized applicant lists and lists nested within school district. The selection of predictor variables focused on the timing variables shown in Table 3, all of which were included. Moreover, because the rate of change may have been different for the TN-VPK participants and nonparticipants during the lag time prior to pretest, an interaction term was included for lag time crossed with

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

baseline scores on the WJ Composite⁶ achievement measure. Also included was a selection of the descriptive variables for children and families shown in Table 2 (age, gender, race/ethnic subgroup, home literacy index, mother's education, and number of working parents). In recognition of the varying consent rates across the randomized lists and the two cohorts, the propensity score model also included Level 2 variables for cohort and the proportions of TN-VPK participants and nonparticipants within each randomized list that were represented in the ISS, along with the interaction between the rates for those groups.

The propensity scores generated by this procedure overlapped completely between the participant and nonparticipant groups, providing a broad range of common support that required no trimming at the extremes. They also showed linear relationships with the composite achievement measures across the longitudinal waves, and we elected to use the propensity score variable as a covariate in the analyses estimating intervention effects. A check on the extent to which the propensity scores used in this manner reduced the baseline differences of concern was made by re-estimating those differences with the propensity scores as the sole covariate in the regression models. The last two columns of Tables 2 and 3 show the *p*-values and effect sizes that resulted with these propensity score adjustments. With the propensity score covariate in the model, there were no statistically significant differences on any baseline variable, and the corresponding propensity score-adjusted effect sizes were quite small with none exceeding .10 and most well below that.

RESULTS

TN-VPK Effects at the End of the Pre-K Year

The first research question this study addressed was whether TN-VPK improved the school readiness of the participating children over the course of the pre-k year. The

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

indicators of school readiness for this purpose were the WJ achievement measures of early literacy, language, and math skills and the ratings made by kindergarten teachers near the beginning of the kindergarten year. TN-VPK effects on the achievement measures were estimated in three level models with children nested in their randomized applicant lists and lists nested in districts. The propensity scores were used as a covariate along with the pretest of the respective outcome measure and a selection of baseline child and family characteristics. Table 4 shows the full analysis results for the WJ Composite6 outcome that characterizes the overall pattern of achievement effects.³ Table 5 provides additional detail about this finding and summarizes the results of analogous analyses for each of the individual WJ scales. As indicated there, the effects on all the measures except Oral Comprehension were statistically significant at the .05 level, and the *p*-value for Oral Comprehension fell under .10. Table 5 also shows the standardized mean difference effect sizes that correspond to the regression coefficients that estimate the difference between the posttest means for the TN-VPK participants and nonparticipants in WJ W-score units.

Standardized effect sizes are one way to characterize the magnitude of the TN-VPK effects, but they compare the participants and nonparticipants only on the posttest and, as such, provide no indication of the nature of the relative improvements by each group over the pre-k year. Table 5, therefore, also presents a variant on the effect size picture that is more informative. The covariate-adjusted pretest and posttest means for each group were extracted from the analysis; these involve the same covariates, other than the pretest itself,

³ The results presented here and in the sections below are somewhat different from those reported earlier in technical reports (Lipsey et al., 2011, 2013a, 2013b), though their pattern is much the same. These differences stem from improvements in the imputation procedure and refinements in the propensity scores and other aspects of the analytic models aimed at better controlling the influence of baseline differences, especially regarding timing of measurement.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

and thus are comparable. By standardizing those pre-post mean differences with the same pooled posttest standard deviation used for the conventional effect size index, differential growth as well as the posttest differences it produces can be depicted.

The last three columns of Table 5 show these effect sizes for pre-post gain. They reveal, first, that both groups of children showed performance improvements during the pre-k year, though the magnitude of the gains varied for the different achievement measures. The pre-post gains on the language measures, for instance, were smaller than those on the literacy and math measures. Relative to the gains by the nonparticipants, those of the TN-VPK participants were proportionately greater on all these measures, with increases ranging from 20% to 83%. However, one of the largest proportionate gains was made on a measure that did not improve very much for either group—Picture Vocabulary.

Another way to characterize the findings on achievement measures is to compare them with the results of other studies of pre-k effects. Summarizing the immediate academic effects of 84 pre-k programs, Duncan and Magnuson (2013) estimated the mean effect size at the end of the pre-k year as .35. However, that includes earlier studies going back to the 1960s; programs researched since the 1980s had an average effect size of .16. Thus the effects found for TN-VPK at the end of the pre-k year are as large as or larger than those typically found in other studies of a wide variety of pre-k programs.

Teacher Ratings

Kindergarten teachers in classrooms with ISS children were asked to rate those children near the beginning of the kindergarten year on the rating scales described earlier. No information was provided to the teachers about which children had participated in TN-VPK. These ratings were requested a few weeks past the start of the school year, lagged so

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

the teachers would have a chance to become familiar with the children but not so much that the kindergarten experience itself was expected to have much effect on their behavior. Analysis of these ratings was analogous to that described above for achievement: multilevel models with children nested in their randomized applicant lists and lists nested in districts. The same covariates were used with two exceptions. The WJ Composite⁶ baseline achievement measure was used in place of pretests (there were no baseline teacher ratings). In addition, a variable representing the timing of the ratings was included, specifically the number of days between September 1 of the pre-k year and the date on which the kindergarten teacher completed the ratings.

Table 6 shows the full model for the analysis of the teachers' ratings of how well prepared the children were for kindergarten. As shown there, the difference between the TN-VPK participants and nonparticipants was statistically significant, with the TN-VPK participants rated as more prepared. Table 7 provides a summary of the results from parallel analyses for all the ratings and includes the standardized mean difference effect sizes for the difference between the TN-VPK participants and nonparticipants. The children who participated in TN-VPK were not only rated as being more ready for school but also as having better social behavior and work-related skills in the classroom. However, the teachers did not see significant differences between the two groups in peer relations, behavior problems, or feelings about school. The effects of exposure to TN-VPK, therefore, were apparent in several ways to kindergarten teachers, and in the areas most closely aligned with the typical focus of pre-k programs.

TN-VPK Effects for Different Subgroups of Children at the End of the Pre-K Year

As reported above, positive and statistically significant overall effects of TN-VPK

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

were found on all but one of the WJ achievement measures and several of the rating scales completed by kindergarten teachers. These findings motivate attention to our second research question, whether there are differential effects for different subgroups of children. This question was addressed using analytic models similar to those described above with the addition of interaction terms between TN-VPK participation status and variables representing the various subgroups of children. The specific variables used as moderators in these analyses were the following:

- The WJ Composite6 baseline measure, included to examine differential effects for children who began the pre-k year with higher or lower achievement performance.
- Age, indexed as age on September 1 of the pre-k year for the respective cohorts.
- Gender, represented by a dummy code distinguishing boys from girls.
- Race/ethnicity and whether children were native English speakers or not. These were not entirely distinct categories because most of the non-native English speaking children were Hispanic. A more differentiated set of subgroup dummy codes was therefore defined for these analyses as follows:
 - Black native English speakers (N=233)
 - Hispanic native English speakers (N=34)
 - Children with English as a second language irrespective of race/ethnicity (N=215).The remaining 594 children were White with a sprinkling of Asian and others and all native English speakers. That category was used as the reference value.
- Family background, including the home literacy index, mother's education, and number of working parents.

Initial analyses estimating effects on the WJ Composite6 achievement measure with

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

each moderator in turn showed statistically significant interactions with baseline achievement, home literacy, mother's education, and English as a second language (ESL). Further exploration of combinations of these moderators, however, revealed that these results were driven by interactions involving the ESL status of the children and mothers' education, particularly mothers with less than a high school education.

To more clearly reveal the nature of these interactions, the TN-VPK effects were examined in relation to child ESL and mothers' education together. These breakouts with the differences between TN-VPK participants and nonparticipants on the WJ Composite⁶ achievement measure along with the corresponding effect sizes are shown in Table 8. For comparability across groups and with the overall effects on the WJ Composite⁶ reported in Table 5, the effect sizes are standardized on the pooled standard deviations for the overall participant and nonparticipant groups. Table 5 shows that TN-VPK effects on achievement were much larger for ESL children than for native English speaking children (effect sizes of .67 vs. .23). Additionally, effects were larger for children of mothers with less than a high school education than for children of more educated mothers (effect sizes of .53 vs. .27). Moreover, the effect size was even larger for ESL children whose mothers had less than a high school education (ES= .88). The largest subgroup, native English speaking children with mothers who had completed high school or more, included 74% of the total sample and had the smallest effect size (ES= .22).

Whether TN-VPK Effects were Sustained through Later School Years

The results described above demonstrate positive TN-VPK effects at the end of the pre-k year on nearly all of the outcome variables included in this study. Given those results, the next question is whether those effects are sustained beyond the pre-k year.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Analysis of TN-VPK effects on the follow-up measures used the same multilevel models, propensity scores, and covariates employed in the analysis of the end of pre-k effects with only minor variations (e.g., dropping the rating time lag covariate that applied only to teacher ratings at the beginning of kindergarten). Also, the WJ Passage Comprehension and Calculation measures added at the end of kindergarten did not have baseline pretest measures to use as covariates and the baseline WJ Composite6 measure was used instead.

Table 9 shows the results of the analysis of TN-VPK effects on the WJ achievement measures at the end of the kindergarten and 1st, 2nd, and 3rd grade years, with the end of pre-k results repeated for ease of comparison. In contrast to the effects found at the end of pre-k, there were no statistically significant differences between TN-VPK participants and nonparticipants on any of the achievement measures at the end of kindergarten or the end of 1st grade. By the end of the 2nd and 3rd grade years, the effects were reversed for all the scales, reaching statistical significance for the WJ Composite6 and WJ Composite8 summary measures as well as several individual scales, notably those assessing math achievement. That is, the children who had not attended TN-VPK outperformed the children who had attended on these measures during these later years.

This pattern of positive TN-VPK effects during the pre-k year that rapidly diminish, then reverse, can be seen in Figure 1 where the covariate-adjusted WJ Composite6 W-score outcomes are plotted for each year for each group. As Figure 1 shows, both the TN-VPK participants and nonparticipants made achievement gains each year. However, the early advantage of the TN-VPK participants disappeared as the nonparticipating children caught up during the kindergarten year, matched the performance of the TN-VPK participants through the end of 1st grade, and then edged ahead in 2nd and 3rd grade.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

A different frame of reference for the achievement trajectories of the TN-VPK participants and nonparticipants is provided when the WJ standard scores are examined in place of the *W*-scores graphed in Figure 1. The standard scores are normed so that a score of 100 represents the mean for the norming sample, presumed to be representative of the national population of children at each respective age. Figure 2 shows the covariate-adjusted standardized scores from pre-k through 3rd grade for TN-VPK participants and nonparticipants. The same pattern of reversing differences between participants and nonparticipants evident in Figure 1 is also apparent in the standard score trajectories. In addition, Figure 2 shows that, relative to national norms, the early gains made by both groups began to flatten out in 1st grade and turned downward in 2nd and 3rd grade.

Moderator Relationships with Follow-up Achievement Outcomes

The analysis of TN-VPK effects at the end of pre-k reported earlier identified two significant moderators of effects on the WJ Composite6 measure. Larger effects were found for ESL children and for children of mothers with less than a high school education, and even larger effects were found for ESL children with less educated mothers. Analysis of the follow-up waves of outcome measures also examined the two-way and three-way interactions between these moderators and TN-VPK participation. However, no significant effects were found in those later years for any of these interactions.

In light of the overall finding of the difference between TN-VPK participants and nonparticipants on achievement measures reversing in 2nd and 3rd grade, it is informative to consider whether that pattern characterizes the native English speaking and ESL children when considered separately. Table 10 reports the mean *W*-scores on the WJ Composite6 outcomes from baseline to end of 3rd grade for these two subgroups of

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

children, further divided into TN-VPK participants and nonparticipants. The mean observed scores are reported for the TN-VPK participant groups; for the nonparticipant groups, the means are covariate adjusted to match the characteristics of the respective participant group. The only statistically significant interaction between native language status and TN-VPK participation was the one that occurred at the end of the pre-k year described earlier, but the large baseline differences are evident.

Figure 3 shows the trajectories on the WJ Composite6 W-scores for the ESL vs. native English speakers graphically from baseline through 3rd grade. The lower starting point and especially strong gains made by the ESL children during the pre-k year can be clearly seen. As with the overall sample, however, this early TN-VPK advantage for the ESL children has disappeared by the end of the kindergarten year and reverses after that. Perhaps most striking in Figure 3 is the performance of the ESL children in the later grades. Though they began with lower achievement scores than the native English speaking children, they had closed much of that gap by the end of kindergarten and, for the TN-VPK nonparticipants, even more of it by the end of 3rd grade. The native English-speaking children, by contrast, showed smaller effects of TN-VPK participation and smaller differences between participants and nonparticipants at the end of the 2nd and 3rd grade years. Recall that the WJ measures were administered in English, so much of the early gain for the ESL children likely reflected their increased mastery of English language.

Teacher Ratings

The results of the analysis of teacher ratings at the end of 1st, 2nd, and 3rd grade are shown in Table 11 along with those for the beginning of kindergarten. As with the achievement measures, some of the positive effects found at the end of the pre-k year

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

reversed in the later years. At the end of the 1st grade year, teachers rated TN-VPK participants significantly lower than nonparticipants on work-related skills, feelings about school, and preparedness for grade. Indeed, all of the effect estimates are negative at that point, though only those three reached statistical significance (marginally for preparedness for grade). However, by the end of the 2nd grade there were no longer any significant differences and that pattern continued into 3rd grade with the exception of a marginally significant positive effect for the TN-VPK participants on teachers' ratings of peer relations.

DISCUSSION

As noted at the beginning of this paper, research specifically on the effects of scaled-up state-funded pre-k programs is remarkably thin. Despite acknowledged limitations resulting from the modest and differential post-randomization consent rates, the study presented here has several characteristics that make it less vulnerable to bias than most, if not all, of the currently available longitudinal research on the effects of state pre-k programs. A large majority of the children whose outcomes were compared acquired their status as TN-VPK participants or nonparticipants as a result of the randomization process implemented in the larger study of which they are a subsample. Moreover, those groups were substantially similar on an array of baseline variables that included achievement pretests, family background, and child demographics as well as whatever unobserved variables were associated with the initiative their parents took to enroll them in the pre-k program. Additionally, propensity scores and selected covariates were used as statistical controls in the analysis to further reduce any bias associated with the few baseline variables on which the groups were not closely comparable.

These methodological credentials make the results of this study especially

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

discouraging in relation to the high expectations for publicly funded pre-k held by advocates. Like the Head Start Impact study—the only other relatively well-controlled longitudinal study of publicly funded pre-k—we found immediate advantages on achievement outcomes for TN-VPK participants during the pre-k year that were not sustained as nonparticipants rapidly caught up in the following years. However, while the Head Start study found that those early achievement effects had mostly disappeared by the end of 3rd grade, the most striking and unexpected finding for TN-VPK was that the early achievement effects became negative in the 2nd and 3rd grades. Also like the Head Start Impact Study, we did not find sustained effects on the noncognitive social-emotional outcomes that have been hypothesized to be the mediators for the positive life outcomes found in the Perry Preschool and Abecedarian studies. While the Head Start study found mostly negative effects on those outcomes by 3rd grade, we found only transitory negative effects in 1st grade that diminished thereafter. These results have led us to think about the many challenges associated with scaling up state-funded pre-k programs, some of which seem especially pertinent to the ongoing national interest in such programs.

Defining Pre-k

The benefits of pre-k are often touted as if the term *pre-k* refers to a well-defined program, but pre-k takes many different forms and there is no reason to believe they all have the same effects. The TN-VPK program is similar to other state initiatives in that its classrooms are primarily located in public schools, in effect defining pre-k as the school grade below kindergarten. However, this is not the only way states have provided preschool programs for children. Florida, for example, relies entirely on private providers, giving families a voucher they can use at any approved program. In North Carolina, Smart

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Start, begun in the early 1990s, did not focus on classrooms at all. Instead, funding was allocated to counties to create high quality and seamless services for children aged 0-5, and it was left to the counties to determine how to do that (Ladd, Muschkin & Dodge, 2014).

As Quinton (2014) recently noted: “...while there’s a growing consensus on the value of preschool, states disagree on where the programs should be based, who should run them, or how the government should support them” (p.2). There is no inherent necessity for state-funded pre-k to be housed in elementary schools and overseen by departments of education. From a policy perspective, it is important to differentiate different types of early childhood programs and to better understand their characteristics, costs, and effects.

Determining Quality

Pre-k advocates have attributed the disappointing findings of the TN-VPK study to the alleged low quality of the Tennessee program and asserted that high quality pre-k, by contrast, can be expected to have much more favorable effects (e.g., Kirp, 2015). Whether judged by the NIEER standards, conventional classroom observation measures, or the magnitude of the effects found at the end of the pre-k year, there is no evidence that the quality of TN-VPK is much different from that of other state programs for which these indicators are available (Farran & Lipsey, 2015). While no one would argue that public pre-k programs should not aspire to high quality, it is not at all apparent what that means.

When Tennessee began its pre-k program, it looked for guidance, as many states do, to the benchmarks established by NIEER (Barnett, et al, 2014). TN-VPK was set up to meet 9 of those 10 benchmarks and is among the states with top rated programs by those standards. Our TN-VPK findings add to the questions that have been raised about whether

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

those benchmarks prescribe features of pre-k programs that are linked to sustained effects on achievement or behavior (Mashburn et al., 2008). Another approach is to rely on rating systems to determine the quality of early childhood classrooms, e.g., the Early Childhood Environmental Rating Scale (ECERS; Harms & Clifford, 1980; with later editions) or the Classroom Assessment Scoring System (CLASS; LaParo & Pianta, 2003) now required of Head Start classrooms. However, Weiland et al. (2013), among others, have found that classroom quality as measured by these instruments had very small or no relationships to children's developmental outcomes. Fundamental empirical work is needed to identify the classroom environments and instructional practices that actually influence young children's development as a basis for defining pre-k quality in meaningful and actionable ways that can be used at the scale of statewide programs.

Alignment with K-3

Our findings highlight the importance of the K-3rd grade experience for children, especially children from low-income backgrounds. The fade out of pre-k effects could, at least in part, be due to failure of kindergarten teachers to build on the skills children bring with them from pre-k. This might happen, for example, if teachers mainly direct their attention to the children who need it the most, thus helping them catch up with those who have been in pre-k. Indeed, some explorations of what kindergarten teachers cover in their classrooms suggest that they may be out of touch generally with the skills their children possess (Claessens, Engel, & Curran, 2014), and thus their instruction thus may not be directed specifically to what any of the children are prepared to learn.

Conclusion

As we noted at the beginning of this paper, increasing numbers of children are living

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

in impoverished circumstances that have both immediate and long lasting adverse consequences for them. Pre-k intervention has been advocated as one way to address this problem and is expanding quickly in many states. However, the idea that pre-k can be easily scaled up in ways that will have the lifelong effects and return on investment found in the small, intensive programs implemented long ago that are so often cited as the models for effective pre-k is questionable at best. We do not yet have contemporary pre-k models that have been implemented at scale and convincingly shown to have enduring positive effects. It is therefore not at all obvious that the rush to implement pre-k widely without the necessary attention to identifying the characteristics that constitute program quality provides worthwhile benefits to children living in disadvantaged environments.

The TN-VPK program saturates the state; every county and all school districts except one have at least one classroom. Thus, the structural support exists in the state to continue to explore pre-k as a means for preparing children for success in school, but we need to think carefully about what the next steps should be. It is apparent that the term *pre-k*, or even *high-quality pre-k*, does not convey actionable information about what the critical elements of the program should be. Now is the time for policymakers and researchers to pay careful attention to the challenge of serving the country's youngest and most vulnerable children well in the pre-k programs that have been developed and promoted with their needs in mind.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

ACKNOWLEDGEMENTS

This research was supported by Grant #R305E090009 from the Institute of Education Sciences, U. S. Department of Education. It would not have been possible without the assistance of the Tennessee Department of Education, especially Connie Casha, Director of the Office of Early Learning; Bobbi Lussier, former Assistant Commissioner of Special Populations; and Robert Taylor, consultant and former Superintendent of Bradley County Schools. The Tennessee's Consortium on Research, Evaluation, and Development provided essential access to the Tennessee education data system. Special thanks go to colleagues who have served this project well in so many ways: Carol Bilbrey, Project Manager for five years; Janie Hughart, who stepped in when Dr. Bilbrey retired; Nianbo Dong who served as a key data analyst; and Georgine Pion who assisted with the multiple imputations.

REFERENCES

- Almond, D. & Currie, J. (2010). Human capital development before age five. National Bureau of Economic Research, Working Paper 15827.
<http://www.nber.org/papers/w15827>
- Baker, M. (2011). Innis Lecture: Universal early childhood interventions: What is the evidence base? *Canadian Journal of Economics/Revue canadienne d'économie*, 44, 1069-1105.
- Barnett, W. S. (2013). Expanding access to quality pre-k is sound public policy. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. S., Carolan, M., Squires, J., Brown, K., & Horowitz, M. (2014). The state of preschool 2014: State preschool yearbook. New Brunswick, NJ: National Institute for Early Education Research.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

- Barnett, W. S., Jung, K., Youn, M-J., & Frede, M. C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. New Brunswick, NJ: National Institute for Early Education Research.
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37, 231-242.
- Cascio, E. U. & Schanzenbach, D. W. (2013). The impacts of expanding access to high quality preschool education. Paper presented at the fall 2013 Brookings Panel on Economic Activity.
- Claessens, A., Engel, M., Curran, F.C. (2014). Academic content, student learning and the persistence of preschool effects. *American Educational Research Journal*, 51, 403-434.
- Cooper, D. & Farran, D.C. (1991). Cooper-Farran Behavioral Rating Scale. Clinical Psychology Publishing Company, Inc.
- Currie, J. & Rossin-Slater, M. (2014). Early-life origins of life-cycle well-being: Research and policy implications. *Journal of Policy Analysis and Management*, 34, 208-242.
- Duncan, G. & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, 27, 109-132.
- Duncan, G., Ziol-Guest, K., & Kalil, A. (2010). Early childhood poverty and adult attainment, behavior, and health. *Child Development*, 81, 306-325.
- Farran, D. C. (2007). Is education the way out of poverty? A Reflection on the 40th anniversary of Head Start. Center for Research on Child Development and Learning, No. 3.
- Farran, D.C., Bilbrey, C. & Lipsey, M. (2003). Academic and Classroom Behavior Record.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Unpublished scale available from D.C. Farran, Peabody Research Institute, Vanderbilt University, Nashville, TN.

Farran, D. C., & Lipsey, M. W. (2015). Expectations of sustained effects from scaled up pre-k: Challenges from the Tennessee study. *Evidence Speaks*, 1:4, October 8. Washington, DC: Brookings. <http://www.brookings.edu/research/papers/2015/10/08-expectations-of-sustained-effects-from-scaled-up-prek-challenges-tennessee-study-farran-lipsey>.

Farran, D.C. & Lipsey, M. (in press). Evidence for the benefits of state pre-kindergarten programs: Myth and misrepresentation. *Behavioral Science and Policy*.

Fitzpatrick, M. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The B.E. Journal of Economic Analysis & Policy*, 8, 1-38.

Gilliam, W. S., & Zigler, E. F. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly*, 15, 441-473.

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

Grehan, A., Cavalluzzo, L., Gnuschke, J., Hanson, R., Oliver, S, and Vosters, K. (2011). Participation during the first four years of Tennessee's Voluntary Prekindergarten program. (Issues & Answers Report, REL 2011-No. 107). Washington, DC: U.S. Department of Education. <http://ies.ed.gov/ncee/edlabs>.

Harms, T., & Clifford, R. M. (1980). *The Early Childhood Environment Rating Scale*. New York: Teachers College Press.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

children. *Science*, 312, 1900-1902.

Heckman, J. (2008). Schools, skills, and synapses. *Economic Inquiry*, 46, 289-324.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103, 2052-2086.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

Kirp, D. L. (2015). Does pre-k make any difference? *New York Times Sunday Review*, October 3.

Ladd, H., Muschkin, C., & Dodge K. (2014). From birth to school: Early childhood initiatives and third grade outcomes in North Carolina. *Journal of Policy Analysis and Management*, 33, 162-187.

LaParo, K. M., & Pianta, R. C. (2003). *CLASS: Classroom Assessment Scoring System*. Charlottesville: University of Virginia

Lazar, I., Darlington, R., Murray, H., Royce, J., & Snipper, A. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development*, 47:2-3, Serial No. 195.

Lipsey, M. W., Farran, D. C., Bilbrey, C., Hofer, K. G., Dong, N. (2011). Initial results of the evaluation of the Tennessee Voluntary Pre-K Program. Research Report. Nashville, TN: Vanderbilt University, Peabody Research Institute.

https://my.vanderbilt.edu/tnprekevaluation/files/2013/10/April2011_PRI_Initial_TN-VPK_ProjectResults.pdf

Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013a). Evaluation of the

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Tennessee Voluntary Prekindergarten Program: End of pre-k results from the randomized control design. Research Report. Nashville, TN: Vanderbilt University, Peabody Research Institute.

https://my.vanderbilt.edu/tnprekevaluation/files/2013/10/May2013_PRI_EndofPK_TN-VPK_RCT_ProjectResults.pdf

Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013b). Evaluation of the Tennessee Voluntary Prekindergarten Program: Kindergarten and first grade follow-up results from the randomized control design. Research Report. Nashville, TN: Vanderbilt University, Peabody Research Institute.

https://my.vanderbilt.edu/tnprekevaluation/files/2013/10/August2013_PRI_Kand1stFollowup_TN-VPK_RCT_ProjectResults_FullReport1.pdf

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

Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., Burchinal, M., Early, D., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732-749.

Minervino, J. & Pianta, R. (September, 2014). Early learning: The new fact base and cost sustainability. In J. Minervino (Ed.), *Lessons from research and the classroom*. Washington: Bill & Melinda Gates Foundation.

Mistler, S. A. (2013). A SAS® macro for applying multiple imputation to multilevel data. *Proceedings of the SAS Global Forum 2013, San Francisco, California. Statistics and Data Analysis*, 438-2013.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Peisner-Feinberg, E. S., Schaaf, J. M., LaForett, D. R., Hildebrandt, L.M., & Sideris, J. (2014).

Effects of Georgia's Pre-K Program on children's school readiness skills: Findings from the 2012-2013 evaluation study. Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, D., Mashburn, A. & Downer, J.

(2012). Third grade follow-up to the Head Start Impact Study final report. OPRE Report # 2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Quinton, S. (September 4, 2015). States agree on need for 'preschool,' differ on definition.

The Pew Charitable Trusts, Research & Analysis, Stateline. Retrieved September 5, 2012, from <http://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/>

ReadyNation (n.d.). Business case for early childhood investments. Washington, DC:

ReadyNation. Retrieved September 12, 2015, from www.ReadyNation.org.

Reardon, S. (2011). The widening academic achievement gap between the rich and the

poor: New evidence and possible explanations. In G. Duncan & R. Murnane, (Eds).

Whither opportunity: Rising inequality, schools, and children's life chances. New York: Russell Sage Foundation.

Reynolds, A., Temple, J, Robertson, D., White, R., & Ou, S-R (2011). Age 26 cost-benefit

analysis of the Child-Parent Center early education program. *Child Development*, 82, 379-404.

Rosinsky, K. (2014). The relationship between publicly funded preschool and 4th grade

math test scores: A state-level analysis. Master's thesis, Georgetown University,

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Washington, DC. Retrieved August 31, 2015 from

https://m.repository.library.georgetown.edu/bitstream/handle/10822/709852/Rosinsky_georgetown_0076M_12517.pdf?sequence=1&isAllowed=y.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005).

Lifetime effects: the High/Scope Perry Preschool study through age 40. Ypsilanti, MI: High/Scope Press.

SREB (2015). Building a strong foundation: State policy for early childhood education. The report of the SREB Early Childhood Commission. Atlanta, GA: Southern Regional Education Board.

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

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

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

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

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

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 1: Sample Retention for Each Data Collection Wave by Condition

	Year 1 (Pre-K)	Year 2 (K)	Year 3 (1st)	Year 4 (2nd)	Year 5 (3rd)
TN-VPK Participants	773	749 (.97)	738 (.95)	726 (.94)	714 (.92)
Nonparticipants	303	297 (.98)	291 (.96)	290 (.96)	280 (.92)
All Participants	1076	1046 (.97)	1029 (.96)	1016 (.94)	994 (.92)

Note: The proportions retained are shown in parentheses.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 2: Comparison of Participant and Nonparticipant Groups on Baseline Measures

Variable	TN-VPK participants [N=773] Mean (SD)	TN-VPK non-participants [N=303] Mean (SD)	p-value	Effect size	PS p-value ^a	PS adj. ES ^b
Age (years)	4.4 (.28)	4.4 (.29)	.533	-.04	.937	.01
Gender (1=male)	.47 (.50)	.48 (.50)	.932	-.01	.994	.00
Race/ethnicity Black (1=yes)	.21 (.42)	.19 (.43)	.449	.05	.802	-.02
Race/ethnicity Hispanic (1=yes)	.14 (.37)	.15 (.44)	.694	-.03	.303	.08
Native language English (1=yes)	.86 (.37)	.84 (.46)	.571	.04	.461	-.06
Not Hispanic, native English (1=yes)	.83 (.40)	.81 (.47)	.619	.03	.279	-.09
Hispanic, native English (1=yes)	.03 (.17)	.03 (.19)	.721	-.02	.443	.07
Hispanic, not native English (1=yes)	.11 (.34)	.13 (.42)	.639	-.03	.502	.05
Not Hispanic, not native English (1=yes)	.03 (.18)	.04 (.26)	.510	-.05	.849	.02
Library card use (0-2)	.96 (.82)	.89 (.84)	.216	.09	.876	.01
Newspaper subscriptions (0-3)	.38 (.76)	.33 (.75)	.417	.06	.702	.04
Magazine subscriptions (0-2)	.29 (.50)	.26 (.51)	.423	.06	.332	-.09
Home literacy index	.16 (2.03)	-.02 (1.96)	.223	.09	.826	-.02
Mother's education (1-4)	2.16 (.72)	2.02 (.74)	.010	.19	.610	-.04
Number of working parents	1.25 (.62)	1.23 (.62)	.641	.03	.990	.00
WJ Letter-Word Identification	319.2 (27.0)	315.1 (27.2)	.035	.15	.815	-.02
WJ Spelling	350.6 (28.4)	349.3 (28.5)	.534	.04	.880	.01
WJ Oral Comprehension	444.4 (15.6)	442.9 (17.5)	.206	.09	.477	-.06
WJ Picture Vocabulary	457.1 (21.0)	454.4 (27.8)	.088	.12	.329	-.08
WJ Applied Problems	392.1 (26.9)	391.6 (29.9)	.818	.02	.344	-.08
WJ Quantitative Concepts	407.6 (13.9)	407.3 (14.3)	.789	.02	.930	.01
WJ Composite6	395.2 (17.7)	393.6 (19.1)	.202	.09	.561	-.05

Notes: Age on Sept. 1 of pre-k year; 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); Home literacy index = sum of the z-scores for Library card, Newspaper subscriptions, and Magazine subscriptions; Mother's education (1=less than high school, 2=high school diploma/GED, 3=associate's degree, 4=more than associate's degree); WJ= W-scores on the indicated Woodcock Johnson pretests.

(a) p-value for difference between means for participants and nonparticipants with the propensity score as a covariate.

(b) Effect size for the difference between means for participants and nonparticipants with the propensity score as a covariate.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 3: Comparison of Participant and Nonparticipant Groups on Timing Variables

Time from School Start Date	TN-VPK participants [N=773] Mean (SD)	TN-VPK non-participants [N=303] Mean (SD)	<i>p</i> -value	Effect Size	PS <i>p</i> -value ^a	PS adj. ES ^b
Days to pretest	71 (22.8)	86 (30.8)	.000	-.61	.607	.03
Days to pre-k posttest	267 (13.5)	279 (20.2)	.000	-.79	.604	.03
Days to K follow-up	626 (21.4)	629 (22.2)	.111	-.11	.243	.02
Days to 1 st grade follow-up	987 (26.4)	990 (29.0)	.110	-.11	.780	.02
Days to 2 nd grade follow-up	1335 (26.5)	1337 (30.0)	.256	-.08	.505	.05
Days to 3 rd grade follow-up	1695 (28.7)	1696 (43.5)	.910	-.01	.948	.01

(a) *p*-value for difference between means for participants and nonparticipants with the propensity score as a covariate.

(b) Effect size for the difference between means for participants and nonparticipants with the propensity score as a covariate.

Table 4: Full Analysis Results for the WJ Composite6 Outcome Measure at the End of the Pre-k Year

	Coefficient	Standard error	<i>t</i> -value	<i>p</i> -value
Intercept	91.74	7.14	12.86	.000
Propensity score	5.92	1.46	4.06	.000
Composite6 pretest	0.79	0.02	43.65	.000
Age (years)	-0.84	0.95	-0.88	.377
Gender (1=male)	-0.18	0.52	-0.34	.734
Race/ethnicity Black	1.15	0.70	1.65	.100
Hispanic, native English	1.22	1.52	0.80	.423
Hispanic, not native English	2.59	0.93	2.78	.005
Not Hispanic, not native English	0.29	1.38	0.21	.834
Home literacy index	0.05	0.14	0.35	.723
Mother's education	0.42	0.39	1.09	.278
Number of working parents	0.07	0.42	0.16	.876
TN-VPK participation	5.32	0.75	7.06	.000

Notes: Age on Sept. 1 of prek year; Home literacy index = sum of the z-scores for Library card, Newspaper subscriptions, and Magazine subscriptions; Mother's education (1=less than high school, 2=high school diploma/GED, 3=associate's degree, 4=more than associate's degree).

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 5: TN-VPK Effect Estimates for Pre-K Gain on Woodcock Johnson Achievement Measures

Outcome	TN-VPK effect estimate in W-score units	p-value	Effect size	Effect size for non-participant gain	Effect size for TN-VPK participant gain	% Increase in Gain for TN-VPK participants
WJ Composite6	5.32	<.001	.32	.74	1.06	44%
<i>Literacy Measures</i>						
Letter-Word Identification	10.77	<.001	.41	.60	1.01	68%
Spelling	7.22	<.001	.29	.80	1.09	36%
<i>Language Measures</i>						
Oral Comprehension	1.50	.093	.09	.44	.53	20%
Picture Vocabulary	3.66	<.001	.20	.24	.44	83%
<i>Math Measures</i>						
Applied Problems	4.03	.005	.17	.61	.78	28%
Quantitative Concepts	4.32	<.001	.27	.68	.96	40%

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 6: Full Analysis Results for the Kindergarten Teachers' Ratings of How Well Prepared the Children Were for Kindergarten

	Coefficient	Standard error	t-value	p-value
Intercept	-16.49	1.14	-14.57	.000
Propensity score	0.40	0.20	1.99	.046
Rating time lag	-0.00	0.00	-0.60	.547
Composite6 pretest	0.05	0.00	19.28	.000
Age (years)	0.04	0.14	0.31	.754
Gender (1=male)	-0.17	0.08	-2.28	.023
Race/ethnicity Black	0.17	0.10	1.65	.100
Hispanic, native English	0.34	0.22	1.52	.129
Hispanic, not native English	0.90	0.13	6.78	.000
Not Hispanic, not native English	0.48	0.20	2.42	.016
Home literacy index	-0.01	0.02	-0.67	.506
Mother's education	0.02	0.06	0.38	.703
Number of working parents	-0.04	0.06	-0.57	.569
TN-VPK participation	0.30	0.11	2.79	.005

Notes: Age on Sept. 1 of pre-k year; Home literacy index = sum of the z-scores for Library card, Newspaper subscriptions, and Magazine subscriptions; Mother's education (1=less than high school, 2=high school diploma/GED, 3=associate's degree, 4=more than associate's degree).

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 7: TN-VPK Effect Estimates for Kindergarten Teachers' Ratings

Outcome	TN-VPK effect		
	estimate	<i>p</i> -value	Effect size
ACBR Preparedness for K (range 1-7)	.30	.005	.22
ACBR Peer Relations (range 1-7)	.04	.684	.04
ACBR Behavior Problems ^a (range 0-1)	-.01	.757	-.04
ACBR Feelings About School ^a (0-1)	-.00	.767	-.03
Cooper-Farran Interpersonal Skills (range 1-7)	.17	.049	.19
Cooper-Farran Work-Related Skills (range 1-7)	.22	.016	.20

(a) Ratings on these scales were skewed; the analysis was done on log transformed values and those are the results shown here

Table 8: TN-VPK Effects on the WJ Composite6 Achievement Composite for Subgroups of Children Who Differ by English Speaking Status and Mothers' Education

	Mother's education	
	<u>Less than HS (N=178)</u>	<u>HS or more (N=898)</u>
Child Language	T-C diff= 8.74* Effect size= .53	T-C diff= 4.50* Effect size= .27
<u>English as second language (N=215)</u> T-C difference= 11.07* Effect size= .67	T-C diff= 14.57* Effect size= .88 (N=76)	T-C diff= 9.04* Effect size= .55 (N=139)
<u>Native English speaker (N=861)</u> T-C difference= 3.74* Effect size= .23	T-C diff= 4.48 Effect size= .27 (N=102)	T-C diff= 3.63* Effect size= .22 (N=759)

T= TN-VPK participants; C=nonparticipants.

* $p < .05$

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 9: TN-VPK Effect Estimates for the Kindergarten through 3rd Grade Years on the Woodcock Johnson Achievement Measures

Outcome	End of pre-k year		End of kindergarten year		End of 1st grade year		End of 2nd grade year		End of 3rd grade year	
	Effect estimate	Effect size	Effect estimate	Effect size	Effect estimate	Effect size	Effect estimate	Effect size	Effect estimate	Effect size
WJ Composite6	5.32**	.32	.25	.02	-.51	-.04	-2.07*	-.15	-1.83 [†]	-.13
WJ Composite8	N/A	-	-.13	-.01	-.70	-.05	-1.91*	-.15	-1.73 [†]	-.13
<i>Literacy</i>										
Letter-Word ID	10.77**	.41	-.27	-.01	-1.56	-.05	-3.24	-.13	-3.46	-.14
Spelling	7.22**	.29	-.68	-.03	-2.11	-.10	-2.45	-.12	-2.36	-.12
<i>Language</i>										
Oral Comprehension	1.50 [†]	.09	.94	.06	-.90	-.07	-1.43	-.11	-.51	-.04
Picture Vocabulary	3.66**	.20	1.01	.09	.95	.08	-.48	-.04	.77	.07
Passage Comprehension	N/A	-	-2.26	-.10	-1.61	-.08	-2.10 [†]	-.13	-1.13	-.07
<i>Math</i>										
Applied Problems	4.03**	.17	1.17	.07	.55	.04	-2.38 [†]	-.14	-3.76*	-.21
Quantitative Concepts	4.32**	.27	-1.07	-.08	-1.33	-.10	-3.45**	-.25	-2.02 [†]	-.15
Calculation	N/A	-	-.13	-.01	-.70	-.05	-1.91*	-.15	-1.73 [†]	-.13

Notes: Effect estimates are the coefficients on the TN-VPK participation variable indicating the difference between the mean outcomes for T-VPK participants and nonparticipants in W-score units. Effect sizes are those coefficients divided by the pooled participant and nonparticipant group standard deviations on the outcome variable.

** $p < .01$, * $p < .05$, [†] $p < .10$

Table 10: ESL-Native English Moderator of Effects on WJ Composite6

Language	TN-VPK	Baseline	End of pre-k*	End of k	End of 1 st grade	End of 2 nd grade	End of 3 rd grade
Native English	Yes	398.7	414.5	443.1	466.1	479.6	491.1
	No	398.8	411.2	442.4	466.7	481.6	492.9
English as Second Language	Yes	377.7	402.3	434.4	458.1	473.1	484.7
	No	378.1	392.2	436.1	460.0	477.5	489.3

* $p < .05$ for the Language x TN-VPK participation condition interaction term in the regression model.

EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Table 11: TN-VPK Effect Estimates for 1st, 2nd, and 3rd Grade Teachers' Ratings

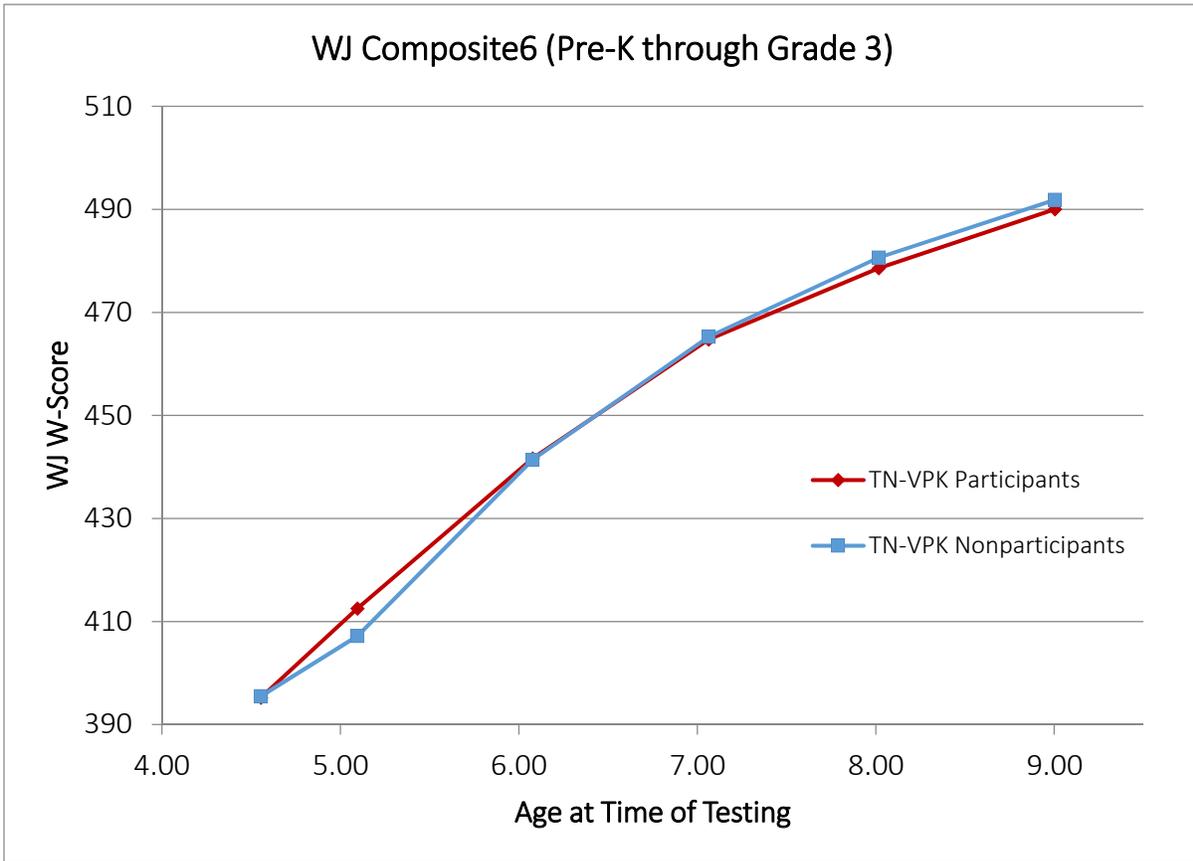
Outcome	Start of kindergarten year		End of 1st grade year		End of 2nd grade year		End of 3rd grade year	
	Effect estimate	Effect size	Effect estimate	Effect size	Effect estimate	Effect size	Effect estimate	Effect size
ACBR Preparedness for Grade	.30*	.22	-.24†	-.17	.07	.05	-.01	-.01
ACBR Peer Relations	.04	.04	-.05	-.05	.04	.04	.21†	.19
ACBR Behavior Problems	-.01	-.04	-.00	-.02	-.02	-.07	-.04	-.16
ACBR Feelings About School	-.00	-.03	-.01*	-.21	.00	.04	.00	.03
CF Interpersonal Skills	.17*	.19	-.15	-.16	.06	.06	.07	.07
CF Work-Related Skills	.22*	.20	-.24*	-.20	.00	-.00	.10	.08

Notes. Scoring range on scales: ACBR Preparedness (1-7); ACBR Peer Relations (1-7); ACBR Behavior Problems (log transformed, 0-1); ACBR Feelings About School (log transformed, 0-1); Cooper-Farran Social Behavior (1-7); Cooper-Farran Work-Related Skills (1-7).

* $p < .05$, † $p < .10$

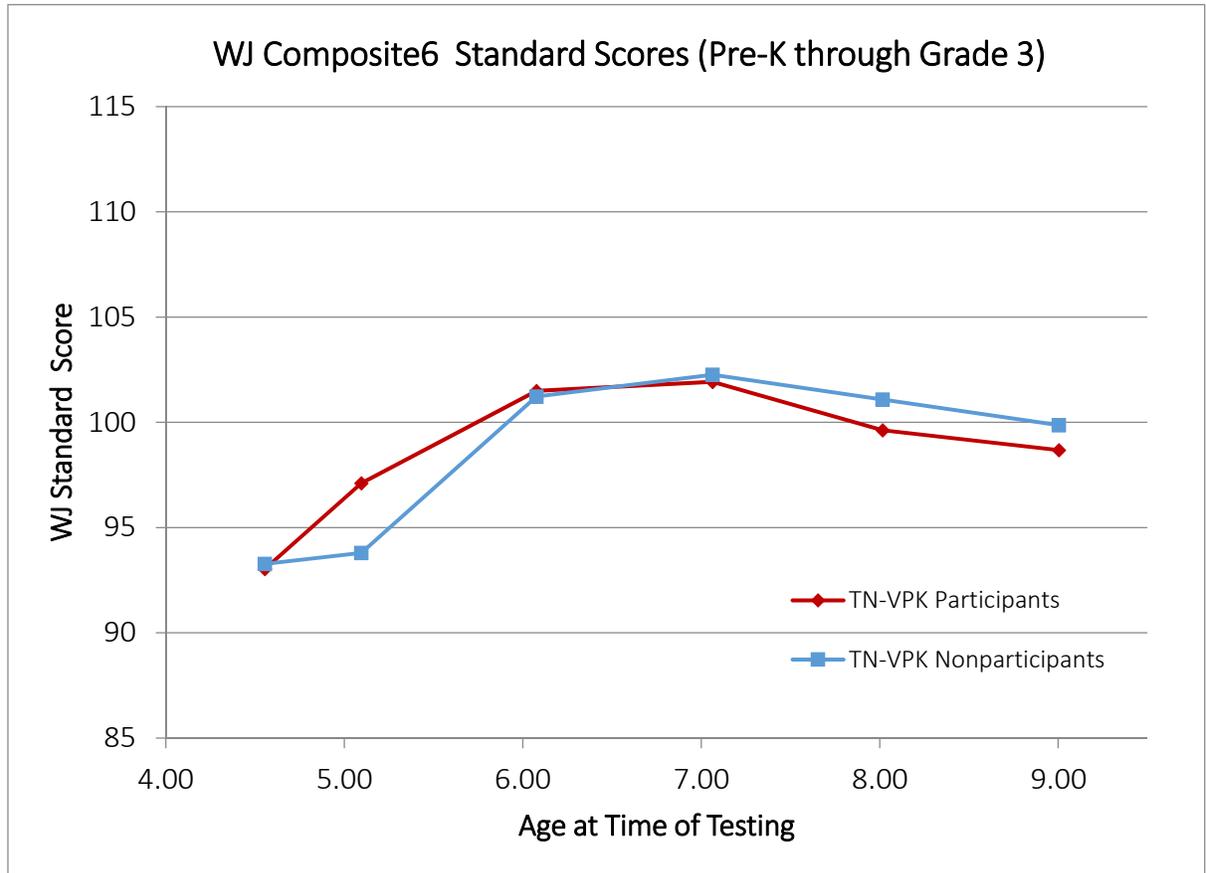
EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Figure 1: W-Scores on WJ Composite6 for the TN-VPK Participant and Non-Participant Groups on Each Wave of Measurement



EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Figure 2: Standard Scores on WJ Composite6 for the TN-VPK Participant and Non-Participant Groups on Each Wave of Measurement



EFFECTS OF A STATE PREKINDERGARTEN PROGRAM

Figure 3: WJ Composite6 for ESL and Native English Speakers at Each Wave of Measurement Broken out by TN-VPK Participation

