

THE EARLY YEARS

Preschool Programs Can Boost School Readiness

William T. Gormley Jr.,* Deborah Phillips, Ted Gayer

Preschool programs in Oklahoma help children to succeed in school.

Early childhood programs have been heralded by both scholars and decision-makers as a promising avenue for fostering school readiness (1–4). In the United States, these sentiments have fueled ambitious preschool initiatives from both state and federal resources.

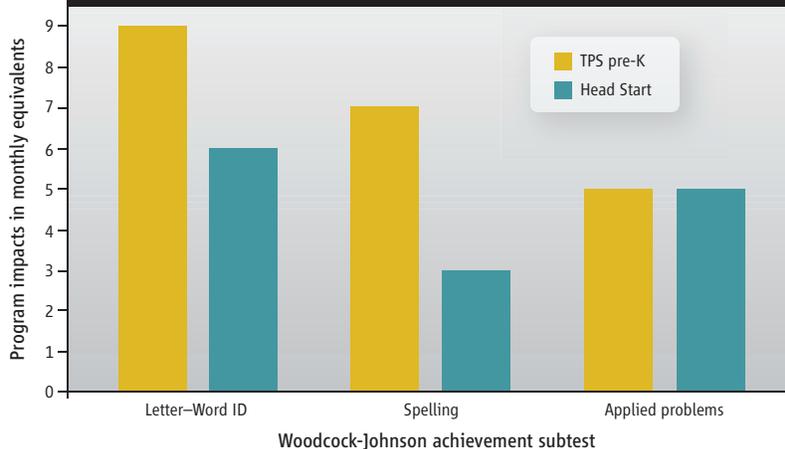
Oklahoma’s prekindergarten (pre-K) program has generated attention because it is universal, is based in the school system, and reaches a higher percentage of 4-year-olds than any other state pre-K program. Oklahoma’s state-funded pre-K program channels aid to local school districts, which are free to run full-day programs, half-day programs, or both. Federally funded Head Start programs, which are targeted to poor or otherwise at-risk children, and private day care centers are also eligible for state funding if they establish “collaborative” relations with their local school district. The Oklahoma pre-K program has relatively high standards compared with those of

other states and offers relatively high pay and benefits to well-qualified teachers. Every lead teacher must have a B.A. degree and must be certified in early-childhood education. Student teacher ratios of 10-to-1 and class sizes of 20 must be maintained. The Community Action Project (CAP) of Tulsa County, whose Head Start program serves the largest number of children in Tulsa, is eligible for state funding. Its teachers meet the same standards as their Tulsa Public Schools (TPS) counterparts and receive similar pay (5).

Here, we estimate the short-term test score gains for children in Tulsa’s pre-K and Head



Preschool programs strengthen reading, writing, and math skills.



Start programs. In August 2006, we administered three subtests of the Woodcock-Johnson Achievement Test to incoming Tulsa students entering pre-K and Head Start programs and kindergarten. The tests were administered just before the commencement of classes by the individual who would be teaching the child that year. About 78% of all pre-K entrants, 69% of all Head Start entrants, and 73% of all kindergarten entrants were tested. The tested students closely mirrored the universe of students, except for the kindergarten cohort, where some differences between the two groups were found (6). While the child was being tested, the parent completed a survey to provide demographic information. Program participation and school lunch eligibility were determined separately (7).

The three subtests we used were the letter-word identification test (a measure of preread-

ing skills); the spelling test (a measure of prewriting skills); and the applied problems test (a measure of premath skills). These subtests have been shown to be especially appropriate for children aged 4 to 5 and have been used by other researchers studying both at-risk and more diverse groups of children. Woodcock-Johnson test scores at or before school entry help to predict later scholastic achievement (8).

A difficult methodological challenge that confronts education researchers is selection bias. Here, the concern is that children whose parents choose for them to participate in the voluntary pre-K or Head Start programs will differ from non-participants and that these differences will affect test scores. To overcome such selection bias, we have used two separate regression-discontinuity estimations. First, the treatment group consisted of 1264 kindergarten students who attended Tulsa’s pre-K program and the comparison group was 1492 children who were about to begin that program. For the second estimation, the treatment group consisted of 327 kindergarten students who attended Tulsa’s Head Start program, and the comparison group was 483 children about to enter that program.

This research design is possible because the TPS and the Tulsa Head Start program strictly enforce a 1 September birthday requirement for enrollment in the 4-year-olds’ program.

Our analysis estimates the continuous relation between age and test score separately on both sides of the age cut-off. This is achieved by regressing test scores against the child’s precise date of birth (the number of days born before or after the cut-off qualification date), an age cut-off indicator variable, and an interaction variable that allows for different slopes on both

PHOTO CREDIT: JUPITERIMAGES

Georgetown Public Policy Institute and Center for Research on Children in the United States (CROCUS), Georgetown University, Washington, DC 20007, USA.

*To whom correspondence should be addressed. E-mail: gormleyw@georgetown.edu

sides of the cut-off point. The coefficient for the cut-off indicator is the estimated treatment effect. This estimated treatment effect is unbiased if there are no discontinuous differences at the cut-off in characteristics that contribute to test scores.

We tested this condition by comparing the regression-adjusted observed characteristics for children in the treatment and comparison groups at the age cut-off limit. For our TPS comparisons, the characteristics were not statistically different at the cut-off. For our Head Start comparisons, the characteristics matched very well, except that children in the comparison group were somewhat more likely to be female, somewhat more likely to be eligible for a reduced price lunch, and somewhat more likely to have a mother with no high-school degree. Consequently, one should be more cautious about interpreting the Head Start results.

In all of our regressions, we also included a wide range of demographic variables, including gender, race and/or ethnicity, school lunch eligibility, mother's education, whether the child lives with his or her biological father, and whether the child has Internet access at home (9). We handle missing data for three variables obtained through the parent survey by using multiple imputation. All of our numbers estimate the effects of treatment on the treated.

The TPS pre-K program has sharply improved students' cognitive development. One way to capture this is to look at the effect sizes: 0.985 for letter-word identification, 0.743 for spelling, and 0.355 for applied problems (10). These effect sizes substantially exceed those reported for pre-K programs generally and are somewhat greater than those reported for five states with relatively high quality pre-K programs (11, 12). The effects of the Tulsa Head Start program, though less spectacular, are also impressive: 0.514 for letter-word identification, 0.334 for spelling, and 0.369 for applied problems. These effect sizes exceed those reported for a national study of Head Start with random assignment of children (13).

Another way to express program effects is to convert test score impacts into monthly equivalents (see chart, page 1723). These gains are above and beyond those that otherwise occur through aging and maturation.

The different estimated test effects across Tulsa pre-K and Tulsa Head Start could be due to differences in the types of children enrolled or to differences in how the programs function. In order to focus on the latter, we trim the Head Start sample to include only children who were eligible for free lunches (typically children whose family income is

less than 130% of the federal poverty level). We then trim the pre-K sample to include only children who were eligible for free lunches and who participated in a full-day program (all Head Start children participated in a full-day program) (14). Even with these modifications, the two samples differ in the race and/or ethnicity of the students, although not in other observable characteristics. We therefore estimate separate program effect sizes for blacks and Hispanics, for both pre-K and Head Start. We do not report results for whites or Native Americans because of small Head Start sample sizes for these two subgroups.

Focusing on the pre-K population that most resembles the population of Head Start students allows us to better compare the two programs. There are larger test impacts for children who are eligible for free lunch in the TPS pre-K full-day program compared with those similarly qualified in the Tulsa Head Start program (a full-day program), for both blacks and Hispanics (fig. S2). The differences are larger for prereading and prewriting skills than for premath skills, where the TPS and Head Start programs are equally effective.

What explains these differences? One possibility is that the programs attract students of different abilities and family circumstances, even though we condition on free-lunch eligibility, full-day program, and race. Another possible explanation may lie in the differing classroom priorities: TPS emphasizes letters and sounds more, whereas Head Start emphasizes fantasy play more (fig. S3). Our analysis of kindergarten students who attended the two types of programs the previous year suggests that these two variables do help to explain part of the difference in verbal test score gains between TPS and Head Start (15). Specifically, the Head Start variable explains less of the variance in letter-word identification and spelling test scores in a model that includes these two classroom variables than in a model that excludes them. This suggests that the two classroom variables account for some of the performance gap between TPS and Head Start. However, we also find that TPS places more emphasis on math, which does not translate into higher pre-math test scores compared to Head Start.

Early childhood education programs in the United States face enormous challenges. The overwhelming majority of Head Start program participants are poor, and many Head Start children face additional risk factors, such as a single-parent home or a home where English is not the primary language spoken. Pre-K programs targeted to poor or otherwise at-risk children face similar challenges. Even

universally available programs, such as Oklahoma's, must cope with the realities of poor families, fragmented families, and immigrant families.

Against this backdrop, it is instructive to compare the potency of program participation variables with that of other variables in our statistical models. For the TPS model, program participation is a more powerful predictor of prereading and prewriting test score outcomes than gender, race and/or ethnicity, free lunch eligibility, mother's education, or whether the biological father lives at home (fig. S4). For the Head Start model, program participation is a more powerful predictor of premath outcomes than gender, free lunch eligibility, mother's education, or whether the biological father lives at home (fig. S5). Early childhood education can therefore make a big difference for short-term test scores, substantially muting the negative effects of family and environmental risk factors.

References and Notes

1. W. S. Barnett, *Am. J. Orthopsychiatry* **63**, 500 (1993).
2. A. J. Reynolds, J. A. Temple, D. L. Robertson, E. A. Mann, *JAMA* **285**, 2339 (2001).
3. J. J. Heckman, *Science* **216**, 1900 (2006).
4. E. Zigler, W. S. Gilliam, S. M. Jones, *A Vision for Universal Preschool Education* (Cambridge Univ. Press, New York, 2006).
5. The starting salary for a TPS teacher was \$29,000 for the 2005–06 school year; the starting salary for a CAP Head Start teacher was \$29,500.
6. Tested kindergarten students were somewhat less likely to be poor and black and somewhat more likely to be middle-class and white than the universe of kindergarten students.
7. Pre-K participation was determined from TPS administrative data; Head Start participation was determined from CAP Head Start of Tulsa County administrative data.
8. G. J. Duncan *et al.*, *Dev. Psychol.* **43**, 1428 (2007).
9. If the covariates balance on each side of the cut-off point, then including them in the regression does not reduce bias of the treatment effect.
10. Experts disagree on the best way to measure effect sizes. We prefer to divide the regression coefficient by the standard deviation of the control group, because the latter is the counterfactual of interest.
11. K. A. Magnuson, C. Ruhm, J. Waldfogel, *Econ. Educ. Rev.* **26**, 33 (2007).
12. V. C. Wong, T. D. Cook, W. S. Barnett, K. Jung, *J. Policy Anal. Manag.* **27**, 122 (2008).
13. M. Puma *et al.*, *Head Start Impact Study: First Year Findings* (Department of Health and Human Services, Washington, DC, 2005).
14. About two-thirds of all TPS program participants enroll in a full-day program.
15. We used hierarchical linear modeling (HLM) to distinguish between organizational-level and individual-level effects on cognitive development. See our SOM for details.
16. This research was supported by generous grants from the Foundation for Child Development, the Spencer Foundation, the David and Lucile Packard Foundation, and the A.L. Mailman Family Foundation. The authors alone are responsible for the contents. For additional information on the Oklahoma project, see www.crocus.georgetown.edu.

10.1126/science.1156019

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5884/1723/DC1