

PRESCHOOL PROGRAMS CAN BOOST
SCHOOL READINESS

William T. Gormley, Jr.

Deborah Phillips

Ted Gayer

Georgetown University

SOM (Supplementary Online Material)

ABSTRACT

A universally-available pre-K program in Tulsa, Oklahoma boosts pre-reading skills by nine months, pre-writing skills by eight months, and pre-math skills by five months. The Tulsa Head Start program boosts pre-reading skills by five months, pre-writing skills by three months, and pre-math skills by five months. Classroom practices help to explain these results. The negative effects of family and environmental risk factors can be mitigated by a strong preschool program.

ACKNOWLEDGMENTS

The authors would like to thank the Foundation for Child Development, the Spencer Foundation, the David and Lucile Packard Foundation, and the A.L. Mailman Family Foundation for their generous financial support. They would also like to thank the Frank Porter Graham Center and the National Institute for Early Education Research for sharing some testing materials at a critical moment. Dan Cullinan provided valuable technical support. Shirley Adelstein, Mireya Almazan, Joy Chen, Ah-Hyun Cho, Cat Edwards, Deanna Ford, Angeles Gottheil, Amy Lowenstein, Katie Newmark, Kate Perper, and Marie Szczurowski coded data and provided research support. Lisa Adams, Sara Campbell, Pam Denton, Sara House, Nicole Larson, Katie Livingood, Jaclyn Niebergall, and Melissa Young observed the classrooms. The authors would like to thank them all. Finally, the authors greatly appreciate the access which the Tulsa Public Schools and the CAP of Tulsa County Head Start program provided to their administrative data, their classrooms, and their students.

PRESCHOOL PROGRAMS CAN BOOST SCHOOL READINESS

Early childhood programs have been heralded by both scholars and decision-makers as an extremely promising avenue for fostering school readiness (1–4). In the U.S., these sentiments have fueled state initiatives and enrollments in pre-K programs across the country. They have also affected recent federal legislative efforts to improve the quality and reach of Head Start, including provisions to foster greater collaboration between pre-K and Head Start programs at the state level. Head Start and most state pre-K programs are targeted to poor or otherwise at-risk children, but a growing number of pre-K programs are universally available to all four-year-olds, irrespective of income. The number of four-year-olds enrolled in state pre-K programs actually now exceeds the number enrolled in Head Start (5).

Oklahoma’s pre-K program has generated considerable attention, partly because it is universal, based in the school system, and reaches a higher percentage of four-year-olds than any other state pre-K program. Importantly, it has a history of collaboration with local Head Start programs. Tulsa has attracted particular interest because its school system is the largest in Oklahoma and because it boasts unusually strong business support for early childhood education (6).

In a series of papers, we have reported on the effectiveness of Tulsa’s universal pre-K program in promoting kindergarteners’ school readiness. We demonstrated that the program produces impressive short-term learning gains in pre-reading, pre-writing, and pre-math skills. While these gains are greatest for disadvantaged children, we have also demonstrated that they apply to children from diverse racial, ethnic, and socio-economic backgrounds (7–9).

In this paper, we extend our findings by estimating the short-term test score gains for children in Tulsa's pre-K program and for children in Tulsa's Head Start program, using newly-gathered data from August 2006. An evaluation of these two large-scale programs in America's heartland may help to illustrate the potential of early childhood programs to enhance school readiness elsewhere.

Oklahoma's state-funded pre-K program, which originated in 1990, became universal in 1998. As determined by the state legislature, the program channels aid to local school districts, which are free to run full-day programs, half-day programs, or both. Although the most common option is for public schools to provide services themselves, Head Start programs and day care centers are also eligible to participate if they establish "collaborative" relationships with their local school district (10). Under these collaborative relationships, the participating program receives 85 to 90 percent of the state subsidy, while the school district receives the remaining 10 to 15 percent.

The Oklahoma pre-K program has relatively high standards compared to other states and offers relatively high pay and benefits to well-qualified teachers. Every lead teacher in every pre-K classroom 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. Of critical importance, pre-K teachers are paid the same wages as other public school teachers. These same rules are supposed to apply to collaborative programs, such as Head Start, although the state Department of Education allowed them a grace period of several years to reach the same standards as the public schools.

In 2005, the Community Action Program (CAP) of Tulsa County, whose Head Start program serves the largest number of children in Tulsa, achieved its goal of having

all of its four-year-olds in “collaborative” programs linked to the Tulsa Public Schools (TPS) and therefore eligible for state funding. As of the 2005-06 academic year, all CAP Head Start four-year-old program teachers, like their TPS counterparts, had a B.A. and were early childhood certified. CAP Head Start teachers were also compensated at rates virtually identical to those of TPS public school teachers (11). In these respects, the TPS pre-K program and the CAP Head Start four-year-old program were indistinguishable.

In August 2006, we administered three subtests of the Woodcock-Johnson Achievement Test to incoming pre-K students, incoming Head Start students, and incoming kindergarten students in Tulsa. The tests were administered just prior to the commencement of classes by the individual who would be teaching the child that year. Teachers were trained to administer the test and were compensated for their participation. Approximately 78 percent of all pre-K entrants, 69 percent of all Head Start entrants, and 73 percent 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 (12). While the child was being tested, the parent completed a two-page survey to provide demographic information on the child. Parent survey response rates ranged from 84 percent for kindergarten entrants to 88 percent for pre-K entrants and Head Start entrants. Program participation and school lunch eligibility were determined from administrative data (13).

The three subtests we used were: the Letter-Word Identification Test (a measure of pre-reading skills); the Spelling Test (a measure of pre-writing skills); and the Applied Problems Test (a measure of pre-math skills). These subtests have been shown to be especially appropriate for relatively young children (ages 4-5) and have been used by

other researchers studying both at-risk and more diverse groups of children (14-16). Woodcock-Johnson test scores at or prior to school entry help to predict later scholastic achievement (17).

The most 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) program will differ, in observable and unobservable characteristics, from those whose parents do not choose for them to participate, and that these differences will affect test scores. To overcome such selection bias, we have employed two separate regression-discontinuity estimations. For the first estimation, the treatment group consists of 1,264 kindergarten students who attended Tulsa's pre-K program and the comparison group consists of 1,492 children who are about to begin that program. For the second estimation, the treatment group consists of 327 kindergarten students who attended Tulsa's Head Start program, and the comparison group consists of 483 children who are about to begin that program. A small number of children whose birthdays fell outside the prescribed range were excluded from the analysis.

This research design is possible because the Tulsa Public Schools and the Tulsa Head Start program strictly enforce a September 1 birthday requirement for enrollment in the four-year-old program. Thus a child born on September 1, 2001 was eligible to enroll in the 05-06 school year, while a child born on September 2, 2001 was not. The thought experiment is to compare the test scores of the former child – who just received the Program X treatment (either pre-K or Head Start) to the test scores of the latter child – who is just about to receive the treatment of Program X. The two children are virtually

the same age, should have no discontinuous differences in their characteristics that influence test scores, and have both selected into treatment of Program X.

Our analysis estimates the continuous relationship 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 (*Qualify*, a measure of the number of days born before/after the cutoff qualification date), an age cut-off indicator variable (*Cutoff*), and an interaction variable (*Cutoff X Qualify*) that allows for different slopes on both sides of the cut-off point. The coefficient for *Cutoff* is the estimated treatment effect, and the constant captures the test score of a student who did not receive the treatment (*Cutoff*) and whose value for each of the other variables in the model equals zero. The 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 Tulsa Public Schools 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 program results as causal impacts.

In all of our regressions, we also included a wide range of demographic variables, including gender, race/ethnicity, school lunch eligibility, mother's education, whether the child lives with his/her biological father, and whether the child has Internet access at

home (18). We handle missing data, for three variables obtained through the parent survey, by using multiple imputation (19). All of our numbers estimate the effects of treatment on the treated (TOT).

As Table S1 indicates, the Tulsa Public Schools 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 I.D., 0.743 for Spelling, and 0.355 for Applied Problems (20). These effect sizes substantially exceed those reported for pre-K programs generally and they are somewhat greater than those reported for five states with relatively high-quality pre-K programs (21, 16). They do not quite reach the short-term effect sizes reported for the Perry Preschool Study (between .77 and 1.16 after two years of preschool) and the Abecedarian Project (on average, 1.08, after three years of preschool) (22, 23). As Table S2 indicates, the effects of the Tulsa Head Start program, though less spectacular, are also impressive: 0.514 for Letter-Word I.D., 0.334 for Spelling, and 0.369 for Applied Problems. These effect sizes exceed those reported for a national random assignment study of Head Start (24, 25). (26)

Another way to express program effects is to convert test score impacts into monthly equivalents, using conversion charts available from the creators of the Woodcock-Johnson Achievement Test. With this method, we find that the TPS pre-K program boosts pre-reading skills by nine months, pre-writing skills by eight months, and pre-math skills by five months. The Tulsa Head Start program boosts pre-reading skills by five months, pre-writing skills by three months, and pre-math skills by five months (see Figure S1). 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 free-lunch-eligible children (typically children whose family income is less than 130 percent of the federal poverty level). We then trim the pre-K sample to include only free-lunch-eligible children who participated in a full-day program (all Head Start children participated in a full-day program) (27). Even with these modifications, the two samples differ in the race/ethnicity of the students, though 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.

One concern when comparing TPS and Head Start students is that the free lunch children in the Head Start sample might still be poorer than the free lunch children in the TPS sample. Under program rules, 90 percent or more of Head Start participants are supposed to have incomes below 100 percent of the federal poverty level (FPL). To qualify for a free lunch, however, students only need to have incomes below 135 percent of the FPL. Unfortunately, we do not have income data for the TPS children. We do know, however, that 11 to 12 percent of the Tulsa Head Start children have incomes between 100 and 135 percent of the FPL. Thus there are some near-poor children in the Head Start free lunch group, which should reduce any discrepancies between the two groups. Our decision to focus on full-day program participants should also help to reduce such discrepancies because full-day pre-K programs in Tulsa are much more likely than

half-day pre-K programs to be located in the poorest neighborhoods, where Title I funding can be used to augment state funding and provide a full day of services.

Another way to think about all this is to consider the test scores of the two reconfigured groups at program entry (August 2006). The pre-reading and pre-writing test scores of these groups at program entry, after focusing only on free-lunch-eligible, full-day program participants, are fairly similar. Tulsa pre-K program entrants who are free-lunch-eligible and enrolled in a full-day program have an average Letter-Word I.D. score of 3.58 and an average Spelling score of 4.14, while Tulsa Head Start program entrants who are free-lunch-eligible and enrolled in a full-day program have an average Letter-Word I.D. score of 3.79 and a Spelling score of 4.60 (28). In short, the pre-reading and pre-writing scores for Head Start program entrants are slightly higher than those of the pre-K program entrants (29). In contrast, a larger gap in pre-math scores runs in the other direction. Tulsa pre-K program entrants who are free-lunch-eligible and enrolled in a full-day program have an average Applied Problems score of 7.18, while Tulsa Head Start program entrants who are free-lunch-eligible and enrolled in a full-day program have an average Applied Problems score of 5.98 (30). One way to think about this is that the Head Start contingent, with somewhat lower pre-math scores, are poised to benefit more from whatever math instruction the Head Start program may provide to four-year-olds, which could influence the comparison of mathematics test impacts between TPS and Head Start.

Even though the selection criteria might still differ across the two programs, focusing on the pre-K population that most resembles the population of Head Start students allows us to better compare the two programs. As Figure S2 indicates, there are

larger test impacts for free-lunch children in the TPS pre-K full-day program compared to free-lunch children in the Tulsa Head Start program (a full-day program), for both blacks and Hispanics. The difference is larger for pre-reading and pre-writing skills than for pre-math skills, where the TPS and Head Start programs are equally effective.

To gain a better understanding of what goes on inside TPS and Head Start classrooms, we deployed highly-trained student observers (eight seniors from the University of Tulsa) to monitor the morning sessions of all TPS pre-K and CAP Head Start four-year-old programs. The observers employed both the Classroom Assessment Scoring System (CLASS) (31) and the Emerging Academics Snapshot (32) protocols. The observers were trained by experts who have been closely involved in the development and use of the two respective protocols. Reliability was established by comparing our observer ratings with expert ratings through the use of classroom videotapes. For CLASS, on average, 88 percent of the observers' responses were either exactly the same as or within 1 scale-point of the expert's responses. All observers achieved at least 81 percent reliability. For the Snapshot, on average, 92 percent of the observers' responses were exactly the same as the expert's responses. All observers achieved at least 90 percent reliability. Ultimately, we were able to observe 77 of 79 pre-K classrooms that met in the morning and 28 of 28 Head Start classrooms, in the spring of 2006.

As reported elsewhere, the TPS pre-K program scored higher on classroom quality on a range of indicators than a multi-state sample of school-based pre-K programs. Similarly, the CAP Head Start program scored higher on classroom quality on a range of indicators than a multi-state sample of Head Start programs (33). When we

compared these two programs within Tulsa, we found similarities in the time allocation practices of the two programs with the exception of three measures. TPS pre-K devotes more time to practicing letters and sounds, more time to mathematics, and less time to fantasy play than Head Start (see Figure S3).

Both the TPS pre-K program and the CAP of Tulsa County Head Start program produce substantial improvements in pre-reading, pre-writing, and pre-math skills. These improvements exceed those reported in multi-state studies of both types of programs (22, 25). The relative efficacy of Tulsa's early childhood programs, as compared to programs elsewhere, could be due to several facts: first, all lead teachers in both programs have a B.A. degree; second, all lead teachers in both programs are early childhood certified; third, all lead teachers in both programs are paid a public school wage; and fourth, all early childhood classrooms in both programs have student-teacher ratios of 10/1 or less and class sizes of 20 or fewer.

Yet, despite striking similarities, the TPS pre-K program is more efficacious than the CAP of Tulsa County Head Start program in enhancing cognitive development. Why should this be the case? 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. Differences in mathematics emphasis do not appear to translate into differences in learning gains for the two programs. This could be because, as noted earlier, the pre-math gap is wider for TPS and Head Start program entrants than either the pre-reading or pre-writing gap. Thus, with a lower starting point, Head Start

participants may benefit more from a given amount of math instruction than TPS pre-K participants do (34). However, an analysis of kindergarten students who attended the two types of programs the previous year suggests that two classroom variables (emphasis on letters and sounds, emphasis on fantasy play) do help to explain part of the difference in test score gains between TPS and Head Start (35). 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 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. A third possibility is that TPS is better able to recruit and retain the best and the brightest teachers than Head Start, either because it offers more opportunities for professional development or because its classrooms may include a smaller proportion of high-risk children. A fourth possibility is that Head Start teachers direct more attention to socio-emotional development than TPS teachers do. However, on measures of emotional support from the CLASS instrument, we see no statistically significant advantage in favor of Head Start.

Early childhood education programs in the U.S. face some 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. Targeted pre-K programs face similar challenges. Even universally available programs, such as Oklahoma's, must cope with the realities of poor families, fragmented families, and immigrant families. In Tulsa, for

example, about three-fourths of pre-K program participants are eligible for a free or reduced-price lunch, and about three-fourths of the Hispanic pre-K program participants come from a home where Spanish is the primary language spoken.

Against this backdrop, it is instructive to compare the potency of program participation variables – either pre-K or Head Start – with that of other variables in our statistical models. For the Tulsa Public Schools model, program participation is a more powerful predictor of pre-reading and pre-writing test score outcomes than gender, race/ethnicity, free lunch eligibility, mother’s education, or whether the biological father lives at home (see Figure S4). For the Head Start model, program participation is a more powerful predictor of pre-math test scores than gender, free lunch eligibility, or whether the biological father lives at home (see Figure S5). In a world where so many children are at risk, it is comforting to know that early childhood education can make such a big difference for short-term test scores. The cumulative effects of family and environmental risk factors are daunting, but their negative impact can be muted substantially by participation in a high-quality early childhood education program (36).

Table S1: Cognitive Outcomes, TPS Pre-K Students

	Letter-Word (Woodcock- Johnson Test 1)	Spelling (Woodcock- Johnson Test 7)	Applied Problems (Woodcock- Johnson Test 10)
Born before cut-off (treated)	3.592***	2.121***	1.826***
	(0.338)	(0.217)	(0.338)
Qualify (days)	0.007***	0.008***	0.012***
	(0.001)	(0.001)	(0.001)
Qualify*Cut-off	-0.000	-0.002*	-0.003**
	(0.002)	(0.001)	(0.002)
Reduced-price lunch	-1.161***	-0.333*	-0.848***
	(0.319)	(0.189)	(0.296)
Free lunch	-1.795***	-0.641***	-1.329***
	(0.235)	(0.160)	(0.244)
Black	-0.125	-0.383***	-2.126***
	(0.209)	(0.146)	(0.217)
Hispanic	-1.470***	-0.164	-3.345***
	(0.234)	(0.155)	(0.272)
Native American	-0.291	0.007	-0.568*
	(0.285)	(0.181)	(0.298)
Asian	0.165	0.739	-2.452***
	(0.605)	(0.456)	(0.662)
Female	0.870***	1.184***	0.747***
	(0.159)	(0.104)	(0.168)
High school degree, Mother	0.187	0.174	0.430
	(0.256)	(0.201)	(0.325)
Some College, Mother	0.662**	0.445**	1.017***
	(0.266)	(0.209)	(0.332)
College Degree, Mother	1.599***	1.121***	2.124***
	(0.376)	(0.237)	(0.413)
Lives with father	0.183	0.177	0.052
	(0.210)	(0.124)	(0.192)
Internet access	0.920***	0.461***	0.829***
	(0.190)	(0.139)	(0.203)
Constant	5.781***	5.392***	11.223***
	(0.417)	(0.290)	(0.489)
Observations	2756	2652	2736
R-squared	0.42	0.50	0.42
Effect Size	0.985	0.743	0.355
Robust standard errors in parentheses			
* significant at 10%; ** significant at 5%; *** significant at 1%			
Effect size calculated by dividing coefficient by standard deviation of comparison group			

Table S2: Cognitive Outcomes, CAP Head Start Students

	Letter-Word (Woodcock- Johnson Test 1)	Spelling (Woodcock- Johnson Test 7)	Applied Problems (Woodcock- Johnson Test 10)
Born before cut-off (treated)	1.820***	0.978**	1.811***
	(0.600)	(0.437)	(0.690)
Qualify (days)	0.006***	0.009***	0.014***
	(0.002)	(0.001)	(0.002)
Qualify*Cut-off	-0.003	-0.005***	-0.007**
	(0.003)	(0.002)	(0.003)
Reduced-price lunch (CAP)	0.098	-0.605	-0.367
	(1.040)	(0.872)	(1.765)
Free lunch (CAP)	-0.104	0.228	-0.074
	(0.801)	(0.671)	(1.515)
Black	-0.307	0.055	-1.019*
	(0.440)	(0.365)	(0.556)
Hispanic	-1.531***	-0.143	-3.518***
	(0.458)	(0.357)	(0.567)
Native American	-1.740***	-0.349	-1.264
	(0.632)	(0.504)	(0.775)
Asian	1.274	1.774**	1.160
	(1.174)	(0.875)	(1.863)
Female	0.548**	0.958***	0.591*
	(0.259)	(0.197)	(0.312)
High school degree, Mother	0.489	-0.077	0.073
	(0.351)	(0.288)	(0.444)
Some College, Mother	1.488***	0.620*	0.586
	(0.428)	(0.367)	(0.548)
College Degree, Mother	3.160***	1.101	0.405
	(1.207)	(0.800)	(1.269)
Lives with father	0.241	0.613**	0.139
	(0.323)	(0.267)	(0.390)
Internet access	0.372	-0.068	0.916**
	(0.375)	(0.262)	(0.463)
Constant	4.471***	5.050***	9.749***
	(0.939)	(0.797)	(1.677)
Observations	810	776	800
R-squared	0.29	0.36	0.39
Effect Size	0.497	0.337	0.363
Robust standard errors in parentheses			
* significant at 10%; ** significant at 5%; *** significant at 1%			
Effect size calculated by dividing coefficient by standard deviation of comparison group			

Figure S1: Tulsa Public Schools (TPS) Pre-K and CAP of Tulsa County Head Start, Program Impacts in Monthly Equivalents

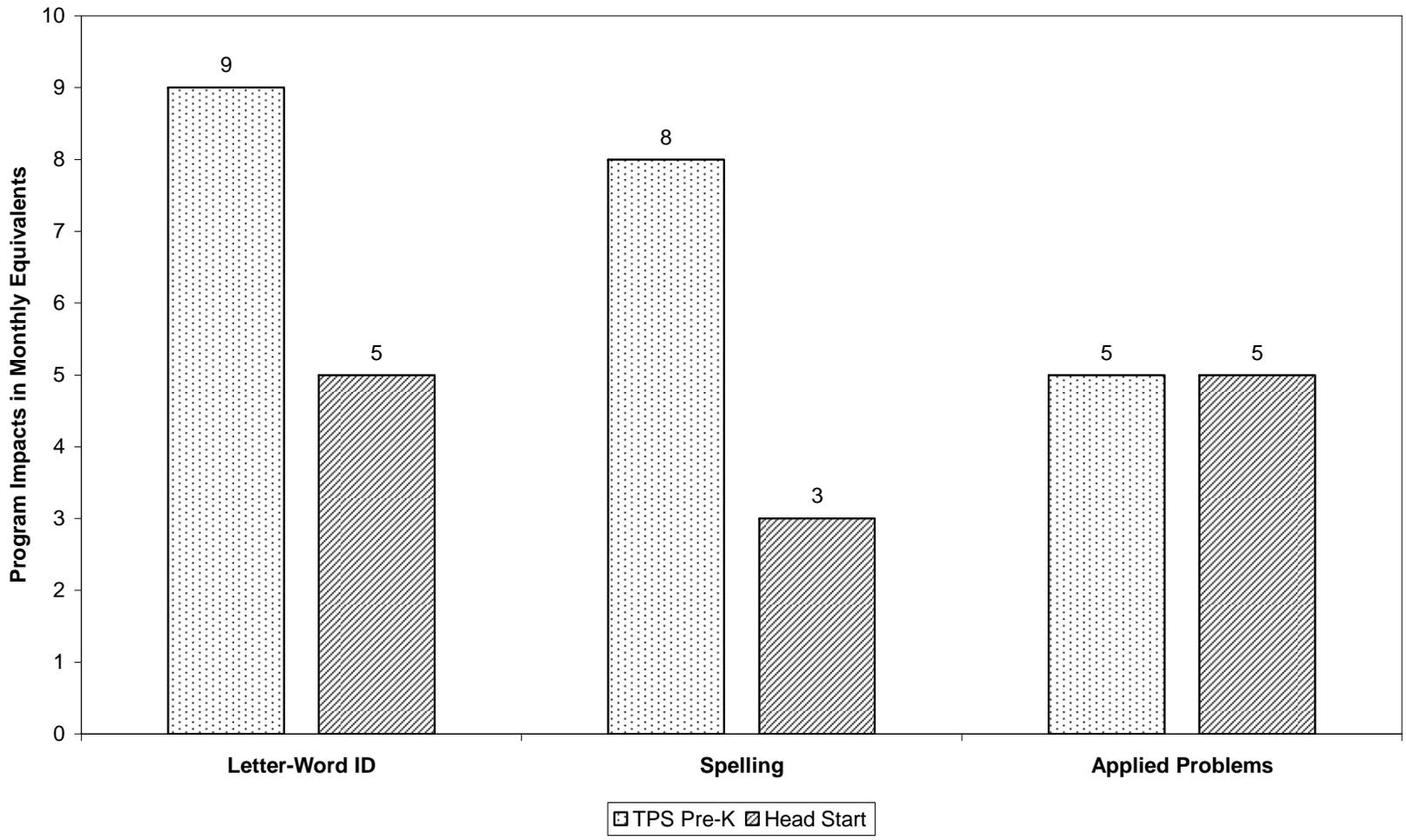


Figure S2: Effect Sizes for Free-Lunch-Eligible, Full-Day Students by Race, Tulsa Public Schools (TPS) Pre-K and CAP of Tulsa County Head Start

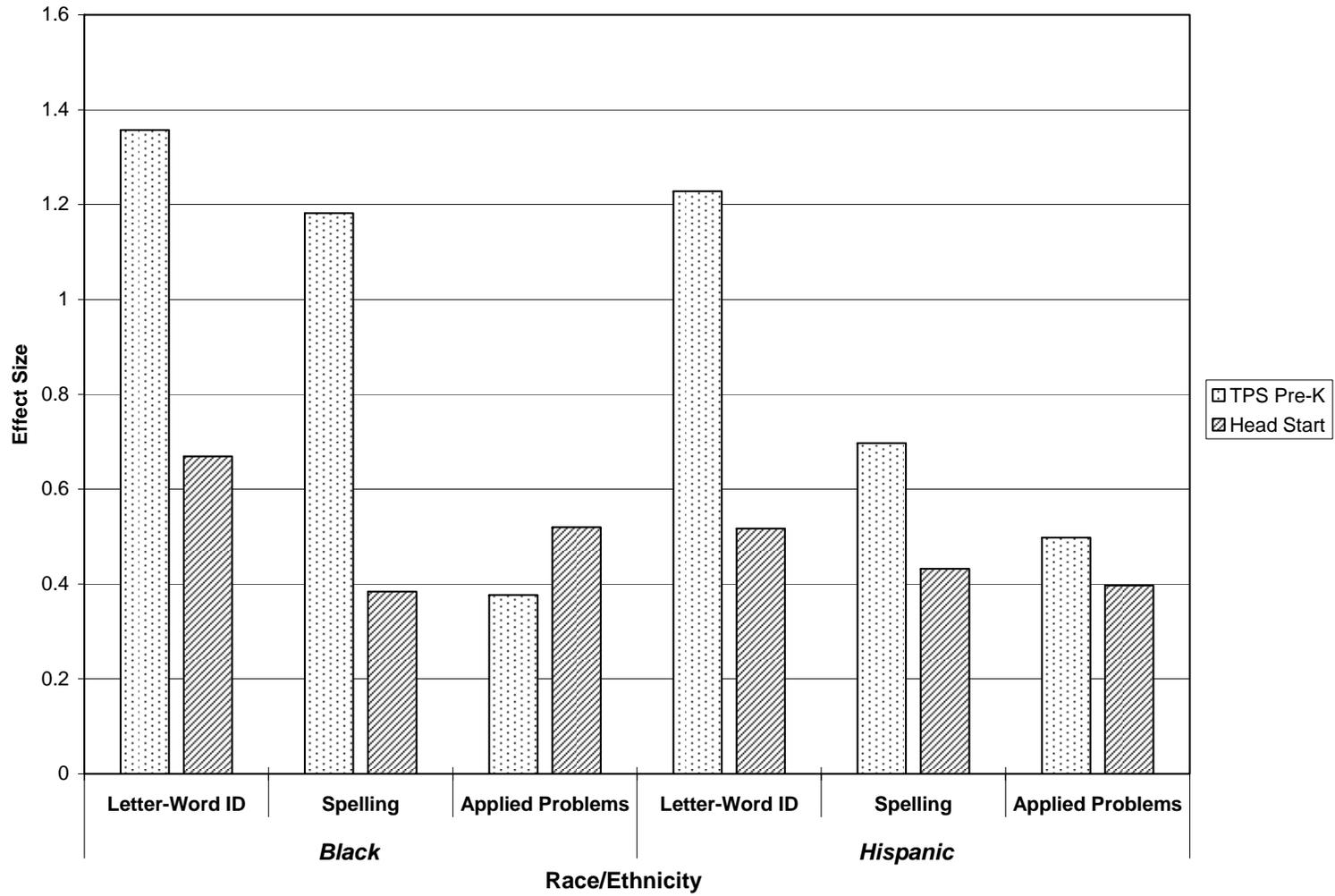


Figure S3: Differences in Classroom Time Allocation, Tulsa Public Schools (TPS) Pre-K and CAP of Tulsa County Head Start Teachers

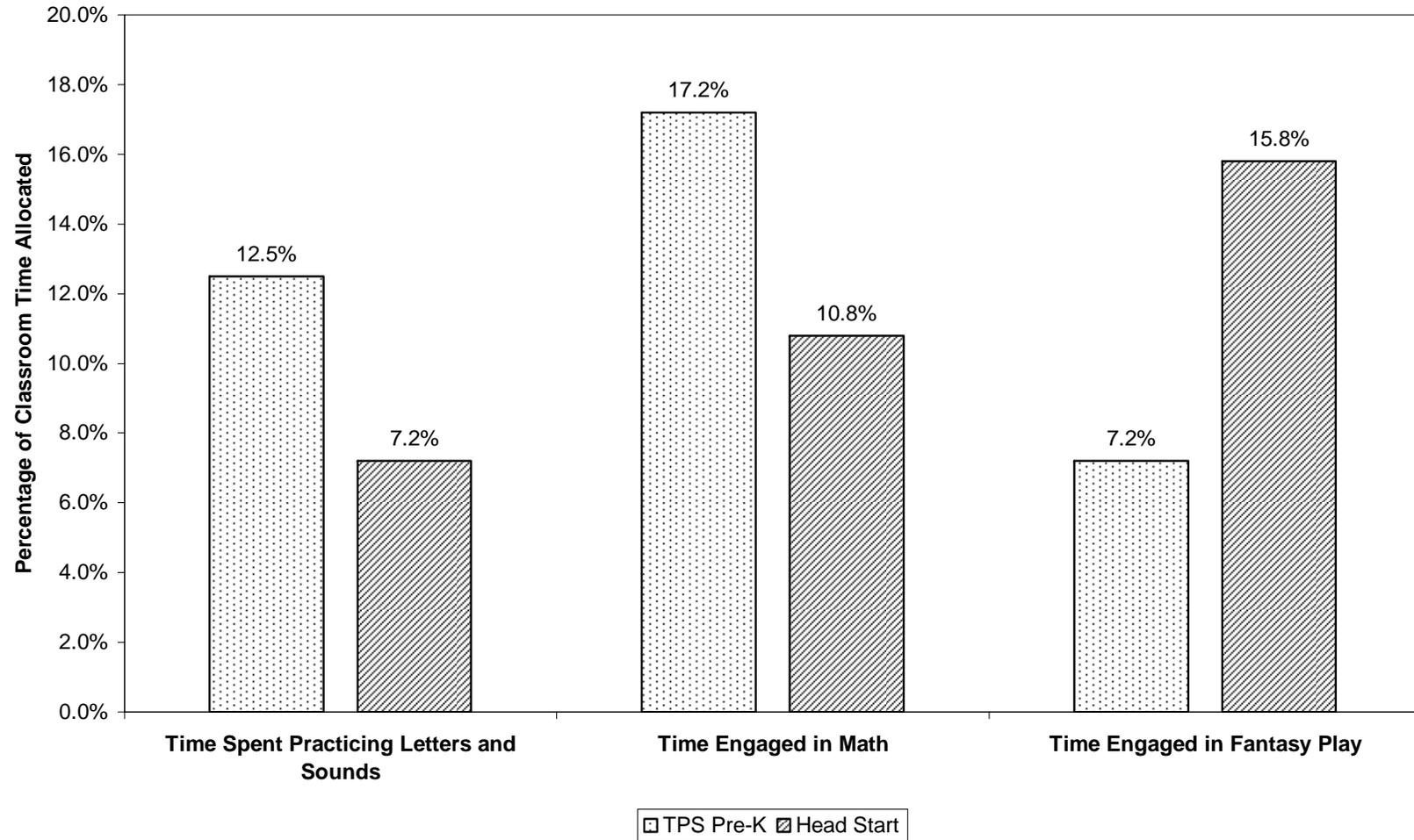
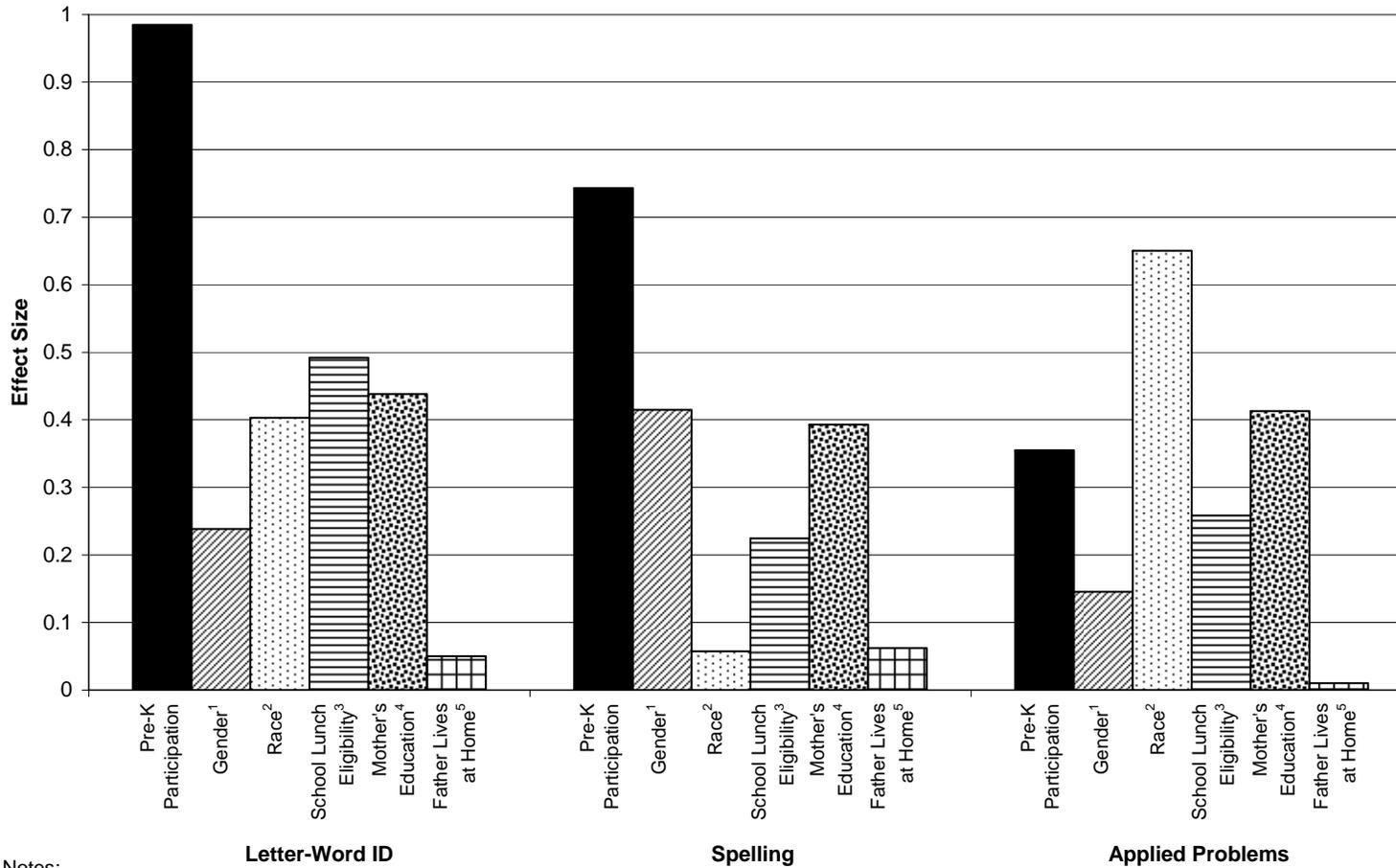


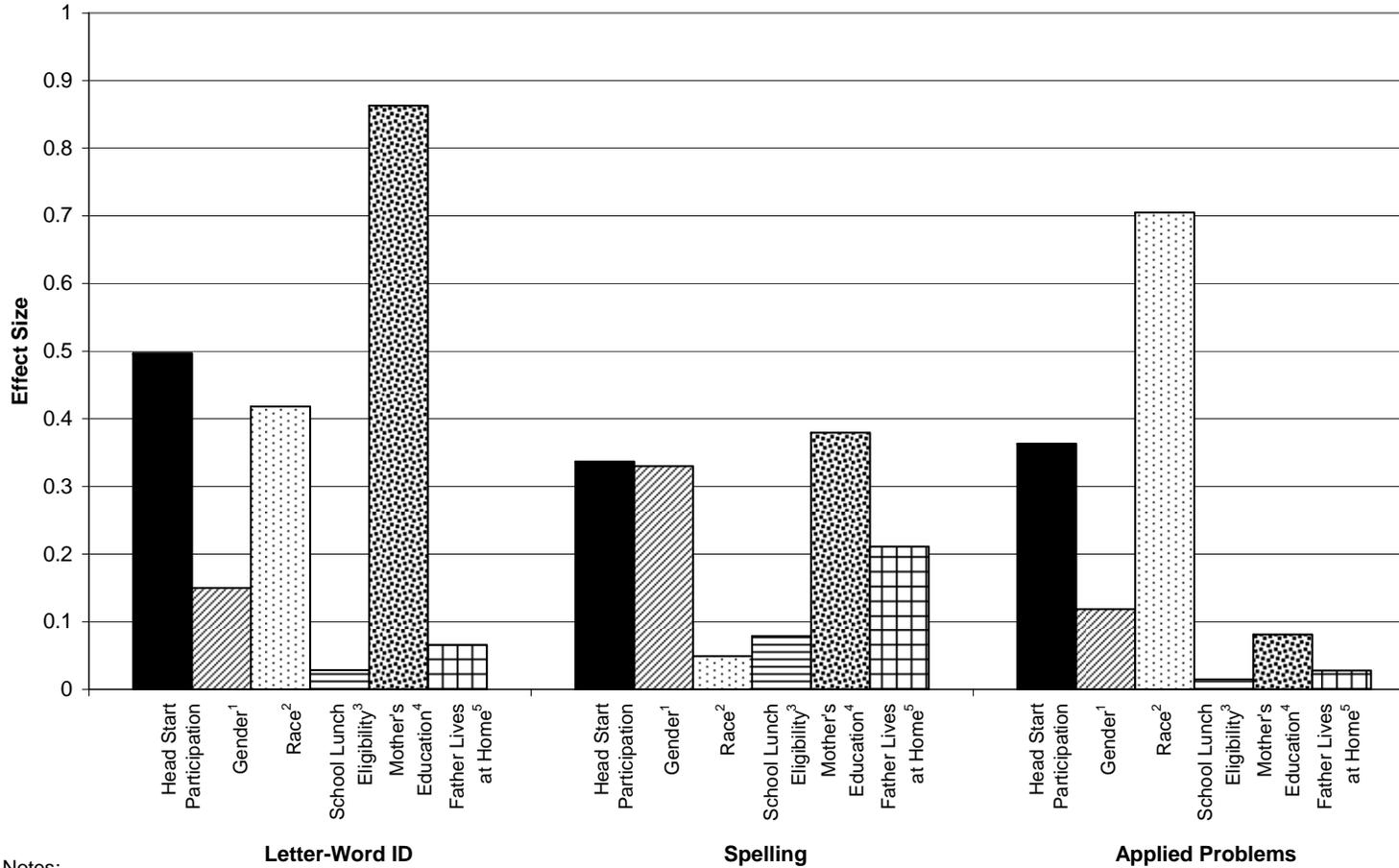
Figure S4: Relative Impacts of Tulsa Public Schools (TPS) Pre-K Participation, Demographic Characteristics



Notes:

1. Effect of female, with male as comparison category
2. Effect of Hispanic, with white as comparison category
3. Effect of free lunch eligibility, with paid lunch as comparison category
4. Effect of mother's college education, with no high school degree as comparison category
5. Effect of biological father living at home, with no as comparison category

Figure S5: Relative Impacts of CAP Head Start Participation, Demographic Characteristics



Notes:

1. Effect of female, with male as comparison category
2. Effect of Hispanic, with white as comparison category
3. Effect of free lunch eligibility, with paid lunch as comparison category
4. Effect of mother's college education, with no high school degree as comparison category
5. Effect of biological father living at home, with no as comparison category

References and Notes

1. W. S. Barnett, *Am. J. Orthopsychiatry* **63**, 500 (1993).
2. A. Reynolds, *J. Am. Med. Assoc.* **285**, 2339 (2001).
3. J. Heckman, *Science* **30**, 1900 (2006).
4. E. Zigler, W. Gilliam, S. Jones, *A Vision for Universal Preschool Education* (Cambridge University Press, New York, 2006).
5. W. S. Barnett, J. Hustedt, L. Hawkinson, K. Robin, *The State of Preschool 2006* (National Institute for Early Education Research, New Brunswick, N.J., 2006).
6. D. Leonhardt, "Bridging Gaps Early on in Oklahoma," *New York Times*, 7 February 2007, p. C1.
7. W. Gormley, T. Gayer, *J. Human Res.* **40**, 533 (2005).
8. W. Gormley, D. Phillips, *Policy Studies J.* **33**, 65 (2005).
9. W. Gormley, T. Gayer, D. Phillips, B. Dawson, "The Effects of Universal Pre-K on Cognitive Development," *Dev. Psychol.* **41**, 872 (2005).
10. The federal Head Start program permits such affiliations with local public schools, as it also permits a relatively wide range of decisions on curriculum, staffing, and the length of the school day. A collaborative relationship between a public school system and a Head Start program (or a day care center) simply means a formal agreement between both parties. Some collaborative agreements focus primarily or exclusively on how state revenue will be shared by the two parties. Other collaborative agreements are more substantive, as when a Head Start program agrees to enroll special education students who would otherwise have enrolled in a public school.
11. 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. More experienced teachers, of course, earn more, and salary step increases are strikingly similar for the two systems. Health and pension benefits are also similar.
12. Tested kindergarten students were somewhat less likely to be poor and black, somewhat more likely to be middle-class and white, than the universe of kindergarten students.
13. Pre-K participation was determined from TPS administrative data; Head Start participation was determined from CAP of Tulsa County Head Start administrative data.
14. P. L. Chase-Lansdale *et al.*, *Science* **299**, 1548 (2003).
15. G. Henry *et al.*, *Report of the Findings from the Early Childhood Study: 2001-02* (Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, 2003).
16. V. Wong, T. Cook, W. S. Barnett, K. Jung, *J. Policy Anal. Manag.* **27**, 122 (2008).
17. G. Duncan *et al.*, *Devel. Psych.* **43**, 1428 (2007).
18. 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. However, the coefficient estimates for these covariates might be of interest in and of themselves.
19. D. Rubin, *J. Am. Stat. Assoc.* **91**, 473 (1996).
20. 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.
21. K. Magnuson, C. Ruhm, J. Waldfogel, *Econ. Educ. Rev.* **78**, 33 (2007).
22. L. Schweinhart, J. Montie, Z. Xiang, W. S. Barnett, C. Belfield, M. Nores, *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40* (Ypsilanti, Mich.: High/Scope Press, 2005), 61.
23. C. Ramey, F. Campbell, M. Burchinal, M. Skinner, D. Gardner, S. Ramey., *Appl. Dev. Sci.* **4**, 2 (2000), 5.
24. M. Puma *et al.*, *Head Start Impact Study: First Year Findings* (Department of Health and Human Services, Washington, DC, 2005).
25. Because we have used treatment on the treated (TOT) estimates, it is fairest to compare the Tulsa results with TOT estimates for four-year-olds who participated in Head Start in the national study. These effect sizes are: .319 (Woodcock Johnson, Letter-Word I.D. Test), .239 (Woodcock-Johnson, Spelling Test), and .147 (Woodcock-Johnson, Applied Problems Test). For the TOT numbers, see: J. Ludwig, D.

Phillips, *The Benefits and Costs of Head Start*, NBER Working Paper 12973 (Cambridge, MA, March 2007), p. 38.

26. The positive test impacts could be underestimated if the pre-test scores were truncated at zero. However, only 5.0 percent of TPS pre-K entrants and only 14.5 percent of Head Start entrants received a zero score for the Letter-Word Identification test. For the Spelling test, the percentages are 4.7 percent for TPS pre-K and 8.9 percent for Head Start, and for the Applied Problems test, the percentages are 8.8 percent for TPS pre-K and 22.4 percent for Head Start. Thus, except for the least prepared of students, these tests capture differences in skill levels. The percentages are even lower for TPS pre-K alumni and Head Start alumni.

27. Approximately two-thirds of all TPS program participants enroll in a full-day program. Full-day programs could be more efficacious because they provide greater exposure to early learning activities. On the other hand, full-day programs produce stress in some children. In either case, it is useful to control for this variable.

28. The difference between the Head Start entrants and TPS pre-K entrants is not significant for Letter-Word Identification, significant at the .01 level for Spelling.

29. One reason for this may be that some of the Head Start program “entrants” may already have benefited from the Head Start program as three-year-olds.

30. The difference between the Head Start entrants and TPS pre-K entrants is significant at the .001 level for Applied Problems.

31. K. La Paro, R. Pianta, B. Hamre, M. Stuhlman, *The Classroom Assessment Scoring System (CLASS). Pre-K Version*. (Charlottesville, Va.: University of Virginia, 2002).

32. S. Ritchie, C. Howes, M. Kraft-Sayre, B. Weiser. *Emerging Academics Snapshot* (Los Angeles: University of California at Los Angeles, 2002).

33. D. Phillips, W. Gormley, A. Lowenstein, *Classroom Quality and Time Allocation in Tulsa’s Early Childhood Programs*, Paper delivered at the meeting of the Society for Research in Child Development, Boston, MA, March 30, 2007.

34. It is also worth noting that the Tulsa Head Start program devotes more time to math than the average Head Start program.

35. We used hierarchical linear modeling (HLM) to distinguish between organizational-level and individual-level effects on cognitive development. Individual-level (or level 1) variables included date of birth, gender, race/ethnicity, mother’s education, and school lunch eligibility. Organizational-level (or level 2) variables included a measure of dosage (half-day v. full-day) plus a Head Start dummy. The two classroom time allocation variables were included in one model but not the other.

36. For additional information on the Tulsa project, see <http://www.crocus.georgetown.edu>.