

Social-Emotional Effects of Early Childhood Education Programs in Tulsa

William T. Gormley, Jr.
and Deborah A. Phillips
Georgetown University

Katie Newmark
Cesar Chavez Prep School

Kate Welti
Child Trends, Inc.

Shirley Adelstein
Georgetown University

This article assesses the effects of Tulsa, Oklahoma's early childhood education programs on social-emotional outcomes, examining teacher ratings of children's behavior from the Adjustment Scales for Preschool Intervention and a measure of attentiveness using fixed effects regressions with propensity score matching. The sample includes 2,832 kindergarten students in 2006, of whom 1,318 participated in the Tulsa Public Schools (TPS) pre-K program and 363 participated in the CAP of Tulsa County Head Start program the previous year. Program participation was associated with lower timidity and higher attentiveness for TPS pre-K alumni and a marginally significant reduction in timidity for Head Start alumni. Results were similar for the free lunch-eligible subsample. We conclude that high-quality, school-based preschool programs can enhance social-emotional development.

In recent years, several studies have concluded that state-funded pre-K programs enhance the cognitive development of children. These studies, which use a variety of methods to assess pre-K in a variety of settings, have found positive impacts on prereading, prewriting, and premath skills in Georgia, Oklahoma, Michigan, South Carolina, New Jersey, and West Virginia (Barnett, Lamy, & Jung, 2005; Gormley, Gayer, Phillips, & Dawson, 2005; Henry et al., 2003; Wong, Cook, Barnett, & Jung, 2008).

In contrast to this increasingly rich body of studies on pre-K and cognitive development, far less is known about the effects of pre-K on social and emotional development. This is a serious gap in knowledge given rapid growth in pre-K enrollment

in the United States and substantial evidence that young children's social-emotional development sets the stage for subsequent social-emotional functioning and plays a role in later academic achievement. Just between 2002 and 2008, the percentage of our nation's 4-year-olds who were enrolled in state-funded pre-K programs rose from 14% to 24% (Barnett, Epstein, Friedman, Boyd, & Hustedt, 2008). For very low-income children, this trend can be traced back to the establishment of Head Start in 1965, a program that now enrolls over 10% of the nation's 4-year-olds.

This study examines the effects of Tulsa, Oklahoma's pre-K and Head Start programs on social-emotional outcomes at kindergarten entry. As such, it extends our prior work demonstrating substantial positive impacts of the Tulsa pre-K and Head Start programs on cognitive development, including prereading skills, prewriting skills, and premath skills (Gormley, Phillips, & Gayer, 2008). We focus on children who were enrolled in either the Tulsa Public Schools (TPS) pre-K program or the Community Action Project (CAP) of Tulsa County Head Start program during the year prior to kindergarten. Oklahoma's pre-K program has received national

This research was supported by grants from the Foundation for Child Development, the David and Lucile Packard Foundation, the Spencer Foundation, and the A.L. Mailman Family Foundation. We are grateful to the leaders and teachers of the Tulsa Public Schools and the Community Action Project of Tulsa County, without whom this research would not have been possible. We would also like to thank Carolyn Hill, Ronna Cook, Phil Fletcher, Camilla Heid, Katherine Magnuson, Rebecca Bulotsky-Shearer, Bill Jacoby, and Howard Bloom for helpful advice. Finally, we would like to thank Dan Cullinan, Amy Lowenstein, Catherine Shaw, Emily Holcombe, and Emily Page for their invaluable research assistance. The first two authors contributed equally to the writing of this article.

Correspondence concerning this article should be addressed to William T. Gormley Jr., Georgetown Public Policy Institute, 3520 Prospect Street, NW, Washington, DC 20007. Electronic mail may be sent to gormleyw@georgetown.edu.

© 2011 The Authors

Child Development © 2011 Society for Research in Child Development, Inc.

All rights reserved. 0009-3920/2011/8206-0027

DOI: 10.1111/j.1467-8624.2011.01648.x

attention because, as one of a handful of programs with universal eligibility, it reaches a higher percentage of 4-year-olds (71%) than any other program in the nation (Barnett et al., 2008). It also offers atypically high quality preschool education (Phillips, Gormley, & Lowenstein, 2009), perhaps because Oklahoma requires a lead teacher with a B.A. degree who is early childhood-certified in every classroom and pays these teachers regular school system wages. In Tulsa, the CAP Head Start program follows the same guidelines. As a result, this investigation may be seen as offering a "best case scenario" look at the potential contribution of high quality school-based pre-K and Head Start programs to children's social-emotional development.

Social-Emotional Development

Young children's social-emotional development captures a broad swath of specific outcomes, ranging from the ability to identify and understand one's own and others' feelings, establish and sustain relationships with both peers and adults, and regulate one's behavior, emotions, and thoughts (National Scientific Council on the Developing Child, 2005). The importance of these foundational capacities has been well documented. Having behavior problems in early childhood, for example, is associated with low peer acceptance, maladaptive teacher-child relationships, and antisocial disorders and delinquency in middle childhood and adolescence (Brody et al., 2003; Ladd & Burgess, 1999; Nagin & Tremblay, 2001; Shaw, Owens, Giovannelli, & Winslow, 2001; White, Moffitt, Earls, Robins, & Silva, 1990). Early childhood behavior that is more internalizing in nature, such as fearfulness or behavioral inhibition, is also associated with the development of serious anxiety problems in middle childhood and beyond (Fox et al., 2005; Schwartz, Wright, Shin, Kagan, & Rauch, 2003; Tincas, Benga, & Fox, 2006).

In light of teachers' concerns about social-emotional development and emerging evidence that both academic and social-emotional aspects of school performance may be affected by children's early social-emotional competencies, those who study early childhood programs have increasingly turned their attention to this domain of behavior. There is, in fact, a long-standing child-care literature in this area growing out of concerns that early nonmaternal care would undermine parents' role in fostering compliance and prosocial behavior (National Institute of Child Health and Human

Development Early Child Care Research Network [NICHD ECCRN], 1998). On the one hand, evidence has documented that children who spend more time in nonparental child care, especially center-based care, during the early childhood years display higher levels of externalizing and aggressive behavior, as well as more adult-child conflict, at 54 months and at kindergarten age and behavior problems through sixth grade (Belsky et al., 2007; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; NICHD ECCRN, 2003, 2005). Yet, other evidence suggests that when quality is high, spending more hours in nonmaternal care is not associated with increased behavior problems for low-income children and may actually reduce such problems (Loeb, Fuller, Kagan, & Carrol, 2004; Lowenstein & Phillips, 2010; Votruba-Drzal, Coley, & Chase-Lansdale, 2004). Moreover, exposure to group child care during the early childhood years appears to contribute to reduced levels of internalizing and inhibited behavior, perhaps especially if the quality of care is high and for children who are at risk of highly inhibited behavior (Crockenberg & Leerkes, 2005; Fox, Henderson, Rubin, Calkins, & Schmidt, 2001).

Research specifically on Early Head Start and Head Start, which predominantly serve very low-income children, has revealed both negative and positive impacts on social-emotional behavior. Negative effects on self-control, interpersonal skills, and increases in externalizing behavior have been found in secondary analyses of large, national data sets that rely on parental reports of where their preschoolers received care (Loeb et al., 2007; Magnuson, Ruhm, & Waldfogel, 2007). In contrast, positive immediate behavioral impacts have been reported from recent randomized trials of Head Start and Early Head Start (U.S. Department of Health and Human Services, 2005) and, in terms of long-term impacts on criminal behavior, from sibling comparison studies (Garces, Thomas, & Currie, 2002). Specifically, Head Start participation was related to modest reductions in parent-reported overall problem behaviors and hyperactivity for 3- but not for 4-year-olds. However, by the end of first grade, there were few significant impacts of Head Start participation (U.S. Department of Health and Human Services, 2010).

Others have turned to large, national data sets—notably the Early Childhood Longitudinal Study-Kindergarten cohort—to examine the social-emotional consequences of exposure to early childhood programs using ordinary least squares (OLS) regression, instrumental variable, and propensity score techniques. Collapsing children who were

reported by parents to be in a day-care center, pre-school, or pre-K program into a broad "center care" category, Loeb et al. (2007) found a positive association between center enrollment and behavior problems. Negative behavioral effects were greater for children who started center care at an earlier age and were greater for Whites than for Blacks and were nonsignificant for Hispanics. Magnuson et al. (2007) also found a positive association between pre-K participation, as defined by parents, and externalizing behavior. In addition, they found a negative relation between pre-K participation and self-control, although this relation disappeared when they switched from OLS regression to other estimating techniques. Magnuson et al. found similar effects when they limited their analyses to children from disadvantaged families. Interestingly, they found that social-emotional effects became nonsignificant if a child's pre-K program and kindergarten class were situated in the same school. It is important to note that because parents' retrospective reports on their children's early childhood program types are of unknown reliability (Lopez & Barrueco, 2005), firm conclusions from these studies about subsets of programs within the spectrum of early childhood options should be approached with caution.

Pre-Kindergarten Programs and Social-Emotional Development

Beyond the work of Magnuson et al. (2007) on pre-kindergarten programs, there is a dearth of research on social-emotional impacts specific to school-based pre-K experience. Reynolds (1989), who analyzed data from 1,539 ethnic minority children, some of whom had enrolled in Chicago Parent Child Center (PCC) preschools administered by the Chicago public schools since 1967, found that pre-school participation had no direct effects on social-emotional maturity at first grade but that it did have positive indirect effects, mediated through higher levels of parent involvement. Graduates from this program also had higher rates of high school completion and lower rates of official juvenile arrests (Reynolds, Temple, Robertson, & Mann, 2001).

In a recent study of pre-K in 11 states, based on a predominantly low-income sample of children, Howes et al. (2008) found that pre-K attendance was associated with a small increase in children's social skills between the fall and spring of the pre-K year, and a small decrease in behavior problems among children of mothers with low levels of education. Larger gains in social skills were associated

with teacher reports of warmer relationships with the child. The importance of the teacher-student relationship and of the emotional climate in the classroom as they relate to the development of social competence and the prevention of problem behaviors in preschoolers was also reported by Mashburn et al. (2008).

The implications of this extensive, but balkanized, literature on the social-emotional impacts of early childhood programs for newly emerging school-based pre-K programs for 4-year-olds are difficult to discern. Work on child care is sometimes discouraging, although not necessarily for low-income children, while experimental research on Head Start offers some glimmer of hope. Work on preschool programs is in its infancy, but it does suggest that when preschools are connected to elementary schools, negative effects are avoided, and when the classrooms are characterized by emotionally supportive teacher-child interactions, benefits accrue to social-emotional development. The Tulsa pre-K program is both school based and offers emotionally supportive educational environments (Phillips et al., 2009).

Research Questions

The current study addresses three questions: (a) What are the effects of school-based and Head Start-based preschool programs on social-emotional behavior at kindergarten entry? (b) Do these effects differ if we focus on poor children in particular? and (c) Do program impacts depend on the classroom micro-context (interactions with adults or peers, or during learning tasks) in which social-emotional behavior manifests itself? We address these questions in the context of the relatively high quality Tulsa pre-K program, using propensity score methodology and then teacher fixed effects to minimize selection bias in our comparisons between children who attended and did not attend these programs.

We hypothesize, based on the relative high quality of the Tulsa pre-K program and its strong links to elementary schools, that it will generate positive effects for children's social-emotional development. Moreover, the program's strong emphasis on academic instruction could help to prepare children for the educational challenges of kindergarten and thus reduce timidity, apathy, and acting out behaviors, although it is also plausible that excessive attention to academic instruction—or developmentally inappropriate practices—could generate or aggravate social-emotional distress.

We further hypothesize, based on the child care and Head Start literatures, that stronger links between Tulsa pre-K experience and social-emotional development will be found for low-income children. In this context, we do not anticipate differing outcomes for the TPS and Head Start programs. They follow the same quality guidelines, both offer emotionally supportive classroom environments, and, while the TPS classrooms place a greater emphasis on academic instruction (which may facilitate children's adjustment to the social challenges of schooling or may generate distress, as noted above), the Tulsa Head Start program, like Head Start programs elsewhere, is explicitly committed to social-emotional development as one of several program goals, whereas the TPS pre-K program is not. As such, it is not obvious that one program or the other is better equipped to foster social-emotional development.

Finally, because evidence regarding social-emotional impacts of pre-K education has not taken classroom micro-contexts into consideration, hypotheses regarding different impacts for teacher, peer, and task-oriented interactions are not warranted.

Method

Sample

A total of 186 TPS kindergarten teachers provided ratings of the social-emotional development of the children in their classrooms in October 2006. We received completed forms for 77% of the kindergarten students. As a result, the total sample consisted of 3,166 kindergarteners: 1,337 TPS alumni, 366 Head Start alumni, and 1,463 children who attended neither program (referred to as controls). Our final analytic sample declined slightly when we eliminated a few children who were too old or too young for their cohort and, more substantially, when we utilized propensity score matching, as discussed below (also see the Appendix).

Based on parent reports that included a large number of nonresponses (and, as such, do not offer systematic data on the experience of the control children), we know that some children in the control group were in another school district's pre-K program or another agency's Head Start program prior to entering kindergarten in Tulsa; some attended a day-care center or family day-care home; others remained at home. For our full sample of children assessed by their teachers, 82% of the TPS control group children and 79% of the Head Start control group children were in some

other kind of institutional care (a group day-care center, a family day-care home, or another preschool), according to their parents.

As Table 1 indicates, the assessed sample of children closely resembled the universe of children in most respects (there were some minor differences in free lunch eligibility). Table 2, discussed below, provides descriptive information on the matched TPS Pre-K and Head Start alumni, as well as the matched control children for both of these groups. Given that we conducted analyses not only on the total sample, but also on the children with free lunch eligibility, it is important to note that 65% of the TPS alumni and 89% of the Head Start alumni fell into this subgroup.

Measures

To analyze the effects on social-emotional development of the preschool programs in Tulsa run by the public schools and by Head Start, we chose the Adjustment Scales for Preschool Intervention (ASPI) instrument developed by researchers at the University of Pennsylvania (Lutz, Fantuzzo, &

Table 1
Comparison of Assessed Children and the Universe of Children, Kindergarten

Variable	Assessed (M)	Universe (M)
Female	0.48	0.47
<i>n</i>	3,166	4,114
White	0.36	0.35
Black	0.30	0.31
Hispanic	0.23	0.23
Native American	0.10	0.10
Asian	0.01	0.01
<i>n</i>	3,144	4,059
Free lunch	0.66*	0.69
Reduced-price lunch	0.11	0.10
Full-price lunch	0.23	0.22
<i>n</i>	3,142	4,061
No high school	0.19	0.20
High school	0.26	0.26
Some college	0.39	0.39
College degree	0.16	0.16
<i>n</i>	1,953	2,155
Lives with father	0.60	0.59
<i>n</i>	2,175	2,404
Internet at home	0.51	0.50
<i>n</i>	2,191	2,422

Note. Results are based on original data, exclusive of multiple imputation. Difference between assessed and universe significant at **p* < .05.

Table 2

Comparison of Observable Characteristics Among Matched Kindergarteners With and Without Exposure to TPS Pre-K and CAP Head Start

Variable	Tulsa Pre-K			CAP Head Start		
	Comparison	Treatment	Std. diff. (%)	Comparison	Treatment	Std. diff. (%)
Age ^a	185.70	180.94	-4.5	197.04	190.65	-6.0
Free lunch	.66	.66	-1.8	.91	.89	-2.9
Reduced-price lunch	.12	.12	0.8	.06	.06	0.0
Full-price lunch	.22	.22	1.4	.04	.05	3.3
White	.33	.36	5.2	.11	.11	0.7
Black	.37	.34	-6.0	.44	.41	-5.2
Hispanic	.21	.20	-2.8	.38	.40	5.4
Native American	.08	.09	4.4	.08	.07	-1.0
Asian	.01	.01	0.0			
Female	.47	.48	0.7	.48	.50	3.0
No high school	.22	.20	-4.8	.28	.26	-3.6
High school	.29	.27	-3.2	.29	.32	5.9
Some college	.40	.41	3.6	.33	.34	0.6
College degree	.10	.12	4.6	.09	.08	-4.4
Lives with father	.58	.59	2.8	.53	.54	1.2
Internet at home	.46	.49	6.6	.36	.34	-4.9

Note. None of the differences between matched comparison and treatment observations within each sample is statistically significant. Comparisons are from analysis of phenotypes and situtypes using the first impute for each sample. Difference of means tests were based on weighted samples to account for matching with replacement. The standardized difference for each covariate was calculated as the difference of the sample means in the matched treatment and comparison groups as a percentage of the square root of the average of the sample variances in the treatment and comparison groups (Rosenbaum & Rubin, 1985).

^aThis variable is a measure of the number of days born before or after the September 1, 2001, birthday cutoff enforced by the Tulsa Public Schools and Tulsa Head Start programs for enrollment in their 4-year-old programs in 2005–2006. A positive value for age indicates that a child was born before the cutoff (eligible for enrollment), while a negative value indicates that a child was born after the cutoff (ineligible for enrollment).

McDermott, 2002). The ASPI instrument consists of 144 statements describing behaviors that children may display. The assessor checks any descriptor that applies to the child being assessed. For example, with regard to the question, “Does this child pay attention in the classroom?” the assessor checks or does not check each of the following specific behavioral descriptors: generally listens well; talks, gazes around, plays with things; sits so quietly you don’t know if he or she is attending or not; lacks interest, “just sits”; and appears to live in a dream world. As these examples make clear, the descriptors capture both good behaviors and problem behaviors, although the majority of the items do focus on problem behaviors. In addition, the rated behaviors are organized into subsets defined by classroom context (or “situtypes”): interactions with the teachers, interactions with peers, and task-related behavior. With regard to the child’s relationship with the teacher, items assess, for example, how the child greets the teacher, seeks his or her help, and answers teacher questions. For peer relationships, items assess, for example, how well the child gets along with others his or her age and how

she or he handles peer conflicts. For task-related behavior, the items inquire about, for example, how well the child pays attention, copes with new learning tasks, and his or her extent of involvement in classroom activities. As a result, the ASPI ratings can be examined in two ways: first, as phenotypes of children across classroom contexts, and second, as behaviors within specific classroom micro-contexts or situtypes.

Adjustment Scales for Preschool Intervention was developed in close consultation with Head Start teachers, in an effort to ensure that it would be user friendly. Teacher ratings on the ASPI at the beginning of the Head Start year have been reported to differentiate children who display socially disruptive behavior at the end of the school year and to be associated with early math ability and general classroom competencies (Fantuzzo et al., 2007). For additional information on the construct validity of the ASPI, see Lutz et al. (2002) and Fantuzzo, Bulotsky, McDermott, Mosca, and Lutz (2003).

As a supplement to the ASPI, we also asked teachers to assess each child’s attentiveness by

completing a four-item Likert scale. The four items, which together constitute a self-regulation subscale, were extracted from an 18-item instrument known as the Instrumental Competence Scale for Young Children (Adler & Lange, 1997).

Factor Analysis

To assess the internal structure of our ASPI social-emotional assessments, we utilized common factor analysis (Snook & Gorsuch, 1989) based on all kindergarten children. We generated five factors, using orthogonal equamax rotation for the loadings: Disobedient (misbehaves and fails to follow rules), Aggressive (provokes other children, throws objects, fights), Attention Seeking (attempts to gain teacher's attention and impulsive), Apathetic (lacks energy and displays low classroom engagement), and Timid (shy toward teacher and displays low levels of participation). Factor scores were then calculated using factor weights. The five factors were then standardized to have a mean of 50 and a standard deviation of 10.

Although all five factors capture negative behaviors, it is important to note that the children—those who attended pre-K or Head Start and those who did not—received proportionately many more affirmative ratings (checks) from their kindergarten teachers for positive behavior than for problematic behavior. The vast majority of the children were obedient rather than disobedient, and engaged rather than apathetic, for example. Thus, if pre-school enrollment contributed to lower apathy ratings, for example, it would be most appropriate to interpret this as preventing apathy or enhancing engagement.

In addition to these five factors from the ASPI that seek to capture social-emotional dispositions (phenotypes), we identified three factors that seek to capture the classroom contexts in which social-emotional dispositions are manifested (situtypes). Although we used our own factor analysis strategy to identify such situtypes, the three factors we identified (Inappropriate Task-Related Behavior, Inappropriate Interactions With the Teacher, and Inappropriate Interactions With Peers) precisely parallel the three identified by the University of Pennsylvania researchers who conceived of this interesting strategy (Bulotsky-Shearer, Fantuzzo, & McDermott, 2008). As with the phenotypes, the situtypes capture what might be viewed as negative or dysfunctional behaviors in specific micro-contexts (e.g., teacher vs. peer interactions and academic vs. interpersonal situations).

Procedure

After receiving permission from the TPS administration and CAP of Tulsa County Head Start, we distributed the ASPI forms (and the supplementary attentiveness questions) to all kindergarten teachers. We asked the teachers to complete the forms for each child in her or his classroom during the week of October 2, 2006, approximately 40 days after the commencement of classes. We deliberately selected a time period that was early enough in the school year to approximate a pretest but late enough to ensure that the teacher knew each child well enough to fill out the form.

In addition to the social-emotional assessment data, we obtained demographic data for each child from TPS and Head Start administrative data and from a parent survey administered in August 2006, when the same children took a cognitive development test, as arranged by us. Administrative data yielded valuable information on the child's date of birth, gender, and race and ethnicity, as well as enrollment in the TPS pre-K or CAP Head Start program. The parent survey yielded additional information on the mother's education, the child's place of birth, the parent's place of birth, the primary language spoken at home, Internet access at home, whether the child's biological father lives at home, and other variables.

Analytic Approach

We used propensity score matching and a weighted teacher fixed effects model to assess the impact of the TPS pre-K program and the Tulsa Head Start program on social-emotional development—both phenotypes and situtypes—for both the total sample of children and the subset who were eligible for free lunch. Propensity score matching has been widely used by scholars assessing a variety of questions, including the effects of a job training program (Dehejia & Wahba, 1999), the effects of neighborhood poverty on dropping out of school and teenage pregnancy (Harding, 2003), and the effects of substance abuse programs (Guo, Barth, & Gibbons, 2006). As noted above, it has also been used to assess the developmental effects of pre-K (Magnuson et al., 2007).

The motivation for propensity score matching is to identify a control group that resembles the treatment group as much as possible on observable characteristics, in order to minimize selection bias. Members of the treatment and control groups are matched based on having a similar likelihood of

being in the treatment group, a measure known as the propensity score, which is estimated from a wide variety of observable characteristics. That is, treated individuals are compared to individuals who “look” like members of the treatment group, but who did not actually choose the treatment.

Before settling on propensity score matching to estimate program effects, we actively considered an alternative—a regression-discontinuity design (RDD) we have used to estimate the effects of TPS pre-K on cognitive development (Gormley et al., 2005). We rejected this alternative, which involves comparisons across cohorts, primarily because we feared that kindergarten teachers would expect a higher level of maturity than pre-K teachers, which could make for unfair comparisons across cohorts. Although propensity score matching has its drawbacks (Wilde & Hollister, 2007), it seems to be more appropriate when the comparison group is local and when short-term effects are being assessed (Bloom, Michalopoulos, Hill, & Lei, 2002). Both of these conditions applied to our Tulsa data.

The use of multivariate regression modeling with teacher fixed effects further strengthened our propensity score matching methodology by accounting for inexact matching and unobserved heterogeneity across classrooms. Because propensity score matching usually produces very similar, but not identical, treatment and comparison groups, analyzing the matched samples using regression models with controls for a variety of covariates helps minimize any bias due to inexact matching. Using matching as a preprocessing step prior to parametric analysis also reduces model dependency (Ho, Imai, King, & Stuart, 2007).

An additional concern is unobserved heterogeneity, or the possibility that unobserved differences between classrooms are not fully captured by the independent variables in the regressions. In other words, there may be something about a given teacher that made her students’ values for the dependent variable higher or lower on average than other classrooms, and a failure to account for this can induce omitted variable bias. For example, it is possible that teachers differed in their application of the ASPI instrument and or that TPS and Head Start alumni are not randomly sorted into kindergarten classrooms. As a result, classroom means of the dependent variables may vary. A teacher fixed effects model addresses this by including a dummy variable for each teacher (save one), thereby allowing the intercept to vary by classroom. These dummy variables absorb any between-classroom variation in the dependent variable, and the esti-

mated coefficients thus reflect individual-level effects.

To address missing data, we employed multiple imputation. Multiple imputation has been shown to perform better than other common methods of addressing missing data (Croy & Novins, 2005; Rubin, 1996; Sinharay, Stern, & Russell, 2001), and its superior performance has also been demonstrated specifically in the context of propensity score methods (Mattei, 2009). We implemented multiple imputation by creating five imputed data sets using Royston’s (2005) *ice* program in Stata. After applying our propensity score matching technique to each of the five data sets, we used the *mi*-combine command to produce our regression results. In this procedure, regressions were estimated separately for each of the matched samples and then combined to produce our final results. Final parameter estimates reflect averages across the five regression analyses, and standard errors were calculated following the rules developed in Rubin’s (1987) seminal work. For more on our application of multiple imputation and its consequences, sample sizes, and results, see a longer version of this article (<http://www.crocus.georgietown.edu>; see “working papers”).

To set up the propensity score matching for the TPS pre-K program participants, we used a wide variety of covariates (see working paper) to estimate a logit regression of the likelihood that TPS kindergarten students attended the TPS pre-K program the previous year. The estimated logit equation was then used to predict each individual’s propensity score, namely his or her probability of having attended pre-K. Next, we used Stata’s *PSMATCH2* command to match treatment and control individuals based on their propensity score values (Leuven & Sianesi, 2003). We employed one-to-one nearest neighbor matching with replacement within a caliper of .001 to achieve the best match. We followed the same procedures for kindergarten students who attended the CAP of Tulsa County Head Start program the previous year, except with a .005 caliper. We used a slightly different caliper for the two populations because it was more difficult to find suitable matches for the Head Start population and we wanted to avoid dropping a substantial portion of our treatment group. Matching was done within program auspice.

If the propensity score matching process works well, then members of the treatment group and the matched comparison group should have similar observable characteristics. For illustrative purposes, Table 2 provides differences for our first imputation

in the full sample analyses of phenotypes and situations, including a measure of the standardized difference between the matched treatment group and the matched control group. As Table 2 indicates, we were able to match demographically similar samples, as evidenced by the absence of any statistically significant imbalances and standardized differences well below 10%. Across five imputations, propensity score matching eliminated almost all significant differences between the matched treatment and comparison groups (see working paper for details).

Through propensity score matching, we designed very similar, although not identical, treatment, and control groups. To err on the side of caution, we ran teacher fixed effects regressions for the matched samples that controlled for several child-level covariates (gender, race, date of birth, free lunch eligibility, mother's education, whether the child lives with the biological father, and Internet access at home) rather than utilizing a simple difference in means between our matched treatment and control groups. Statistically significant *F* tests confirmed the presence of classroom-level variation in social-emotional outcomes, underscoring the importance of controlling for teacher fixed effects. To account for the fact that some comparison observations were matched more than once due to matching with replacement, we weighted our regressions such that treatment observations received a weight equal to one and comparison observations received a weight equal to the total times each was matched. Robust standard errors were adjusted to account for clustering by student.

Of course, the two treatment groups differed in their demographic characteristics. Most notably, Head Start alumni were poorer than TPS pre-K alumni, which could help to account for any differences in program impact. For this reason, we also conducted separate propensity score matching and regression estimates for free-lunch-eligible children only, which should make the TPS and Head Start samples more similar—albeit imperfectly—with regard to family income. These results are reported below, after our primary results.

Results

We first present results for the social-emotional outcomes (phenotypes) for all children, followed by results for the subsample of low-income children. Results for the classroom context analyses (situations) are then presented, followed by tests for program

differences in impacts (TPS vs. Head Start). Alpha was set at .05, and thus all results reported as significant are $p < .05$ or lower. Results referred to in the text as “marginally significant” do not meet this standard but are $p < .10$.

Social-Emotional Outcomes

As Table 3 indicates, children who participated in the TPS pre-K program were less timid than control group children who attended neither the TPS pre-K program nor Head Start. Pre-K and control group children did not differ in four other areas: disobedience, aggressiveness, attention seeking, and apathy. On the four-item attentiveness scale, children who participated in the TPS pre-K program were more attentive. The effects were quite modest, with effect sizes of 0.15 for timidity and 0.19 for attentiveness (see Table 3). In contrast, children who participated in the CAP of Tulsa County program did not differ from their peers who did not attend this program. However, there was a marginally significant reduction in timidity for these children.

Because the TPS pre-K program and the CAP of Tulsa County Head Start program serve rather different children (e.g., the Head Start children are poorer, more likely to be Black or Hispanic, and have more poorly educated mothers), we reran the analyses on samples restricted to low-income children as proxied by free lunch status. This restricted sample was still characterized by significant differences in race ethnicity (more Black and Hispanic children in Head Start), but not in free lunch status or maternal education. For low-income TPS pre-K alumni, there was a marginally significant reduction in timidity and a significant increase in attentiveness (see Table 3). For Head Start alumni, there were no significant differences. It should be noted that our working sample size for TPS declined considerably when we switched from all children to free-lunch-eligible children only, but that the TPS free-lunch-eligible sample is still much bigger than the Head Start free-lunch-eligible sample.

Situation Findings

As noted earlier, we also analyzed the same data by focusing on the classroom micro-contexts in which the children's behaviors were manifested (situations). TPS pre-K participation was associated with a marginally significant reduction in teacher interaction problems (see Table 4). Head Start participation was associated with no significant differences. When the sample of children was restricted

Table 3

Effects on Social-emotional Development of Participation in the TPS Pre-K and CAP Head Start Programs, Matched Full and Free Lunch Samples

Social-emotional factor	Tulsa Pre-K				CAP Head Start			
	Full sample		Free lunch sample		Full sample		Free lunch sample	
	<i>B</i> (SE)	Effect size	<i>B</i> (SE)	Effect size	<i>B</i> (SE)	Effect size	<i>B</i> (SE)	Effect size
Disobedient	0.16 (0.72)	0.02	-0.65 (0.92)	-0.07	1.51 (1.27)	0.15	1.32 (1.42)	0.13
Aggressive	0.41 (0.92)	0.04	0.74 (1.02)	0.07	0.24 (1.30)	0.02	0.15 (1.76)	0.02
Attention seeking	-0.92 (0.65)	-0.09	-0.82 (0.80)	-0.08	-0.50 (1.31)	-0.05	-1.07 (1.38)	-0.11
Apathetic	-0.92 (0.69)	-0.09	-1.06 (1.05)	-0.11	-0.13 (1.05)	-0.01	-1.97 (1.50)	-0.20
Timid	-1.55* (0.68)	-0.15	-2.09 (1.21)	-0.21	-2.72 (1.46)	-0.27	-1.67 (1.60)	-0.17
Treatment cases	1,316		850		363		322	
% matched	87%		80%		93%		94%	
Unique control cases	587		353		216		191	
Attentiveness index	0.15** (0.05)	0.19	0.19* (0.09)	0.26	0.04 (0.11)	0.06	0.10 (0.09)	0.12
Treatment cases	1,305		840		357		317	
% matched	85%		80%		93%		94%	
Unique control cases	564		347		214		188	

Note. Results are from ordinary least squares regressions with teacher fixed effects. Models controlled for race/ethnicity, gender, mother's education, free lunch status, whether the child lives with his or her father, and whether the child has home Internet access. Observations were weighted to account for matching with replacement, such that each treatment observation received a weight equal to 1 and each comparison observation received a weight equal to the number of times matched. Robust standard errors were adjusted for clustering by student. Reported sample sizes are from the first imputed data set in each analysis.

* $p < .05$. ** $p < .01$.

to those of low income (free lunch status), we found no significant differences for either TPS pre-K or Head Start (see Table 4).

Comparison of Program Effects

Although we deliberately analyzed the TPS pre-K and CAP Head Start samples separately, we were interested in knowing whether the programs differed, in statistically significant terms, in their social-emotional impacts. To determine whether program impacts estimated using separate models were different, we calculated z scores following a method recommended by Paternoster, Brame, Mazerolle, and Piquero (1998). For the full sample and for the free lunch sample, we found no statistically significant differences.

Statistical Power Analysis

We have noted that the absence of statistically significant findings, for Head Start in particular, could be due to the relatively small sample size for this population. Our matched Head Start sample size was 674 (337 treatment cases plus 337 control group cases). Because we employed a teacher fixed effects model, the number of predictors in our model was fairly large (17 covariates plus 175 tea-

cher dummies, for a total of 192 predictors). If we assume an alpha level of .05, then a statistical power analysis yields observed power of 1.00, which exceeds the standard threshold of 0.80. Arguably, this makes it possible to detect fairly subtle effects (e.g., effect sizes of 0.10). However, if we had a larger sample or if we did not use a fixed effects model, it would have been easier to discern statistical significance.

Discussion

Our findings indicate that high-quality, school-based pre-K programs can support the development of some social-emotional skills that enable children to enter kindergarten ready to learn. In response to the concern that children's social-emotional skills may be slighted at the expense of their cognitive skills when they attend school-based pre-K programs, our findings clearly demonstrate that such trade-offs need not occur. The children in the TPS pre-K and Head Start programs failed to demonstrate the increases in aggressive and disobedient behavior seen in the child-care research literature, and those who attended TPS pre-K were portrayed by their kindergarten teachers as exhibiting less timidity, as well as higher levels of attentiveness,

Table 4

Effects on Situtypes of Participation in the TPS Pre-K and CAP Head Start Programs, Matched Full and Free Lunch Samples

Situtype	Tulsa Pre-K				CAP Head Start			
	Full sample		Free lunch sample		Full sample		Free lunch sample	
	<i>B</i> (<i>SE</i>)	Effect size	<i>B</i> (<i>SE</i>)	Effect size	<i>B</i> (<i>SE</i>)	Effect size	<i>B</i> (<i>SE</i>)	Effect size
Learning task problems	-0.60 (0.76)	-0.06	-1.35 (0.98)	-0.01	0.91 (1.22)	0.09	0.33 (1.36)	0.03
Teacher interaction problems	-1.18 (0.69)	-0.12	-1.64 (1.35)	-0.16	-2.09 (1.40)	-0.21	-2.17 (1.51)	-0.22
Peer interaction problems	-0.34 (0.68)	-0.03	-0.17 (1.10)	-0.02	0.35 (1.33)	0.03	-0.80 (1.62)	-0.08
Treatment cases	1,316		850		363		322	
% matched	87%		80%		93%		94%	
Unique control cases	587		353		216		191	

Note. None of the treatment effects is statistically significant. Results are from OLS regressions with teacher fixed effects. Models controlled for race/ethnicity, gender, mother's education, free lunch status, whether the child lives with his/her father, and whether the child has home Internet access. Observations were weighted to account for matching with replacement, such that each treatment observation received a weight equal to 1 and each comparison observation received a weight equal to the number of times matched. Robust standard errors were adjusted for clustering by student. Reported sample sizes are from the first imputed data set in each analysis.

than Tulsa 4-year-olds who did not experience this program. Effect sizes were 0.15 and 0.19, respectively—comparable in magnitude, but not necessarily direction, to those found in other efforts to assess social-emotional impacts of early childhood programs (see, e.g., Howes et al., 2008; Loeb et al., 2007; Magnuson et al., 2007). When the analyses were restricted to children who were eligible for free lunches, only the results for enhanced attentiveness remained significant. Head Start failed to show these positive impacts.

Importantly, the distribution of scores on these outcomes placed the majority of children in this study (both controls and pre-K alumni) within the outgoing (rather than timid) and attentive (rather than inattentive) range of teacher ratings. Thus, within the context of relatively well-behaved children, pre-K experience is most appropriately interpreted as both preventing negative social-emotional outcomes and fostering positive outcomes. Taking timidity as an example, 84.7% of TPS pre-K alumni appeared to “get along with two or more companions,” as opposed to 79.5% of our matched control group (see working paper for more examples). In effect, children who experienced TPS pre-K were socialized into the attentive and interactive role of kindergarten student to a greater extent than were children who did not experience this program.

The absence of a similar pattern of findings among the children who had experienced Head Start as 4-year-olds is somewhat surprising in light of prior evidence that classroom quality, as mea-

sured by the CLASS, in both TPS and Head Start classrooms in Tulsa exceeded that of similar pre-K classrooms in other states and did not differ from each other (see Phillips et al., 2009). However, it is worth noting that the Tulsa Head Start program does devote significantly less time to certain academic subjects (practicing letters and sounds, mathematics) than the Tulsa pre-K program. Because some of these differences give TPS pre-K alumni a cognitive advantage (Gormley et al., 2008), they may also give these same students a greater comfort level in kindergarten than Head Start alumni. This in turn could make it easier for TPS pre-K alumni to relate to teachers and peers in a school setting. Alternatively, it could be that the Head Start program's smaller sample size obscured some social-emotional effects.

Beyond differences in pedagogy, three other factors help to distinguish TPS and Head Start classrooms: (a) Unlike the TPS classrooms, only two of the Head Start classrooms were colocated with elementary schools, a situation that Magnuson et al. (2007) reported to foster positive social outcomes in their examination of preschool effects; (b) Head Start clearly served a more disadvantaged group of children and, even in the free lunch sample, a sample with a very different racial/ethnic profile as compared to TPS pre-K; and (c) Head Start programs were more likely to be full-day, a feature that, in some of the prior child-care literature, has been associated with more negative social developmental outcomes (see Phillips, McCartney, & Sussman, 2006). There is a pressing need to understand the

specific conditions under which Head Start helps or hinders children's social-emotional development.

Turning back to the TPS findings, their consistency with prior evidence of positive social-emotional effects of high-quality center-based child care, Head Start and Early Head Start, the Chicago PCC, and pre-K programs characterized by supportive teacher-child interactions and emotional climates (Howes et al., 2008; Mashburn et al., 2008; Reynolds et al., 2001; U.S. Department of Health and Human Services, 2005) is notable. They also run counter to prior evidence of detrimental social-emotional impacts of exposure to group care arrangements across the early childhood years (Belsky et al., 2007; Loeb et al., 2007; NICHD ECCRN, 2003, 2005). Many of the negative findings on socio-emotional impacts have focused on full-day, as opposed to half-day programs, and TPS sponsors both. Still, 66% of the TPS pre-K alumni in our kindergarten sample participated in a full-day pre-K program (approximately 6½ hr vs. 3¼ hr for half-day programs) the previous year. It is also important to note that some prior studies, from an earlier era, featured a comparison group that included a substantial number of children who stayed at home with their mothers. In contrast, our comparison group consisted primarily of children who were exposed to some other kind of early care and education program.

We also examined whether the effects of pre-K experience were more apparent in certain classroom micro-contexts than in others. There was virtually no evidence of such context effects in either TPS pre-K or Head Start in either the full or low-income samples. In light of prior evidence that the quality of the teacher-student relationship is among the strongest predictors of growth in children's social skills during the pre-K year (Howes et al., 2008; Mashburn et al., 2008), it is somewhat surprising that we did not find a significant association (beyond the marginal level) between pre-K experience and the children's behavior in the context of teacher-child interactions. It may be that the extensive group care experience of the control group also promoted their comfort and skill with teacher-child interactions.

These findings highlight the importance of expanding the typical range of social-emotional behaviors that are examined in research on early childhood programs. The early childhood field has paid much greater attention to externalizing behavior, particularly aggression and other behavior problems, than it has to either internalizing behavior, such as timidity, or regulatory behavior,

such as attentiveness. And yet, research that seeks to understand which social-emotional competencies affect school performance is increasingly converging on regulatory behavior, including executive functioning and attentional capacities (Blair, 2002; Diamond, Barnett, Thomas, & Munro, 2007; Duncan et al., 2007), and other aspects of behavior that promote participation and engagement in learning (Bierman et al., 2008; Domitrovich, Cortes, & Greenberg, 2007; Fantuzzo et al., 2003; Webster-Stratton, Reid, & Hammond, 2004). These are precisely the behaviors that were most consistently affected by the Tulsa pre-K experience as captured in our outcomes of timidity and attentiveness.

There are several limitations to the current study. First, our exclusive reliance on teacher ratings of the children's social-emotional outcomes was not ideal. In particular, it is difficult to know how individual biases of various kinds may have affected the ratings. We have controlled for this, in part, through the use of teacher fixed effects. The addition of more objective, observational measures of children's behavior would have strengthened our study still further. Second, because the teachers needed to be familiar with their students prior to making the ratings, as discussed earlier, we were not able to utilize the relatively rigorous RDD methodology used with our examination of cognitive outcomes (Gormley et al., 2008) when examining this domain of outcomes. This increases the possibility that the inferences we have drawn from this study are threatened by the potential omission of unobserved variables. We did, however, analyze the cognitive outcomes using our propensity score methodology and found reasonably similar results to RDD, lending credibility to our propensity score matching technique (results available upon request). Third, we have controlled for some but not all forms of teacher bias. If kindergarten teachers were, for some reason, biased against Head Start alumni, despite strong ties between TPS and Head Start, this could affect our results. Fourth, it is important to keep in mind that Tulsa's early childhood education programs are not typical of early childhood education programs across the country. As a result, caution is needed in generalizing our findings to preschool or early childhood programs more generally.

Conclusion

As the pressures on preschools to prepare children for school mount, there is growing recognition that this involves both exposure to preacademic

learning and support for the range of capacities that enable children to engage in learning, attend to instruction, interact effectively with teachers and peers, and manage themselves with growing independence in the classroom (Raver & Knitzer, 2002). This study demonstrates that state-funded pre-K programs, with a strong emphasis on academic content, can simultaneously support the development of emerging social-emotional competences. The stronger capacities to pay attention and the lower levels of timidity among the children who attended the TPS pre-K program suggest that these children are entering kindergarten better prepared to engage in learning than are their peers who did not attend pre-K.

It is evident that even without any clear focus on attention or interpersonal skills, the good instruction and supportive emotional environments that characterize Tulsa's school-based pre-K classrooms (Phillips et al., 2009) are producing important advancements in social-emotional development. At the same time, a number of early childhood interventions that focus explicitly on self-regulatory and attentional skills, and that attend to links between these behaviors and early learning, are now emerging and being refined for use with children who are making the transition from pre-K to kindergarten (see Bierman et al., 2008; Diamond et al., 2007; Raver et al., 2008, 2009). Several of these interventions have been launched in Head Start programs, where they may be especially needed. An interesting next step would involve extending this work to state-funded pre-K programs to see if, by adding an explicit focus on self-regulatory skills to the general pre-K curriculum, even stronger effects are found (see Bogard & Takanishi, 2005).

The evidence that high quality pre-K education reduced timidity for alumni is also important from a mental health perspective. The teacher ratings of timidity capture dimensions of a temperamental disposition known as social reticence (Fox et al., 2001; Rubin, Bukowski, & Parker, 2006), which defines a pattern of fearful, withdrawn behavior that is a risk factor for later anxiety disorders (Ladd & Burgess, 1999; Schwartz et al., 2003; Tincas et al., 2006). Prior research has reported that exposure to group care settings contributes to a reduction in social reticence over time (Fox et al., 2001). Our evidence suggests that exposure to peers in the context of emotionally supportive pre-K experiences may play a similar role for 4- and 5-year-olds. Whether these experiences are powerful enough to disrupt worrisome pathways from early inhibition and fearfulness to subsequent anxiety disorders remains

to be seen, but examinations of these types of pathways are highly deserving of empirical attention. This line of inquiry holds the potential to link what has been a highly educationally oriented pre-K literature to an equally large body of empirical research on childhood mental health (National Scientific Council on the Developing Child, 2008). It also fits well within a framework that increasingly views emotional and behavioral problems seen in young children as arising in the context of transactions between children and the contexts in which they spend their early years (Cicchetti & Sroufe, 2000; Fantuzzo et al., 2003).

It is difficult to generalize these findings from the relatively high quality Tulsa pre-K program to other areas of the country or to other types of early childhood settings, including Head Start. It may be the somewhat unusual features of the TPS pre-K program that account for the generally positive story told by our data. These include high teacher education requirements that are matched with relatively high salaries, colocation of the pre-K and elementary school classrooms, strong community support for the program, and, within the classrooms, unusually strong attention to classroom management, and high-quality feedback regarding learning. Some, but not all, of these features also characterize the Tulsa Head Start program. Clearly, we need to disentangle the programmatic features of the TPS pre-K program in order to determine which of them account for the program's success in enhancing children's social-emotional development. Pending that investigation, the Tulsa pre-K program continues to offer the country a promising example of how to foster both cognitive and social-emotional development as children embark on their critically important journey through school.

References

- Adler, F., & Lange, G. (1997). *Children's mastery orientations and school achievement in the elementary grades*. Poster presented at the biennial meetings of the Society for Research in Child Development, Washington, DC.
- Barnett, W. S., Epstein, D., Friedman, A., Boyd, J., & Hustedt, J. (2008). *The state of preschool 2008*. New Brunswick, NJ: National Institute for Early Education Research.
- Barnett, W. S., Lamy, C., & Jung, K. (2005, December). *The effects of state prekindergarten programs on young children's school readiness in five states*. New Brunswick, NJ: National Institute for Early Education Research.
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Tresch-Owen, M., et al. (2007).

- Are there long-term effects of early child care? *Child Development*, 78, 681–701.
- Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., et al. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, 79, 1802–1817.
- Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist*, 57, 111–127.
- Bloom, H., Michalopoulos, C., Hill, C., & Lei, Y. (2002, June). *Can nonexperimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* (MDRC Working Paper on Research Methodology). New York: Manpower Demonstration Research Corporation.
- Bogard, K., & Takanishi, R. (2005). PK-3: An aligned and coordinated approach to education for children 3 to 8 years old. *Social Policy Report*, 19, 1–23.
- Brody, L. M., Nagin, D. E., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developing trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, 39, 222–245.
- Bulotsky-Shearer, R., Fantuzzo, J., & McDermott, P. (2008). An investigation of classroom situational dimensions of emotional and behavioral adjustment and cognitive and social outcomes for Head Start children. *Developmental Psychology*, 44, 139–154.
- Cicchetti, D., & Sroufe, L. A. (2000). The past is prologue to the future: The times they've been a changing [Editorial]. *Development and Psychopathology*, 12, 255–264.
- Crockenberg, S. C., & Leerkes, E. M. (2005). Infant temperament moderates associations between childcare type and quantity and externalizing and internalizing behaviors at 2½ years. *Infant Behavior and Development*, 28, 20–35.
- Croy, C. D., & Novins, D. K. (2005). Methods for addressing missing data in psychiatric and developmental research. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 1230–1240.
- Dehejia, R., & Wahba, S. (1999). Causal effects in non-experimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 1053–1062.
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, 318, 1387–1388.
- Domitrovich, C. E., Cortes, R., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool PATHS curriculum. *Journal of Primary Prevention*, 28, 67–91.
- Duncan, G., Dowsett, C., Claessens, A., Magnuson, K., Huston, A., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428–1446.
- Fantuzzo, J., Bulotsky, R., McDermott, P., Mosca, S., & Lutz, M. N. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review*, 32, 185–203.
- Fantuzzo, J., Bulotsky-Shearer, R., McDermott, P., McWayne, C., Frye, D., & Perlman, S. (2007). Investigation of dimensions of social-emotional classroom behavior and school readiness for low-income urban preschool children. *School Psychology Review*, 36, 44–62.
- Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development*, 72, 1–21.
- Fox, N., Nichols, K., Henderson, H., Rubin, K., Schmidt, L., Hamer, D., et al. (2005). Evidence for a gene-environment interaction in predicting behavioral inhibition in middle childhood. *Psychological Science*, 16, 921–926.
- Garces, E., Thomas, D., & Currie, J. (2002). Longer term effects of Head Start. *American Economic Review*, 92, 999–1012.
- Gormley, W., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, 41, 872–884.
- Gormley, W., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, 320, 1723–1724.
- Guo, S., Barth, R., & Gibbons, C. (2006). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children and Youth Services Review*, 28, 357–383.
- Harding, D. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology*, 109, 676–719.
- Henry, G., Henderson, L., Ponder, B., Gordon, C., Mashburn, A., & Rickman, D. (2003). *Report of the findings from the Early Childhood Study: 2001-02*. Atlanta, GA: Andrew Young School of Policy Studies, Georgia State University.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15, 199–236.
- Howes, C., Burchinal, B., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23, 27–50.
- Ladd, G. W., & Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive-withdrawn children during early grade school. *Child Development*, 70, 910–929.
- Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing (Version 3.0.0). Retrieved August 30, 2011, from <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. (2007). How much is too much? The influence of

- preschool centers on children's social and cognitive development. *Economics of Education Review*, 26, 52–66.
- Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, 75, 47–65.
- Lopez, M., & Barrueco, S. (2005, October). *In search of meaning: Disentangling the complex influences on children's school readiness*. Paper presented at the National Symposium on Family Issues, State College, PA.
- Lowenstein, A. E., & Phillips, D. A. (2010). *Child care, household composition, and aggressive behavior: An examination of risk and protective factors in fragile families*. Manuscript submitted for publication.
- Lutz, M. N., Fantuzzo, J., & McDermott, P. (2002). Multi-dimensional assessment of emotional and behavioral adjustment problems of low-income preschool children: Development and initial validation. *Early Childhood Research Quarterly*, 17, 338–355.
- Magnuson, K., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, 26, 33–51.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., et al. (2008). *Pre-k program standards and children's development of academic, language and social skills*. Unpublished manuscript.
- Mattei, A. (2009). Estimating and using propensity score in presence of missing background data: An application to assess the impact of childbearing on wellbeing. *Statistical Methods and Applications*, 18, 257–273.
- Nagin, D. S., & Tremblay, R. E. (2001). Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school. *Archives of General Psychiatry*, 58, 389–394.
- National Institute of Child Health and Human Development Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development*, 69, 1145–1170.
- National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, 74, 976–1005.
- National Institute of Child Health and Human Development Early Child Care Research Network. (2005). Early child care and children's development in the primary grades: Results from the NICHD study of early child care. *American Educational Research Journal*, 43, 537–570.
- National Scientific Council on the Developing Child. (2005). *Excessive stress disrupts the architecture of the developing brain* (Working Paper No. 3). Retrieved August 30, 2011, from <http://www.developingchild.net>
- National Scientific Council on the Developing Child. (2008). *Mental health problems in early childhood can impair learning and behavior for life* (Working Paper No. 6). Retrieved August 30, 2011, from <http://www.developingchild.net>
- Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36, 859–866.
- Phillips, D., Gormley, W. T., & Lowenstein, E. A. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly*, 24, 213–228.
- Phillips, D., McCartney, K., & Sussman, A. (2006). Child care and early development. In K. McCartney & D. Phillips (Eds.), *Handbook of early childhood development* (pp. 471–489). Malden, MA: Blackwell.
- Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, 23, 10–26.
- Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 77, 302–316.
- Raver, C. C., & Knitzer, J. (2002). *Ready to enter: What research tells policymakers about strategies to promote social-emotional school-readiness among three- and four-year old children*. New York: National Center for Children in Poverty.
- Reynolds, A. (1989). A structural model of first-grade outcomes for an urban, low socioeconomic status, minority population. *Journal of Educational Psychology*, 81, 594–603.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association*, 285, 2339–2346.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, 39, 33–38.
- Royston, P. (2005). Multiple imputation of missing values: Update of ice. *Stata Journal*, 5, 527–536.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91, 473–489.
- Rubin, K., Bukowski, W., & Parker, J. (2006). Peer interactions, relationships, and groups. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 571–645). New York: Wiley.
- Schwartz, C., Wright, C., Shin, L., Kagan, J., & Rauch, S. (2003). Inhibited and uninhibited infants "grown up": Adult amygdalar response to novelty. *Science*, 300, 1952–1953.
- Shaw, D. S., Owens, E. B., Giovannelli, J., & Winslow, E. B. (2001). Infant and toddler pathways leading to early

- externalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 44–51.
- Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, 6, 317–329.
- Snook, S., & Gorsuch, R. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106, 148–154.
- Tincas, I., Benga, O., & Fox, N. (2006). Temperamental predictors of anxiety disorders. *Cognition, Brain, Behavior*, 10, 489–515.
- U.S. Department of Health and Human Services. (2005). *Head Start Impact Study: First year findings*. Washington, DC: Author.
- U.S. Department of Health and Human Services. (2010). *Head Start Impact Study and follow-up, 2000-2010*. Washington, DC: Author.
- Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, 75, 296–312.
- Webster-Stratton, C., Reid, J., & Hammond, M. (2004). Treating children with early onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33, 105–124.
- White, J. L., Moffitt, T. E., Earls, F., Robins, L., & Silva, P. A. (1990). How early can we tell? Predictors of childhood conduct disorder and adolescent delinquency. *Criminology*, 28, 507–533.
- Wilde, E., & Hollister, R. (2007). How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management*, 26, 455–477.
- Wong, V., Cook, T., 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.

Appendix: Kindergarten Sample Sizes

These are sample sizes for age-appropriate kindergarteners. We compared Tulsa Public Schools (TPS) Pre-K alumni with kindergarten children who were in neither TPS Pre-K nor CAP Head Start. We also compared CAP Head Start alumni with kindergarten children who were in neither program.

Table A1
Sample Sizes for Age-Appropriate Kindergarteners

	Tulsa Public Schools		CAP Head Start	
	Alumni	Comparison nonalumni	Alumni	Comparison nonalumni
Full sample	1,565	1,594	470	1,594
Complete social-emotional survey (ASPI)	1,318	1,151	363	1,151
Complete ASPI survey, received parent survey	995	871	244	871
Matched children, phenotype/situtype PSM	1,144	587	337	216
Matched children, attentiveness PSM	1,105	564	331	214

Note. Matched sample sizes are from the first imputed data set in each analysis. In propensity score matching, individuals in the comparison group(s) are sometimes used more than once to provide the best possible match for treatment group children. With matching, the number of observations in the comparison group equals the number of observations in the treatment group. For example, for our first imputation, the working sample size for matched children, phenotypes and situtypes, for Head Start was 674 (337 treatment group children plus 337 control group children, 216 of whom were unique). Because the statistics in this table are based on children who were age appropriate for kindergarten, given Oklahoma's birthday cutoff date of September 1, the sample sizes for assessed children are somewhat smaller than the 1,337 TPS alumni, 366 Head Start alumni, and 1,463 comparison children described in the Method section of this article.