2013

Effects of the Pre-K Program of Kalamazoo County Ready 4s on Kindergarten Entry Test Scores: Estimates Based on Data from the Fall of 2011 and the Fall of 2012

Timothy J. Bartik
W.E. Upjohn Institute, bartik@upjohn.org

Upjohn Institute working paper ; 13-198

W.E. UPJOHN INSTITUTE
FOR EMPLOYMENT RESEARCH
Researching the causes and consequences of unemployment

Effects of the Pre-K Program of Kalamazoo County Ready 4s on Kindergarten Entry Test Scores: Estimates Based on Data from the Fall of 2011 and the Fall of 2012

Upjohn Institute Working Paper 13-198

Timothy J. Bartik
W.E. Upjohn Institute for Employment Research
e-mail: bartik@upjohn.org

June 2013

ABSTRACT

This paper uses a regression discontinuity model to examine the effects on kindergarten entrance assessments of the Kalamazoo County Ready 4s (KC Ready 4s) program, a half-day pre-K program for four-year-olds in Kalamazoo County, Michigan. The results are based on test scores and other characteristics of up to 220 children participating in KC Ready 4s, with data coming from both 2011–2012 and 2012–2013 participants in the program. The estimates find consistently statistically significant effects of this pre-K program on improving entering kindergartners' math test scores. Some estimates also suggest marginally statistically significant effects of KC Ready 4s on vocabulary test scores. No statistically significant effects are found on letter-word identification test scores, due in part to the small available sample size, but some of the point estimates are large. The program does not appear to have large or statistically significant effects in improving children's behavioral assessments. The overall average effects of KC Ready 4s on the three academic test scores are large, at an effect size of at least 0.52. This is toward the high end of effects found in previous studies of short-term effects of pre-K programs. These estimates also are consistent with program benefits exceeding program costs.

JEL Classification Codes: I21, I28, H75

Key Words: Pre-K, Preschool, Regression discontinuity, Evaluation of education programs

Acknowledgments: I thank Wei-Jang Huang and Claire Black for their assistance with this paper. I also thank Sandy Standish and Donna Neevel for their assistance and persistence in organizing this data collection, and Kristal Ehrhardt, Regena Nelson, and Patti Cornelius for their help with data collection and organization. I thank Sandy Standish and Regena Nelson for comments on a previous draft. I serve on the Board of KC Ready 4s. The conclusions of this analysis are my own, and should not be interpreted as reflecting anyone else's views. In particular, they should not be interpreted as official views of the Upjohn Institute or KC Ready 4s.

Effects of the Pre-K Program of Kalamazoo County Ready 4s on Kindergarten Entry Test Scores: Estimates Based on Data from the Fall of 2011 and the Fall of 2012

This paper provides preliminary estimates of test score effects for a relatively new preschool program, Kalamazoo County Ready 4s (KC Ready 4s). These estimates use a methodology, regression discontinuity, that is widely regarded as providing rigorous estimates of cause-and-effect relationships, and that has been extensively used in previous preschool studies. The estimates have some limitations in that I am examining a new program that is still changing, with a small sample size. Therefore, the estimates do not have sufficient sample size to be as precise as one would like, and the changing nature of the program also leads to some uncertainty.

However, the preliminary estimates do suggest that the program has large test score effects. The effects are large compared to how much test scores typically vary across kindergartners, compared to other preschool programs, or compared to what students ordinarily learn in the year before kindergarten. The test score effects are also large enough that we would predict important percentage increases in participants' future earnings. Estimated effects on vocabulary, literacy, and math skills average out to more than a 50 percent improvement in what students typically learn during the prekindergarten year, and would be predicted to increase future earnings by at least 9 percent. On the other hand, I don't find statistically significant effects of KC Ready 4s on improving child behavior, although these measures of child behavior may be more imprecise and biased.

As we will review in the next section, these new estimates are consistent with previous preschool research. The estimated short-term effects of KC Ready 4s on test scores are toward the upper end of the range we observe for most other comparable preschool programs.

After reviewing the prior research literature, I briefly summarize the history and goals of KC Ready 4s. The program has been able to achieve short-term results on test scores, even though most of the program's attention has been devoted to the long-term goal of developing a better preschool system of services in Kalamazoo County.

I then move on to describing the regression discontinuity methodology. Although regression discontinuity methods are not quite as immune to bias as random assignment experiments, the current results do seem to be robust to a wide variety of modeling variations.

Following the methodological discussion, I review the tests used and the data. The tests used are similar to those used by many previous preschool studies, which should make results more comparable.

Finally, I present the results. I provide estimates that suggest that these estimated effects are likely to be associated with program benefits that significantly exceed costs.

The conclusion discusses the implications of these results, as well as plans for future research related to this project.

REVIEW OF PREVIOUS PRESCHOOL RESEARCH

Several well-known randomized experimental studies of preschool have found large short-term and long-term effects. For example, the Perry Preschool program in Ypsilanti, Michigan, conducted in the 1960s, examined the effects of a high-quality half-day preschool program conducted for two years, at ages three and four. Perry studies have found short-term test-score effects at the end of the program with an average effect size of 0.93 (Schweinhart et al.

2005).[1] Although these test score effects fade considerably over time (for example, they are 0.24 in effect-size units by the end of kindergarten), Perry Preschool ended up having profound long-term effects on the life fortunes of former participants, who now have been followed up to age 40. Arrests up through age 40 are reduced by over a third, and average earnings from age 28 to age 40 are increased by over one-fourth (Heckman, Moon et al. 2010). As has been argued by Heckman, Malofeeva et al. (2010), these long-term effects of Perry, despite the fading of test score effects, may be due to Perry's effects in improving participants' "social skills" or "soft skills," which are harder to measure.

The Abecedarian project, another random assignment study, examined a high-quality full-time child care and preschool program that lasted from birth to age five. This study also found large short-term test score effects—for example, the effect size for test score effects at 60 months of age—the end of the program—is 0.52 (Ramey and Campbell 1991). Similar test score effects persisted during the K–12 period. The long-term educational attainment effects of the Abecedarian program would imply lifetime earnings increases of a little over 10 percent (Bartik, Gormley, and Adelstein 2012; Campbell et al. 2012).

Additional long-term evidence of preschool effects is provided by various studies of the Child-Parent Center (CPC) program (Reynolds et al. 2011a,b). This program was a half-day preschool program run by Chicago Public Schools, with about half the children participating just at age 4, and the other half participating at both ages three and four. The one-year program, which is most similar to the KC Ready 4s program, showed an effect size at kindergarten entrance of 0.64. (The two-year program had an effect size of 0.89 [Reynolds 1995].) Based on

[1] "Effect size" is standard education research jargon for expressing the magnitude of estimated test score effects of some program, event, or characteristic. Because test score measures are arbitrary and vary widely, there is an issue in comparing effects across studies. "Effect size" standardizes the actual test score effect by dividing that effect by the standard deviation of the test score across students in a cross section of students at the same grade level or in the same class.

follow-up data through age 26 on effects of the program on increasing educational attainment, the program is estimated to increase lifetime earnings by around 8 percent (Reynolds et al. 2011b).[2]

On the other hand, research on Head Start has been more mixed. Head Start research suggests significant short-term test score effects. The recent random assignment experiment found short-term test score effects in effect-size units that averaged 0.22 (Puma et al. 2012). A meta-analysis by Shager et al. (2013) found short-term Head Start effects with an average effect size of 0.31. But the random assignment experiment also found that these test score effects significantly faded over time, to an effect size that averaged 0.06 in third grade, and these diminished test score effects were no longer statistically significant. On the other hand, research by Ludwig and Miller (2007), Deming (2009), and Currie and others (Currie and Thomas 1995; Garces and Currie 2002) suggests that even though test score effects fade, Head Start has considerable effects on adult outcomes. For example, Deming finds fading test score effects in his study comparing siblings in the same family who differed in whether they participated in Head Start, but then finds large adult effects. Based on the change in adult outcomes, he predicts that Head Start increases former participants' earnings by 11 percent.

At least two recent meta-analyses have tried to summarize test score effects of preschool. Camilli et al. (2010) find test score effects that average 0.48 in effect-size units at the end of preschool, and that average 0.24 in short-term follow-up (ages 5–10). Leak et al. (2010) find short-term effects of age 4 preschool programs that average 0.28.

The regression discontinuity methodology used in my analysis in this paper of KC Ready 4s has been applied in previous studies of state and local preschool programs. Wong et al. (2008)

[2] Direct estimates of the CPC's effects on annual earnings from ages 24 to 27 find an increase of around 7 percent (Reynolds et al. 2011a).

apply this methodology to estimate the effects of preschool at kindergarten entrance in five states. These effect sizes average 0.37 across the various states and tests. Wong et al.'s five-state study includes Michigan's Great Start Readiness Program. Results for this program at kindergarten entrance average an effect size of 0.42. Some of these same authors have also done regression discontinuity studies of Arkansas and New Mexico, with average effect sizes of 0.45 (Arkansas) and 0.55 (New Mexico) (Hustedt et al. 2007, 2009). Bill Gormley at Georgetown and his colleagues have done a variety of studies of the universal pre-K program in Tulsa, Oklahoma (Gormley and Gayer 2005; Gormley et al. 2005; Gormley, Phillips, and Gayer 2008; Phillips, Gormley, and Lowenstein 2009; Gormley 2010; Gormley et al. 2011; Bartik, Gormley, and Adelstein 2012). The Tulsa program has been documented by these researchers to be a high-quality program, based not only on its high pay and high credential requirements for teachers, but also on outside classroom observations of the content of daily interactions between teachers and children in the classroom (Phillips Gormley, and Lowenstein 2009). The most recent Tulsa regression discontinuity study of short-term test score effects, by Bartik, Gormley, and Adelstein (2012), finds an effect size at kindergarten entrance for a half-day preschool program of 0.59, and an effect size for a full-day program of 1.07. A recent study of Tennessee's pre-K program finds an effect size of 0.34 when using random assignment, and of 0.64 using regression discontinuity (Lipsey et al. 2011).[3] Finally, a recent regression discontinuity study of Boston's preschool program, by Weiland and Yoshikawa (2013), finds an average effect size of 0.54.

[3] As will be mentioned later, it is not surprising that regression discontinuity would yield a larger effect size than random assignment, as the comparison group in random assignment is more likely than the comparison group in regression discontinuity to attend some other preschool program in the preceding year, given that families know the child will enter kindergarten the next year. Therefore, the types of analyses are estimating somewhat different effects. In addition, in the case of the Tennessee program (Lipsey et al. 2011), the experiment had a low consent rate for collecting data, which led to some significant differences between the treatment and control groups despite random assignment.

Table 1 lists some of the preschool effects discussed above. The table emphasizes short-term test score effects, because that is what will be estimated here. Also, in cases where data are available for different program variants, I emphasize programs that are most similar to the half-day age-four-only program that is considered in the current paper.

Table 1 Summary of Short-Term Test Score Effects of Various Pre-K Programs

Program	Type of effect considered	Effect size
Perry Preschool	End of pre-K	0.93
Perry Preschool	End of kindergarten	0.24
Abecedarian project	Effects when child is 60 months of age	0.52
Child-Parent Centers, Chicago Public Schools	Kindergarten entry, one-year version of program	0.64
Head Start	Experimental estimates at end of Head Start	0.22
Head Start	Meta-analysis of effects at HS end/kindergarten entrance	0.31
Summary of pre-K studies	Meta-analysis by Camilli et al. (2010) of many pre-K studies, end of program effects	0.48
Summary of pre-K studies	Meta-analysis by Camilli et al. (2010) of many pre-K studies, short-term follow-up, ages 5-10	0.24
Alternative summary of pre-K studies	Meta-analysis by Leak et al. (2010) of many age 4 pre-K studies, short-term effects	0.28
State-funded pre-K in 5 states	Avg. regression discontinuity results at kindergarten entrance, 5 states	0.37
Michigan's state-funded pre-K	Regression discontinuity results at kindergarten entrance	0.42
Arkansas's state-funded pre-K	Regression discontinuity results at kindergarten entrance	0.45
New Mexico's state-funded pre-K	Regression discontinuity results at kindergarten entrance	0.55
Tulsa pre-K, funded by state/school district	Regression discontinuity results at kindergarten entrance, half-day program	0.59
Tennessee's state-funded pre-K	Random assignment results, end of pre-K year	0.34
Tennessee's state-funded pre-K	Regression discontinuity results at kindergarten entrance	0.64
Boston Public Schools' pre-K	Regression discontinuity results at kindergarten entrance	0.54

SOURCE: Perry: Schweinhart et al. (2005); Abecedarian: Ramey and Campbell (1991); CPC: Reynolds (1995); Head Start experiment: Puma et al. (2012); Head Start meta-analysis: Shager et al. (2013); 5-state and Michigan RD (regression discontinuity) results: Wong et al. (2008); Arkansas: Hustedt et al. (2007); New Mexico: Hustedt et al. (2009); Tulsa: Bartik, Gormley, and Adelstein (2012); Tennessee: Lipsey et al. (2011); Boston: Weiland and Yoshikawa (2013).

As the table shows, short-term test score effects for programs that have at least some similarity with KC Ready 4s tend to be in the range of 0.20 to 0.70. Most effect sizes seem to be below 0.60. So, prior to examining any data on KC Ready 4s, it is perhaps a reasonable expectation that the program can achieve some effect in the 0.20 to 0.60 range.

BACKGROUND ON KALAMAZOO COUNTY READY 4s (KC READY 4s)

Kalamazoo County Ready 4s is a nonprofit organization seeking to ensure universal access to pre-K education for all four-year-olds in Kalamazoo County. Kalamazoo County is a county of a little over 250,000 population in southwest Michigan, about equidistant between Chicago and Detroit. The county includes a mixture of urban areas (the cities of Kalamazoo and Portage and surrounding areas) and rural areas, with an economic base that includes higher education (Western Michigan University and Kalamazoo College) as well as biotech and manufacturing (Pfizer and Stryker Medical Instruments). The county has an overall poverty rate of around 19 percent, and the population mix includes 11 percent African American and 4 percent Hispanic. Much of this poverty, as well as much of the minority population, is concentrated in the urban core (the city of Kalamazoo, population around 75,000, has a 34 percent overall poverty rate, with 22 percent of residents being African American and 6 percent Hispanic), but there also is considerable rural poverty.

Kalamazoo City and County have a long history of investing in education. A lawsuit known as the Kalamazoo School Case, from 1874, involved a taxpayer suing to prevent the city from providing free publicly supported high schools; the Michigan Supreme Court's ruling in favor of taxpayer-supported public high schools was influential in encouraging other state and local areas to expand publicly supported K–12 education. The Kalamazoo area aggressively lobbied the state to recruit a state teacher's college (now Western Michigan University) to locate in the area in 1903. More recently, the Kalamazoo Promise has gained national attention, including visits by national media and a commencement speech at Kalamazoo Central High School by President Obama. Starting in 2005, the Kalamazoo Promise, funded by anonymous private donors, has provided funding for up to 100 percent of in-state college tuition for graduates of Kalamazoo Public Schools.

Kalamazoo County Ready 4s came out of this longstanding local tradition of seeing investment in education as a means of both individual betterment and local economic development. KC Ready 4s developed out of an initiative by a local interfaith community organizing group called ISAAC.[4] At a large community meeting in the fall of 2009, ISAAC obtained public commitments from many local political and community leaders for the goal that all Kalamazoo County four-year-olds should have access to high-quality pre-K programs. These public commitments were followed by a two-year planning process, coordinated by the local United Way and involving participants from local preschools, local school districts, the business community, and researchers in what was then called the Kalamazoo County Committee for Early Childhood Education. Kalamazoo County Ready 4s was then created as an independent nonprofit organization in early 2011.

The long-term goal of KC Ready 4s is to ensure that all Kalamazoo County four-year-olds have access to high-quality pre-K programs. High-quality pre-K is assumed to at a minimum require a half-day pre-K experience for five days a week during one school year. To reach that goal, KC Ready 4s provides financial support for families that cannot afford high-quality pre-K. KC Ready 4s also coordinates and funds efforts to expand the number of high-quality pre-K slots. Even during the two-year planning process, community efforts began to improve local pre-K options. The Kalamazoo County Committee, followed by KC Ready 4s, has funded efforts to provide outside evaluation of local pre-K programs, as well as mentoring,

[4] Interfaith Strategy for Action and Advocacy in the Community, an affiliate of the Gamaliel Foundation, a national network of faith-based community organizing groups that is perhaps best known for having at one time employed a community organizer named Barack Obama.

teacher training, and curriculum assistance to improve program quality. These efforts are expected to continue and intensify in the future.

From the beginning, KC Ready 4s has included both public and private pre-K providers. The thinking is that either public or private pre-K programs can be of high quality, and that it would be foolish both politically and substantively to exclude either group from KC Ready 4s' support and help.

In the 2011–2012 school year, KC Ready 4s began providing subsidies for high-quality slots on a pilot basis. Sixty-five slots were supported by KC Ready 4s in that pilot year. For the 2012–2013 school year, the number of KC Ready 4s slots expanded to 130. These KC Ready 4s slots go only to private or public providers that are rated as being at four or five stars on a quality scale. KC Ready 4s currently is supporting slots at seven private providers and two public providers. KC Ready 4s is also working with many other providers to improve their quality rating to a four- or five-star level so that they are eligible for future funding.

KC Ready 4s hopes to continue to expand in subsequent years. The long-term goal is to provide a sufficient number of supported slots to ensure universal access to pre-K for all four-year-olds in Kalamazoo County. This is expected to require roughly 1,000 slots. With slightly over 3,000 four-year-olds in Kalamazoo County, and with about 1,000 four-year-olds enrolled in Head Start or the state-funded Great Start Readiness Program, the expectation is that an additional 1,000 supported high-quality slots would allow all interested families to access high-quality pre-K.

KC Ready 4s funding for supported slots is provided on a sliding-scale fee basis related to family income. The program assumes that $4,500 is needed to support a half-day preschool slot for five days a week for one school year. Table 2 shows the evolving sliding-scale formula

Table 2 KC Ready 4s Fee Schedules for 2011–2012 and 2012–2013 School Years

Fee schedule for 2011–2012	
Family income in relation to poverty line for that household size (%)	KC Ready 4s pays the below percentage of annual tuition of $4,500 (%)
< 325	100
325–349	95
350–374	90
375–399	85
400–424	75
425–449	65
450–474	55
475–499	45
500–524	30
525–549	15
≥ 550	0
Fee schedule for 2012–2013	
Family income ($)	KC Ready 4s pays the below percentage of annual tuition of $4,500 (%)
< = 50,000	100
50,001–70,000	98
70,001–80,000	90
80,001–90,000	75
90,001–100,000	50
100,001–110,000	25
> 110,000	0

for the 2011–2012 pilot year and for the 2012–2013 school year. The initial fee schedule was extraordinarily generous at high-income levels, which was controversial, and also was more complex because of its ties to family size. The newer fee schedule restricts the income range of families eligible for subsidies and is much simpler. For 2013–2014, the KC Ready 4s fee schedule has been further scaled back at higher income levels. The fee schedule for 2013–2014 eliminates any subsidies for families at greater than $90,000 income, and it imposes at least some modest family copays at all income levels.

KC Ready 4s provides support to approved providers by agreeing to provide tuition subsidies for a certain number of slots, in many cases an entire class. The actual subsidies provided to families depend upon the income mix of those choosing to enroll at that preschool. All those in supported slots are counted as KC Ready 4s participants, even if their family income means that they are ineligible for tuition subsidies. However, even for those families, KC Ready

4s efforts may have meant the creation of that class, and KC Ready 4s efforts may have helped upgrade the quality of that preschool.

KC Ready 4s works closely with the local Head Start program, and with local administrators and participants in the state-funded Great Start Readiness Program. For example, the three programs have adopted a common application for preschool admission. The three programs also are seeking to provide more “braided” funding opportunities, in which children in the same class may receive funding from these three different programs. This will allow greater income integration of local pre-K classrooms.

KC Ready 4s is currently supported by private financial sources. This includes the local United Way, local foundations, local banks and businesses, and private individuals. Achieving the program’s full goal of universal access may at some point require public funding from some as-yet-unknown combination of federal, state, and local sources, although no definite plans have been developed for funding the full-scale program.

From its beginning, KC Ready 4s has aimed to ensure objective evaluation of its program success. Accordingly, starting with the pilot class, pre-tests have been administered to all KC Ready 4s participant children as early as possible in the school year. Post-tests are then administered at the same time of year for program graduates who are beginning kindergarten. The same tests at the same time of year are administered both times. As will be explained later, this evaluation design is needed to implement the regression discontinuity approach to evaluating preschool effectiveness. So far, the program has pre-program test data available for the 2011–2012 and 2012–2013 cohorts. Post-program test data are available for the 2011–2012 cohort. All of these data are used in the current evaluation.

THE REGRESSION DISCONTINUITY METHODOLOGY AND PRESCHOOL PROGRAMS

The regression discontinuity approach is intended to provide estimates that are "almost" as good as random assignment in estimating the true causal impact of some program (Lee and Lemieux 2010). If random assignment experimentation is the "gold standard" of program evaluation, regression discontinuity could be viewed as meeting a "silver standard."

In the absence of random assignment, the problem with evaluating the impact of some program can be described as being due to selection bias. On average, it is plausible that the participants in the program, versus seemingly similar nonparticipants, will differ in unobserved ways. Participants may differ from nonparticipants because of "self-selection"—that is, because of unobserved differences that drive why participants choose to participate. Participants may also differ from nonparticipants because of "program selection"—that is, because of unobserved differences that drive why the program makes choices that lead to the enrollment of certain types of participants.

Either type of selection bias can bias program results in a positive or negative direction. For example, perhaps participants choose a program because they are more ambitious, which may bias the estimates toward finding positive effects of the program. Or perhaps participants may choose a program because they are needier than average in ways we don't observe, which may bias estimates toward finding that the program has zero or negative effects. The program might also choose to help the neediest, which would also bias estimates in a negative direction. Or the program may try to enroll participants with more potential, which would make the program look more effective.

To put this in mathematics, we are trying to estimate the following equation:

(1) $$Y_i = \mathrm{B}_0 + \mathrm{B}_t * T_i + \mathbf{B_x} * \boldsymbol{X_i} + e_i \,.$$

Y_i is the outcome of interest for an individual child, indexed by i. T_i is a zero-one dummy for treatment or participation in the program. $\boldsymbol{X_i}$ is a vector of observed individual characteristics that may affect the outcome variable. B_0, B_t, and $\mathbf{B_x}$ are parameters to be estimated, with B_t being the parameter of most interest for program evaluation, as it represents the effect of the program on the outcome of interest. e_i is the disturbance term.

The estimation problem is that the disturbance term e_i includes unobserved characteristics of the individual that affect outcomes. If these unobserved characteristics are correlated with the treatment dummy/participation dummy T_i , then the estimates of B_t will be biased.

Random assignment solves the selection bias problem by assigning participation randomly. We expect the unobserved attributes of participants versus nonparticipants to tend to be the same on average as the sample size increases, because participants are neither self-selected nor program-selected, but rather are selected randomly. As a result, we can be confident, at least with a large enough sample size, that the resulting postprogram differences between participants and nonparticipants are due to the program and not to differences in preprogram characteristics between participants and nonparticipants. From an econometric standpoint, if T_i is randomly assigned, it would not be expected in a large sample to on average be correlated with unobserved characteristics.

Regression discontinuity is used in a setting where random assignment data are unavailable. However, it exploits a situation where assignment to a program is based on some abrupt threshold or cutoff in some variable Z_i. If we assume that this variable itself has a smooth effect on the outcomes variables we're interested in, then we can separate the effect of the program itself from other variables. Or to put it another way, if variable Z_i by itself does not cause large changes in the outcomes variable, then individuals who are just below or just above the cutoff for program participation are “almost the same” in all unobserved and observed characteristics that might affect potential program outcomes, except for one: whether or not they are assigned to the program.

Mathematically, we include some function of Z_i as a control in the regression equation. We might also interact that function of Z_i with the treatment dummy. It is conventional to first difference Z_i from the cutoff Z_c which determines participation in the program. We then enter that difference from the cutoff as a linear function, a quadratic function, etc. Under this formulation, this function equals zero when Z_i equals Z_c. If this function is linear, we get the following estimating equation:

(2) $$Y_i = \mathrm{B}_0 + \mathrm{B}_t * T_i + \mathrm{B}_z * (\mathrm{Z}_i - \mathrm{Z}_c) + \mathrm{B}_{zt} * T_i * (\, Z_i - Z_c) + \mathbf{B_x} * \boldsymbol{X_i} + e_i \,.$$

More generally, we might also allow for squared or higher-order terms in $(Z_i - Z_c)$, with the coefficients on these higher-order terms also allowed to vary with the treatment dummy.

We then hope that if we control for the influence of Z_i on the outcome variable to the right and left of the cutoff, then we can separate the influence of assignment due to the cutoff from the effect of the Z_i variable by itself. We can more accurately do so if we have a more general functional form for how $(Z_i - Z_c)$ affects the outcome variable, or if we consider only observations closer to the cutoff Z_c. At the extreme, if we had sufficient observations to have a very large sample only a tiny amount in either direction away from Z_i, then the direct effects of any remaining variation in Z_i would be negligible, and the program participation would be arbitrarily close to being randomly assigned.

For example, in an education setting, regression discontinuity has been used to evaluate the effects of a summer school program in the K–12 system in which students are assigned to the summer school program based on falling below some cutoff performance level on a standardized test in the immediately preceding spring (Jacob and Lefgren 2004). We can then look at how performance the following fall on some test is related to the prior spring's test score level. Presumably fall test scores will be positively related to spring test scores. But if we see some "break" in test scores—where students who in the spring scored "just below" the test score cutoff tended to do surprisingly well in the fall relative to students who in the spring scored "just above" the test score cutoff—then this break is reasonably strong evidence of positive effects of the summer school program on test scores.

For pre-K programs at age four, regression discontinuity techniques have been used by exploiting the fact that age is an assignment variable. We administer the same tests at the same time of year to entering pre-K students, as well as to entering kindergartners who participated in that same pre-K program the previous school year. Whether students participated this year or the previous year will be based on their age. The children's dates of birth will extend over a two-year period. Presumably children's test scores will tend to increase with age. But close to the age cutoff for entering kindergartners, there will be some children who just made the age cutoff this year and therefore participated in pre-K the previous year, and other children who just missed the age cutoff this year and therefore are only entering pre-K this year. If pre-K has an effect on test scores, we expect to see an abrupt "jump" in test scores at the age cutoff: the slightly older children will have much larger test scores than we would expect for the slightly younger children, based on their age.

Or, to put it another way, on average the entering pre-K students differ by one year in age from the pre-K graduates entering kindergarten. Otherwise they are similar in having been both self-selected and program-selected to enter the same pre-K program. The one-year-greater average age of entering kindergartners will increase average test scores, even if pre-K has zero effects. But we can control for age by looking at how test scores vary with age within the pre-K entrant group and within the pre-K graduate/kindergarten entrant group. After controlling for age, we assume the remaining test score variation between the two groups is due to the pre-K program.

Regression discontinuity has been used to evaluate many preschool programs. This includes pre-K programs in Michigan, New Jersey, West Virginia, South Carolina, Oklahoma, Arkansas, New Mexico, Tennessee, Tulsa, and Boston (see Table 1 for specific studies).

There are two possible problems with regression discontinuity as an evaluation technique for pre-K. The first is that there may still be some remaining selection bias. This could occur if the selection process varies from year to year. In most of the studies that have been done of pre-K using regression discontinuity, the tests have all been administered the same fall, which means that we are comparing children who were selected this year into pre-K with children selected last year. If the selection process has changed over time, then these changes in the selection process may be responsible for differences in test results, not the program. Selection bias could also occur if there is attrition from the sample due to people moving to different places or schools over time, so that the observed sample of pre-K graduates/entering kindergartners differs in unobserved attributes from the observed sample of pre-K entrants.

The possibility of selection bias in regression discontinuity studies of pre-K can to some extent be ascertained by seeing whether there are significant jumps in observed variables other than test scores at the age cutoff. If the observed characteristics of pre-K entrants versus pre-K graduates/kindergarten entrants abruptly change at the cutoff, this suggests that there may also be abrupt changes in unobserved characteristics as well. In general, in these previous studies, the researchers do not find evidence of abrupt shifts in observed characteristics at the age cutoff, which suggests that selection bias is not a problem. In the present study, we also have another control for selection bias: we have observations before and after pre-K on the same children, for whom we expect unobserved preexisting characteristics to be the same.

The second problem with using age cutoffs to estimate regression discontinuity models of pre-K effects is that participation in pre-K programs is not the only difference between the two groups, the group of entering pre-K students versus the group of entering kindergartners. During the preceding year, parents knew that the one group would be entering kindergarten the next year, whereas the other group would be entering pre-K. This difference may have led to different experiences of the two groups (Whitehurst 2013).

I suspect that this criticism has some truth, but that it only changes the interpretation of the estimates rather than their validity. The main thing that parents do because their child is entering kindergarten next year is to enroll their child in pre-K. One can hypothesize that parents may also devote more time to academic tutoring or encouraging their child to socialize with older kids. But I think such differences in parent behavior other than pre-K enrollment are much more speculative.

This does affect the interpretation of the regression discontinuity pre-K estimates. In a random assignment experiment, we would randomly assign some kids to attend our designated pre-K program, whereas other families would be left to their own devices. In such a study, with all kids in both the treatment and control group entering kindergarten the following year, a considerable percentage of the control group will be enrolled in some other pre-K program. Therefore, in a random assignment study of a pre-K program, the estimated program effects are really the effects of the pre-K program being examined versus the effects of whatever pre-K programs families find on their own. In contrast, regression discontinuity studies of pre-K are really studying the effects of the designated pre-K program versus the much lower frequency of pre-K enrollment for kids who won't be entering kindergarten for another two years. We would therefore expect regression discontinuity studies to yield somewhat larger test score effects than random assignment studies of pre-K. On the other hand, the regression discontinuity studies of pre-K will come closer to estimating the effects of pre-K versus no pre-K. It is these effects that should be compared with the total costs of pre-K.

If one looks back at Table 1, it is not obvious that this is a major problem.[5] The Perry Preschool, Head Start experiment, and Abecedarian results come from random assignment experiments, while the Chicago Child-Parent Center (CPC) study has a nonrandomly chosen comparison group of neighborhoods with no availability of the CPC program. In contrast, the last results in the table are for regression discontinuity studies. It is not obvious that the regression discontinuity studies tend systematically to have larger effect sizes than other studies that use either randomly chosen comparison groups or other comparison groups.[6]

BACKGROUND ON DATA AND TESTS

KC Ready 4s collected test data in the fall of the 2011–2012 year and the fall of the 2012–2013 year for program entrants. KC Ready 4s also collected test data in the fall of the 2012–2013 year at kindergarten entrance for children who had participated in KC Ready 4s in 2011–2012 (program "graduates").

[5] In addition to the results in Table 1, the study by Bartik, Gormley, and Adelstein (2012) uses regression discontinuity methods and finds effects for a full-day pre-K program that are around twice as great as for a half-day program. It is hard to see why biases due to different parental behavior would result in this pattern.

[6] One exception is the Tennessee pre-K program, which shows larger estimates from the experiment than from regression discontinuity, 0.64 versus 0.34 in effect size units. But this experiment has problems stemming from differential attrition in data collected from the treatment versus control groups, so it is unclear whether these differences are due to differences in parent behavior as children approach kindergarten, or due to problems in this experiment.

Among the tests were three academic achievement tests. These three tests were: 1) the Peabody Picture Vocabulary Test, Version 4 (Vocabulary); 2) the Woodcock-Johnson Subtest 1, Letter-Word Identification (Letter-Word ID); and 3) the Woodcock-Johnson Subtest 10, Applied Math Problems (Math). The Peabody Picture Vocabulary Test (PPVT) asks a child to identify which of four pictures corresponds to a word the child has been given. The Woodcock-Johnson Letter-Word ID test shows children letters and words and asks them to say those letters or words. The Woodcock-Johnson Applied Problems test asks a child to solve a very simple arithmetic problem.

These three academic tests were chosen in part because they are nationally normed tests that are widely respected and used. In addition, these tests are designed so that the same tests can be given to children at both preschool entrance and kindergarten entrance; in this way, children are measured on the same scale. Furthermore, these tests have frequently been used in previous studies of pre-K.[7]

Because I want to measure children's test score performance on the same scale, I do not age-norm the resulting test scores using each test's standard age norming procedures. Rather, I use the raw scores.[8] The empirical analysis, by controlling for age, implicitly uses this study's own data to do the age norming.

[7] Some other studies that have used one or more of these tests are as follows: the various Tulsa studies (e.g., Bartik, Gormley, and Adelstein 2012) include Letter-Word ID and Applied Problems tests; the regression discontinuity studies of state programs done by the National Institute of Early Education Research used the PPVT and the Applied Problems test (Hustedt et al. 2007, 2009; Wong et al. 2008); the Tennessee study used the Letter-Word ID and Applied Problems tests (Lipsey et al. 2011); the Boston study used the PPVT, the Letter-Word ID test, and the Applied Problems test (Weiland and Yoshikawa 2013); the Head Start experiment used the Letter-Word ID test and the Applied Problems test (Puma et al. 2012); Deming's study of Head Start used the PPVT (Deming 2009); and the Perry study used the PPVT (Schweinhart et al. 2005).

[8] For the PPVT Vocabulary test, I use the "growth score value" (GSV) transformation of the raw score, which is supposed to equate test scores across different forms and editions of the test, and to report the scores in a form in which equal intervals correspond to equal learning (Dunn and Dunn 2007). In estimates not reported here, I also looked at the PPVT raw score, which gives very similar results to the GSV scores.

These data were collected by trained contractors with KC Ready 4s. The training was supervised by early childhood education professors at Western Michigan University.

Although ideally all these data would have been collected in the first few weeks of school, delays in getting responses from teachers and obtaining parental consent led to some delays in test administration. Tests were generally administered in October and November of each year. This implies that the study is measuring the impact to a small degree of the combination of the KC Ready 4s pre-K program with the first month or so of kindergarten.

KC Ready 4s also collected behavioral data on students. The behavioral data were collected using the Devereux Early Childhood Assessment (LeBuffe and Naglieri 1999a,b). This assessment asks the teacher to assess how often a child has exhibited each one of 37 behaviors in the previous four weeks: never, rarely, occasionally, frequently, or very frequently. Some of these behaviors are positive and some are negative. The positive behaviors are used to calculate a "Total Protective Factor" score, and the negative behaviors are used to construct a "Behavioral Concerns" score. I used the raw version of these two scores rather than any percentiles that can be constructed from the scores. It should be kept in mind that the Behavioral Concerns measure is a "negative" measure, which we would hope would decline with age and with pre-K experience.

One issue with the Devereux behavioral measures is their subjective assignment by teachers. The academic tests were administered by the same trained workers for both pre-K entrants and pre-K graduates/kindergarten entrants. In contrast, the Devereux assessment for pre-K entrants is completed by pre-K teachers, and the Devereux assessment for pre-K graduates is completed by kindergarten teachers. This introduces greater subjective variation in the behavioral scores. This subjective variation differs across pre-K entrants and graduates, which may bias the results. In particular, pre-K teachers and kindergarten teachers may have different standards for behavior. For example, even though in theory the Devereux scores I use are not explicitly age-normed, it is possible that kindergarten teachers have higher standards for behavior than pre-K teachers, which would tend to make the behavioral scores worse for pre-K graduates even if the behavior, viewed objectively, has actually improved over the year.

KC Ready 4s also collected data on the pre-K application form for a number of characteristics of the family and child. These characteristics include the following: child's race, child's Hispanic status, child's gender, whether both parents are currently present in the family, family income, education level of the father, education level of the mother, age of the mother, and age of the father.

Because of the limited sample size of the current study, it seemed prudent to seek to succinctly summarize these variables. Race and Hispanic status were summarized as a zero-one variable for whether the child was nonwhite or Hispanic. Education of the parents was calculated as years of education, where high school dropout = 10, high school graduate = 12, community college graduate = 14, bachelor's degree = 16, master's degree = 18, and doctorate = 20.

Another issue is missing values. For the relatively few cases in which demographic characteristics were missing, I assigned the sample mean to those observations.[9] For income, the missing cases seemed to be households in which income was too high to receive a KC Ready 4s subsidy. For these cases, I assigned income equal to 140 percent of the cutoff for that family and year for receiving a KC Ready 4s subsidy, based on the fee table. The regressions also included

[9] For the discrete variables "minority" and "single parent," I assumed that missing cases were nonminority and both parents were present.

a control for income being assigned because it was missing. Finally, all income was adjusted to 2012 dollars.

Another limitation of the demographic data is that all the data refer to the application form for pre-K. Therefore, any changes in family composition or family income between the application form and the test date, which is over a year later for the kindergarten entrant/preschool graduate tests, would not be reflected in the data.

We ended up with 62 pre-K observations with academic test score data from cohort A, who participated in pre-K in 2011–2012. (For two of these children, we do not have the Devereux data, so for those dependent variables, we have two fewer total observations.) KC Ready 4s was able to track down and collect academic test score data and behavioral assessments from 51 of these children at kindergarten entrance in the fall of 2012. We have 107 observations with valid data on academic test scores from cohort B, who participated in pre-K in 2012–2013.[10] This cohort does not yet have post-pre-K data, although KC Ready 4s plans to collect such data in the fall of 2013. However, this cohort does contribute to the evaluation, because its members' test score data do reveal how test scores vary with age up to the age cutoff.

Table 3 reports some summary statistics for the test score variables used and the demographic variables. As can be seen in the table, academic test scores and behavior both improve from the pre-K entrants to the pre-K graduates. The question is whether the improvements exceed what we would expect because of one year of aging, which I will explore later in the paper.

[10] Why 107, and not the 130 students enrolled in KC Ready 4s in 2012–2013? This is largely due to delays in getting children signed up for KC Ready 4s, so that some 2012–2013 enrollment did not occur until after the testing was completed.

Table 3 Descriptive Statistics

Variable	Mean	Standard deviation
Pre-K entrant test scores		
Vocabulary	129.40	16.83
Letter-Word ID	9.54	5.51
Math	13.13	4.17
Total protective factors	81.47	15.82
Behavioral concerns	7.31	4.29
Pre-K graduate test scores		
Vocabulary	147.47	10.84
Letter-Word ID	21.80	9.65
Math	19.16	3.90
Total protective factors	83.71	15.77
Behavioral concerns	6.35	4.74
Demographic and socioeconomic variables		
Female child (0–1)	0.532	0.500
Minority child (0–1)	0.382	0.487
Single-parent family (0–1)	0.182	0.387
Ln of family income	11.062	0.821
Missing income	0.218	0.414
Education of mother (in years)	15.381	2.776
Education of father (in years)	15.091	2.781
Age of mother (in years)	34.265	5.385
Age of father (in years)	37.099	6.062

NOTE: Minority child is child who is reported as being either nonwhite or Hispanic. If family income is missing, as explained in text, I assume the family is in an income category that does not receive a pre-K subsidy, and I assume income equal to 1.4 times that implicit top code for family income from the fee schedule. Missing values for minority child or single-parent family are assumed to be nonminority and both parents present. For other demographic variables, sample mean is assigned to missing values. Test score data are raw scores, except for Vocabulary, which is the "GSV" transformation of the raw score, as explained in text.

As for the demographics, this is a group that includes a variety of income groups. The mean of log family income corresponds to annual family income of a little over $63,000. This might be seen as a large family income to be a program average, but this includes families who are participants in KC Ready 4s but do not receive any tuition subsidies. Of the 170 children entering KC Ready 4s for whom we have data, 43 did not receive any tuition subsides even though they were in a slot and class sponsored by KC Ready 4s. Seventy-five of the 170 had low-enough income to receive the full $4,500 subsidy. The remaining 52 children were from families that received a partial subsidy for preschool tuition.

If we weight family income by the amount of subsidy received, we find that 25 percent of all KC Ready 4s subsidies went to families with income of $27,033 or less, 50 percent went to families with income of $45,206 or less, and 75 percent went to families with income of $66,006 or less. The income statistics reflect two features of KC Ready 4s and the pre-K environment in Kalamazoo. First, KC Ready 4s is aiming to supplement existing pre-K programs and expand to universal access to high-quality pre-K. Second, under the cooperative arrangements that KC Ready 4s has with the local Head Start and state-funded pre-K programs, children who meet income limits for those programs are first referred to those programs if slots are available. Head Start's income limit is 100 percent of the poverty line, whereas the state-funded Great Start Readiness Program usually has an income limit of 300 percent of the poverty line. Therefore, it is natural that KC Ready 4s will include a wide variety of low-income, working-class, and middle-class families who either are unable to obtain a federal- or state-funded slot, or who are ineligible for such slots.

Based on Table 3's statistics, the sample includes a significant percentage of Hispanic or nonwhite children, at almost 40 percent. A little less than 20 percent are in single-parent families. Average education levels of parents are above average, at more than 15 years (high school plus three years of college). Ages of mother and father seem to be fairly typical.

One implication is that although KC Ready 4s includes a significant number of disadvantaged children, the program also includes a broader population. This differs from many previously studied pre-K programs, which often were targeted at the disadvantaged. If one thinks that pre-K has its greatest impact on the disadvantaged, then this implies that KC Ready

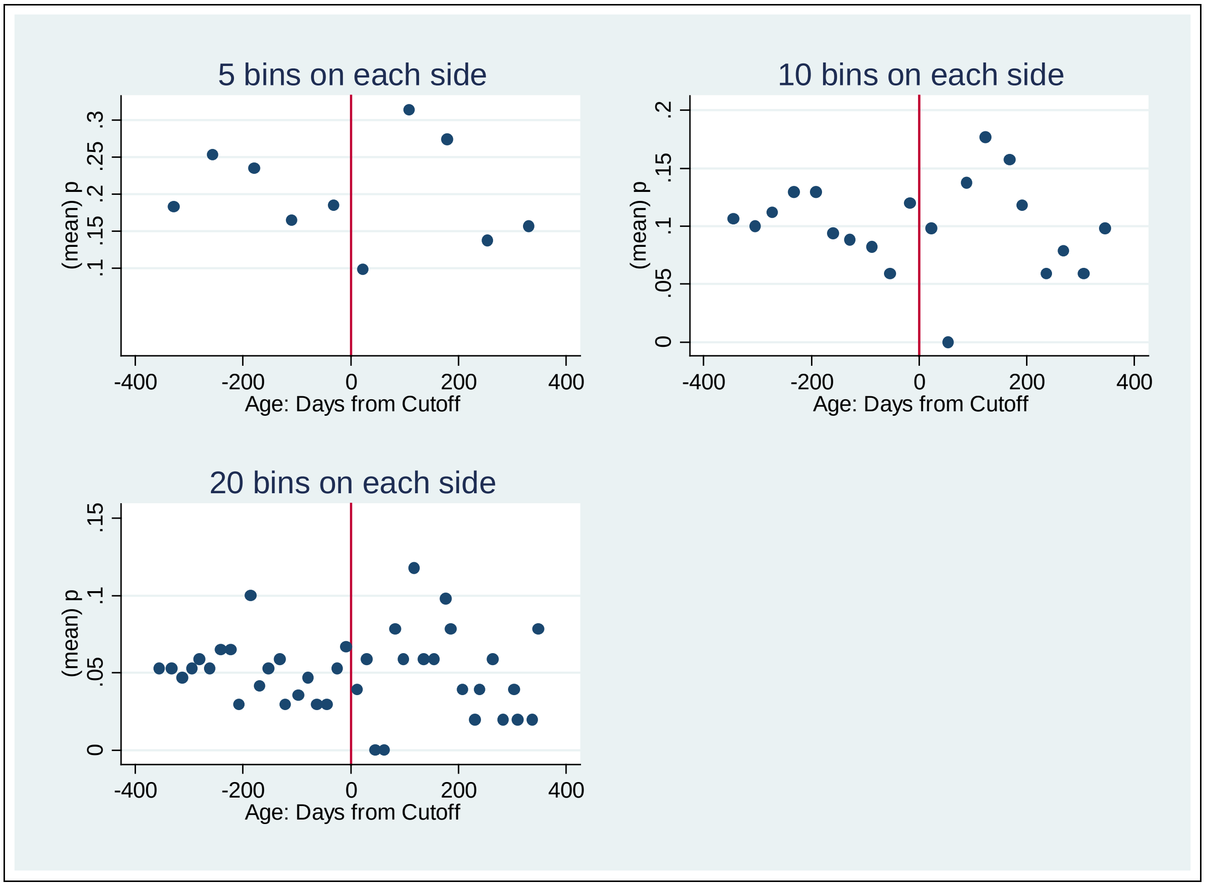
4s might be expected to have lower impacts than some other pre-K programs. On the other hand, Bartik, Gormley, and Adelstein (2012) find similar test score impacts of Tulsa's pre-K program on both disadvantaged and nondisadvantaged children, and Weiland and Yoshikawa (2013) find statistically significant impacts of Boston's pre-K programs for all income groups.

RESULTS

I first consider whether there is any sign of "bunching" of the age variable. If there was a strong sign that individuals of particular ages were disappearing disproportionately from either the pre-K entrant or the pre-K graduate sample, this might be a concern. For example, if children who were doing poorly disappeared from the pre-K graduate sample, particularly if they were relatively young, then this might bias the estimates.

Figure 1 shows the results where the observations used are grouped into 5 bins, 10 bins, or 20 bins on each side. I calculate the percentage in each bin as a percentage of the total on each side of the age cutoff. (There are 169 observations in the pre-K entrant group to the left of the cutoff, and 51 observations in the pre-K graduate/kindergarten entrance group to the right of the cutoff.) As can be seen in the figure, there is no obvious sign of a bunching of observations. There are a few bins near the cutoff on the right-hand side that are underpopulated, but these are not the bins that are closest to the cutoff, and then some bins relatively close to the cutoff on the right-hand side are also overpopulated. All of this could well be the result of chance, given the relatively sparse number of observations (51) in the pre-K graduate sample.

Figure 1 Distribution of KC Ready 4s Children by Age; Children Grouped by Age Bins

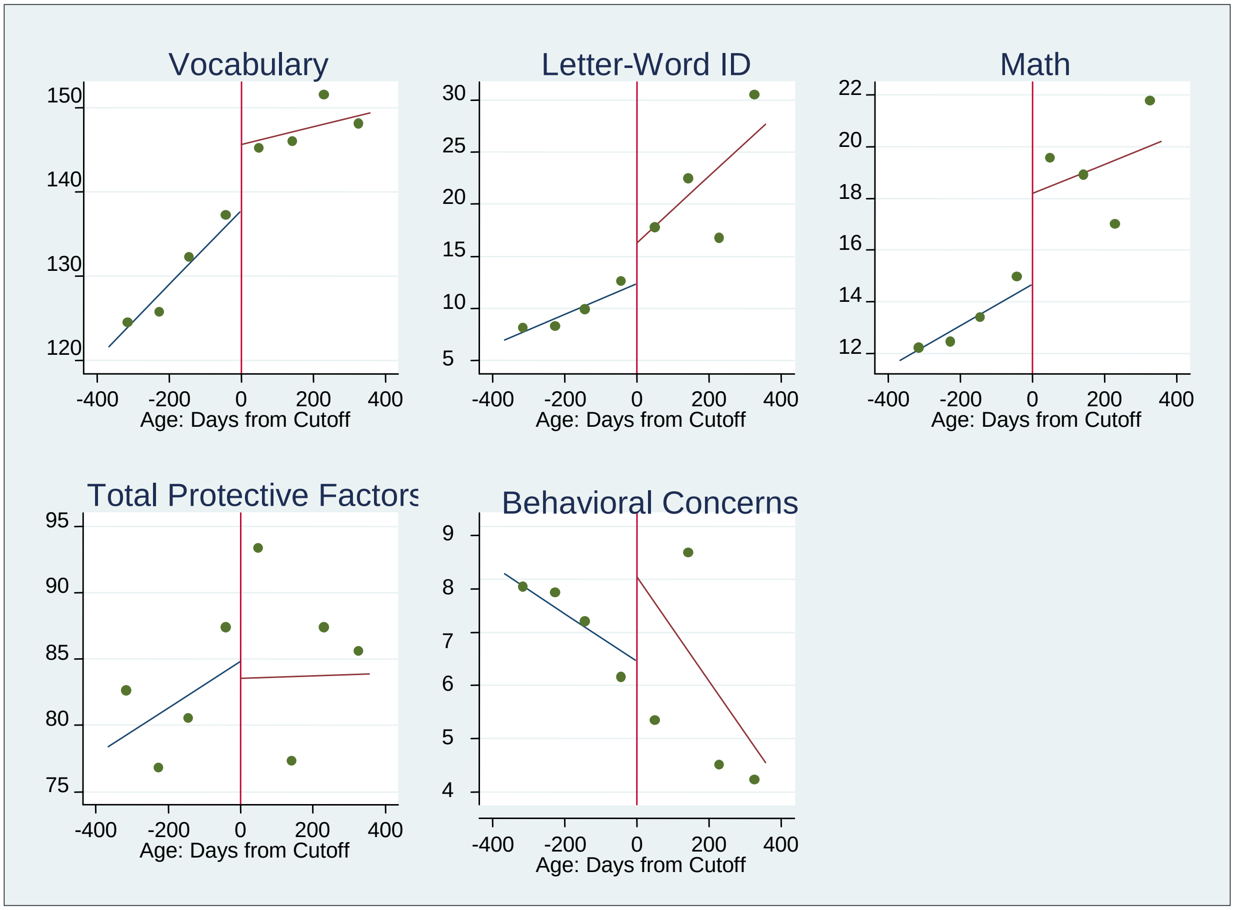


NOTE: This chart provides estimates of the sample frequency by age when I group pre-K entrants (to left of age cutoff) and pre-K graduates (to right of age cutoff) into a certain number (either 5, 10, or 20) of equal-sized ranges of days to either side of the age cutoff. The calculated frequency divides the number of children in each age bin by the total number of children in each group (e.g., the frequency is calculated separately for pre-K entrants versus pre-K graduates). This is done because there is a smaller total sample size for pre-K graduates, both because there is only follow-up data for the first cohort, and because there is some sample attrition from pre-K entrant to pre-K graduate.

I then move on to consider constructing a picture for the relationship between the test score variables and the age variable. Specifically, I consider whether an examination of some simple averages by “bins” to the right and left of the age cutoff reveals a jump at the cutoff. After some experimentation, I settled on four bins to either side of the age cutoff. This was the largest bin size that passed the various tests proposed by Lee and Lemieux (2010). Specifically, after including four bins on each side, additional sub-bins added in were not statistically significant, and adding in a set of interactions between bin dummies and the age variable was also not statistically significant.

Figure 2 shows this simple picture. Accompanying the average test scores by bin size is a simple linear regression of test scores on age, with the slope allowed to vary on either side of the age cutoff.

Figure 2 Variation of Test Scores with Age



NOTE: The points in the figure show average test scores by age range when children are grouped in four equally sized bins (in terms of range of birth dates included) to both left and right of age cutoff. The points show average test score and average age for each bin. The lines shown are from simple linear regression of test score on constant term, treatment dummy, and linear term in age differenced from cutoff, as well as interaction term between that age differenced variable and treatment dummy.

The figure suggests strong evidence that Math test scores jump significantly at the cutoff, which in turn suggests an effect of KC Ready 4s. There also is some evidence of a jump in Vocabulary scores, and perhaps Letter-Word Identification scores. On the other hand, there is little sign of improvement at the cutoff in the behavioral measures.

One assumption of the regression discontinuity methodology is that there will not be a strong jump in unobservable characteristics at the cutoff. In the present case, the assumption is that at the age cutoff, the children and families on either side close to the age cutoff do not differ significantly in unobserved characteristics. This cannot be directly tested because these characteristics are unobserved. But we can test whether there is a jump in observed characteristics.

Table 4 tests for a jump by regressing each test score variable, and all the demographic controls, on a treatment dummy (1 if pre-K graduate, 0 if pre-K entrant), the differential of the age variable from the age cutoff, and an interaction term between the treatment dummy and the age differential. This allows age to linearly affect each variable, with a possible jump at the age cutoff, and with the linear slope of the variable with respect to age being allowed to vary on either side of the age cutoff. For comparison, Table 4 also reports each variable's mean.

As Table 4 shows, there is strong evidence of a significant jump in the Math test scores at the age cutoff. There is somewhat less evidence for a jump in Vocabulary test scores at the age cutoff, and marginal evidence for a jump in Letter-Word ID test scores at the cutoff. There is no sign of any significant change in the behavioral assessment scores at the cutoffs, and the point estimates actually go in the wrong direction.

None of the socioeconomic controls show a statistically significant jump at the age cutoff. The variable with a jump that is closest to statistical significance is the log of family income. But this jump is much less pronounced when examining only cases where the income variable is not missing. Therefore, it is possible that because the pre-K graduate cohort has a higher percentage of missing-income families, this might be biasing the estimated jump in the income variable.

The lack of statistically significant jumps at the age cutoff in observed socioeconomic variables does not mean that unobserved variables might not jump at the age cutoff. But it is consistent with that hypothesis.

Table 4 Variation in Test Scores and Demographic Characteristics at the Age Cutoff for Pre-K Participation the Previous Year

Variable	Mean	Jump at age cutoff	Probability of jump being different from zero
Vocabulary	133.60	7.92	0.043
Letter-word ID	12.39	3.95	0.144
Math	14.53	3.52	0.003
Total protective factors	81.99	−1.30	0.807
Behavioral concerns	7.09	1.59	0.311
Minority	0.53	−0.12	0.503
Female	0.38	−0.15	0.422
Single-parent	0.18	−0.08	0.558
ln(income)	11.06	0.42	0.089
ln(income) (missing excluded)	11.06	0.22	0.390
Missing income	0.22	0.21	0.252
Mother's education	15.38	0.60	0.565
Father's education	15.09	1.43	0.177
Mother's age	34.27	2.29	0.218
Father's age	37.10	3.03	0.215

NOTES: The mean reported is overall sample mean for that variable. The reported jump at the age cutoff is the coefficient on a dummy “treatment” variable for whether that child was old enough to participate in pre-K the previous year. This regression also includes the age differential from the cutoff, and the age differential interacted with the dummy variable for the treatment dummy. The probability that is reported is the probability of a *t*-test that high in absolute value for the *t*-statistic on the coefficient for the dummy treatment variable. No other controls are included in these regressions. The probabilities and *t*-statistics are based on robust standard errors and are clustered at the person level.

Although these results suggest that there may be some significant jumps in test scores at the age cutoff, the results do not control for the observed characteristics of the child or family. As mentioned, these observed characteristics do not show significant jumps at the age cutoff. Therefore, we should not expect large changes in results because of these controls. However, including these controls may be rationalized either to increase precision, or as an extra check against bias due to differences between pre-K entrants and pre-K graduates. Controlling for observed characteristics prevents possible bias due to these observed characteristics or other differences that are linearly predicted by these observed characteristics.

Table 5 prevents the full set of results for all the test score dependent variables. The specification allows for a linear effect of age on test scores, with the linear effect of test scores allowed to differ between the pre-K entrant group and the pre-K graduate group. In addition, this specification uses all the available data, with the ages of children ranging from one year before the age cutoff to one year after the age cutoff. Even with including the entire sample, we only have a maximum of 220 observations, which means that obtaining statistical significance is likely to require quite large estimated effects.

The specification allows for robust standard errors, which will be valid even if the error term is heteroskedastic. In addition, I allow the error term to be clustered for the same person. The 51 pre-K graduates are all also observed as pre-K entrants. Presumably there would be some positive correlation across test scores for the same child. Allowing for clustering by person avoids inappropriately assuming that all the error terms are independent.

Before turning to the effects of the treatment dummies, I consider the effects of age on test scores. Age has the expected effect on all test scores. The effect of age on test scores is statistically significant for all except the Behavioral Concerns assessment. This is consistent with the descriptive picture provided by Figure 2. It is also consistent with the overall regression discontinuity model. (If age did not have a significant effect on test scores, we could instead consider a simpler model that did not control for age.) The age trend is only significantly different before and after the age cutoff for the Vocabulary test score dependent variable, but in that case it is highly significant. For consistency, I included the different linear trends before and after the age cutoff for all the test score dependent variables.

Table 5 Estimated Effects on Test Scores of Pre-K, Regression Discontinuity Design, Controlling for Child's Age and Other Child/Family Characteristics

	Dependent variable				
	Academic tests			Behavioral assessments	
Independent variable	Vocabulary	Letter-word ID	Math	Total protective factors	Behavioral concerns
Treatment dummy	6.499*	2.411	2.738**	−4.411	2.422
	(1.69)	(1.10)	(2.58)	(−0.86)	(1.56)
Age of child (relative to cutoff age, in days; = 0 at age cutoff)	0.0418***	0.0175***	0.00893***	0.0196*	−0.00534
	(3.89)	(4.17)	(2.94)	(1.68)	(−1.57)
Age relative to cutoff age multiplied by treatment dummy	−0.0331**	0.0127	−0.00403	−0.0232	−0.00456
	(−2.02)	(1.18)	(−0.96)	(−1.02)	(−0.69)
Female child	2.137	−1.867*	−0.219	2.648	−1.188*
	(0.93)	(−1.91)	(−0.37)	(1.20)	(−1.92)
Minority child	−5.139**	−0.456	−0.801	−0.415	0.854
	(−2.04)	(−0.49)	(−1.22)	(−0.17)	(1.15)
Single-parent family	−3.382	−1.078	−0.698	−2.054	1.752
	(−0.87)	(−0.88)	(−0.64)	(−0.47)	(1.54)
ln (family income)	1.615	−0.696	0.474	2.853	−0.369
	(0.64)	(−0.81)	(0.82)	(1.19)	(−0.48)
Family income missing	−5.934*	1.078	−0.344	3.450	−0.744
	(−1.77)	(0.65)	(−0.40)	(0.99)	(−0.70)
Education of mom (in years)	1.029*	0.250	0.0253	0.143	0.152
	(1.78)	(1.08)	(0.15)	(0.24)	(0.86)
Education of dad (in years)	0.432	0.858***	0.406**	0.567	−0.164
	(0.90)	(3.45)	(2.60)	(1.00)	(−1.04)
Age of mom	−0.147	0.0947	0.0408	0.401	0.00145
	(−0.51)	(0.87)	(0.41)	(1.29)	(0.02)
Age of dad	0.172	−0.136	−0.0718	−0.186	−0.0968
	(0.74)	(−1.32)	(−0.84)	(−0.72)	(−1.27)
Constant term	99.02***	7.008	5.127	35.27	14.05**
	(3.76)	(0.87)	(0.95)	(1.58)	(1.98)
Sample size	220	220	220	218	218
adj. R-sq	0.310	0.530	0.375	0.115	0.114
Mean of dependent variable	133.6	12.39	14.53	81.99	7.087
Standard deviation of dependent variable	17.39	8.450	4.826	15.80	4.408

NOTE: The table shows coefficient estimates for each independent variable in predicting the test score and behavioral assessment dependent variables. *t*-statistics are in parentheses; * significant at 0.10 level, ** significant at 0.05 level, *** significant at 0.01 level. Standard errors used to calculate *t*-statistics are clustered at person level and robust.

Most of the other individual characteristics were not statistically significant, which may reflect the small sample size. Minority students did somewhat worse on vocabulary, as did students with missing family income. Education of mom or dad was sometimes associated with higher test scores.

The effects of the “treatment dummy”—a zero/one variable for whether the child was a pre-K graduate/kindergarten entrant (treatment = 1) versus the child being a pre-K entrant (treatment = 0)—was highly statistically significant and positive for the Math test scores. Pre-K participation had marginally significant effects on Vocabulary test scores. The pre-K effect on Letter-Word ID was positive but not close to being statistically significant. For the two behavior variables, pre-K participation had the “wrong” sign in the point estimates, but these point estimates were not statistically significantly different from zero.

The linear specification with all data included is not the only possible specification for a regression discontinuity model using these data. I also tried some other plausible models. For instance, I tried a model in which age was allowed to have a quadratic effect on test scores, with the added quadratic terms allowed to differ before and after the age cutoff. These added quadratic terms were never statistically significant.[11] In addition, as another way of avoiding some of the limitations of a linear specification, I estimated another linear specification, but with only observations within 183 days (one-half year) of the age cutoff. The linear specification is less restrictive if we consider a narrower range of dates. Moving to the half-year specification roughly cuts the sample size in half, which reduces the precision of estimates.

Finally, one might be concerned about remaining unobserved variables. To deal with this, we can estimate the model using only the children observed twice in the sample, once as pre-K entrants and the other time as pre-K graduates. Doing this also restricts the sample size (to 51 observations times 2), which makes it harder to detect significant effects.

[11] The probabilities of the *F*-test on the two quadratic terms in age for each test score dependent variable are: Vocabulary, 0.7458; Letter-Word ID, 0.8420; Math, 0.6252; Total Protective Factors, 0.1264; Behavioral Concerns, 0.1835.

Table 6 presents the results for the treatment dummies for these four possible specifications: 1) the one-year linear model already estimated, 2) the quadratic one-year model, 3) the six-month linear model, and 4) the panel model (also assumed to be linear). The estimated models also include controls for all the child and family characteristics used before.

Table 6 "Treatment" Effects of Pre-K Participation—Four Alternative Specifications

	Linear 1-year estimates	Quadratic 1-year estimates	183-day linear estimates	Panel estimates
Vocabulary				
Treatment effect	6.50*	3.33	8.04*	6.24
t-stat	1.69	0.64	1.80	1.05
Effect size	0.60	0.31	0.74	0.58
Percentile (%)	22.6	12.1	27.1	21.7
Days (%)	42.6	21.9	40.3	36.2
Letter-word ID				
Treatment effect	2.41	3.68	2.26	2.54
t-stat	1.10	1.40	0.86	0.99
Effect size	0.25	0.38	0.23	0.26
Percentile (%)	9.9	14.9	9.2	10.4
Days (%)	38	57.8	51.4	40.4
Math				
Treatment effect	2.74**	3.72***	3.79***	3.76***
t-stat	2.58	2.69	2.64	2.95
Effect size	0.70	0.95	0.97	0.96
Percentile (%)	25.9	33.0	33.5	33.3
Days (%)	84.0	114.1	166.3	111.9
Total protective factors				
Treatment effect	−4.41	−0.90	2.31	1.13
t-stat	−0.86	−0.14	0.37	0.16
Effect size	−0.28	−0.06	0.15	0.07
Percentile (%)	−11.0	−2.3	5.8	2.8
Days (%)	−61.7	−13.1	6.5	32.7
Behavioral concerns				
Treatment effect	2.42	0.08	−0.53	1.65
t-stat	1.56	0.04	−0.25	0.80
Effect size	0.51	0.02	−0.11	0.35
Percentile (%)	19.5	0.7	−4.5	13.6
Days (%)	−124.2	−4.4	19.5	−122.0

NOTE: Linear specification already presented in Table 5. Quadratic specification adds in squared terms in age differential from cutoff, and that squared term interacts with treatment dummy. Half-year specification returns to linear specification, but restricts sample to plus or minus 183 days of age cutoff. Panel estimates are also linear but only include children observed at both pre-K entrance and kindergarten entrance. *T*-statistics are based on robust standard errors that allow for correlation of error term for same person. The other statistics presented are explained in the text. * significant at 0.10 level, ** at 0.05 level, *** at 0.01 level.

As shown in Table 6, the math effects seem quite robust to different specifications. In a wide variety of specifications, pre-K participation has statistically significant effects on increasing Math test scores. The Vocabulary effects of pre-K jump around a bit more, and vary from being marginally statistically significant to not at all significant. Letter-Word ID's effects do not vary much by specification, but are never statistically significant. Finally, the behavioral assessment results jump around a great deal in the different specifications.

How big are these effects? There are a variety of ways to assess this, which are also explored in Table 6. I calculated effect sizes for the different estimated effects of treatment size. This simply divides the treatment dummy coefficient by the standard deviation of the dependent variable. The standard deviation used was the standard deviation of the appropriate test score variable among the kindergarten entrants we observe (see Table 3). To give more statistical meaning to this effect size, I translated it into how many percentiles this would represent if this test score increase ended at the fiftieth percentile. Finally, I divided the test score coefficient by the estimated effect of one year of aging during the preschool year. In a linear model, this is the coefficient on age in the pre-K entrant group, multiplied by 365 days.[12]

As one would expect, the estimated math effects imply quite large effect sizes, of 0.70 or greater. These effects also represent a large percentage of what aging by itself would do during the pre-K year. The Math test score effects represent from 84 percent to 166 percent of what aging by itself seems to do during the pre-K year. For example, this 84 percent statistic implies that because of pre-K, math learning during that year is 184 percent of what it otherwise would be without pre-K.

Some of the other point estimates of test score effects are also quite large. For the academic test scores, even though many of the effects are at best marginally significant, we have

[12] In the quadratic model, this is the effect on test scores of going from age = −365 to age = 0.

quite large effect sizes and annual learning gains. For example, for Letter-Word ID, the estimated effect size ranges from 0.23 to 0.38, and pre-K is estimated to increase learning by 38–58 percent. These effects are not statistically significant, but they are large. Similarly large effects are found for Vocabulary.

The behavioral assessment effects range all over the place. These effects are sometimes large and sometimes small, and vary in whether they have the right sign.

I also wanted to compare these test score effects with other studies. Therefore, I averaged the three academic test score effects in Table 7. This is similar to what was done in the previous preschool studies summarized in Table 1.

Table 7 Average Effects of KC Ready 4s on Academic Tests (Vocabulary, Letter-Word ID, Math)

	Averages (3 academic tests)			
	Linear	Quadratic	183 days	Panel
Effect size	0.52	0.55	0.65	0.60
Percentile (%)	19.4	20.0	23.3	21.8
Days (%)	54.8	64.6	86.0	62.8

NOTE: This table reports unweighted means of estimates reported in Table 6 for the three academic tests.

These average academic test score effects are within the range of other pre-K studies, but toward the high end. The average effects size varies from 0.52 to 0.65. And on average, across these three academic tests, pre-K appears to increase student learning during the year before kindergarten by 55–86 percent. There is no sign that the effect size or percentage learning effects diminish as I consider more general specifications that make fewer assumptions; in fact the reverse is the case. Allowing a quadratic, or focusing on just a 183-day window to either side, or allowing for biases due to unobserved effects in persons observed only once, are all moving to more general and less restrictive specifications. The effect sizes and percentage learning boosts seem to go up slightly in these more general specifications.

The percentile effects can be used to give a rough benefit-cost analysis of KC Ready 4s. The percentile effects are in the range of 19 to 23 percent. Based on Chetty et al. (2011), a 1 percentile increase in kindergarten test scores increases adult earnings by about one-half of 1 percent. Therefore, we would expect adult earnings to increase by at least 9 percent. In my Tulsa study with Gormley and Adelstein (2012), we calculated that even for the most low-income groups, expected future adult earnings, discounted back to age four, would be at least $260,000. A 9 percent boost would represent over $23,000 in present-value gains, evaluated at age four. As KC Ready 4s provides a maximum subsidy of $4,500 per child, the implied benefit-cost ratio is over five-to-one. This ignores any possible effects in reducing crime, reducing special education and remedial education costs, etc.

CONCLUSION

This study's results, from a regression discontinuity study of the KC Ready 4s program with a relatively small sample, suggest that the program has some positive effects on a child's skills, and that these effects are sometimes large. The effects are large in that they exceed effects estimated for some other pre-K programs, although the effects are not outside the bounds suggested by previous studies. The effects are also large in that they imply benefits that considerably exceed the costs of KC Ready 4s.

On the other hand, because of the small sample size, the results have much statistical uncertainty. The test score increases are only consistently positive and statistically significant for math. Other positive test score effects are frequently not statistically significant. In addition, no strong evidence of improvements in behavior was found.

KC Ready 4s currently plans to continue collecting pre- and postprogram data on children and families participating in the program. These data will allow for greater precision in estimating the effects of this program. In addition to allowing greater precision in estimating overall effects, this continued data collection may allow better insight into which groups benefit the most from KC Ready 4s.

This study also illustrates that it is possible with a small, community-run program to provide rigorous evidence for program effects. Although collecting data to fit in with the regression discontinuity model has some financial and logistical costs—for example the need to track down program graduates in kindergarten—KC Ready 4s was able to meet this challenge. Local community programs can and will invest in evaluation if they understand how to do so and if the local community wants such information.

REFERENCES

Bartik, T. J., W. T. Gormley Jr., and S. Adelstein. 2012. "Earnings Benefits of Tulsa's Pre-K Program for Different Income Groups." *Economics of Education Review* 31(6): 1143–1161.

Camilli, G., S. Vargas, S. Ryan, and W. S. Barnett. 2010. "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." *Teachers College Record* 112(3): 579–620.

Campbell, F. A., E. P. Pungello, M. Burchinal, K. Kainz, Y. Pan, B. H. Wasik, O. A. Barbarin, J. J. Sparling, and C. T. Ramey. 2012. "Adult Outcomes as a Function of an Early Childhood Educational Program: An Abecedarian Project Follow-up." *Developmental Psychology* 48(4): 1033–1043.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics 126(4): 1593–1660.*

Currie, J., and D. Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85(3): 341–364.

Deming, D. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3): 111–134.

Dunn, L. M., and D. M. Dunn. 2007. *Peabody Picture Vocabulary Test.* 4th ed. Minneapolis: Pearson.

Garces, E., D. Thomas, and J. Currie. 2002. "Longer Term Effects of Head Start." *American Economic Review* 92(4): 999–1012.

Gormley, W. T. Jr. 2010. "Small Miracles in Tulsa: The Effects of Universal Pre-K on Cognitive Development." In *Childhood Programs and Practices in the First Decade of Life: A Human Capital Integration*, A. J. Reynolds, A. J. Rolnick, M. M. Englund, and J. A. Temple, eds. New York: Cambridge University Press, pp. 188–198.

Gormley, W. T. Jr., and T. Gayer. 2005. "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program." *Journal of Human Resources* 40(3): 533–558.

Gormley, W. T. Jr., T. Gayer, D. A. Phillips, and B. Dawson. 2005. "The Effects of Universal Pre-K on Cognitive Development." *Developmental Psychology* 41(6): 872–884.

Gormley, W. T. Jr., D. A. Phillips, and T. Gayer. 2008. "Preschool Programs Can Boost School Readiness." *Science* 320(5884): 1723–1724.

Gormley, W. T. Jr., D. A. Phillips, K. Newmark, K. Welti, and S. Adelstein. 2011. "Social-Emotional Effects of Early Childhood Education Programs in Tulsa." *Child Development* 82(6): 2095–2109.

Heckman, J. J., L. Malofeeva, R. Pinto, and P. A. Savelyev. 2010. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz. 2010. "The Rate of Return to the High/Scope Perry Preschool Program." *Journal of Public Economics* 94(1-2): 114–128.

Hustedt, J. T., W. S. Barnett, K. Jung, and L. D. Goetze. 2009. *The New Mexico Pre-K Evaluation: Results from the Initial Four Years of a New State Preschool Initiative—Final Report*. New Brunswick, NJ: National Institute for Early Education Research.

Hustedt, J. T., W. S. Barnett, K. Jung, and J. Thomas. 2007. *The Effects of the Arkansas Better Chance Program on Young Children's School Readiness*. New Brunswick, NJ: National Institute for Early Education Research.

Jacob, B. A., and L. Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics* 86(1): 226–244.

Leak, J., G. J. Duncan, W. Li, K. Magnuson, H. Schindler, and H. Yoshikawa. 2010. "Is Timing Everything? How Early Childhood Education Program Impacts Vary by Starting Age, Program Duration, and Time Since the End of the Program." Paper presented at the fall 2010 meetings of the Association for Public Policy Analysis and Management, held in Boston, MA, November 4–6.

LeBuffe, P. A., and J. A. Naglieri. 1999a. *Devereux Early Childhood Assessment: User's Guide.* Villanova, PA: Devereux Foundation.

LeBuffe, P. A., and J. A. Naglieri. 1999b. *Devereux Early Childhood Assessment: Technical Manual.* Villanova, PA: Devereux Foundation.

Lee, D. S., and T. Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48(2): 281–355.

Lipsey, M. W., D. C. Farran, K. G. Hofer, C. Bilbrey, and N. Dong. 2011 (March). "The Effects of the Tennessee Voluntary Pre-Kindergarten Program: Initial Results." Paper presented at the annual conference of the Society for Research on Educational Effectiveness, held in Washington, DC, March 3–5.

Ludwig, J., and D. L. Miller. 2007. "Does Head Start Improve Children's Life Chances: Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics* 122(1): 159–208.

Phillips, D. A., W. T. Gormley Jr., and A. E. Lowenstein. 2009. "Inside the Pre-Kindergarten Door: Classroom Climate and Instructional Time Allocation in Tulsa's Pre-K Programs." *Early Childhood Research Quarterly* 24(3): 213–228.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer. 2012. *Third Grade Follow-Up to the Head Start Impact Study: Final Report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Ramey, C. T., and F. A. Campbell. 1991. "Poverty, Early Childhood Education, and Academic Competence: The Abecedarian Experiment." In *Children in Poverty: Child Development and Public Policy*, A.C. Huston, ed. New York: Cambridge University Press, pp. 190–221.

Reynolds, A. J. 1995. "One Year of Preschool Intervention or Two: Does It Matter?" *Early Childhood Research Quarterly* 10(1): 1–31.

Reynolds, A. J., J. A. Temple, S. R. Ou, I. A. Arteaga, and B. A. B. White. 2011. "School-Based Early Childhood Education and Age-28 Well-Being: Effects by Timing, Dosage, and Subgroups." *Science* 333(6040): 360–364.

Reynolds, A. J., J. A. Temple, B. A. B. White, S. R. Ou, and D. L. Robertson. 2011. "Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program." *Child Development* 82(1): 379–404.

Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40.* Ypsilanti, MI: HighScope Educational Research Foundation.

Shager, H. M., H. S. Schindler, K. A. Magnuson, G. J. Duncan, H. Yoshikawa, and C. M. Hart. 2013. "Can Research Design Explain Variation in Head Start Research Results? A Meta-Analysis of Cognitive and Achievement Outcomes." *Educational Evaluation and Policy Analysis* 35(1): 76–95.

Weiland, C., and H. Yoshikawa. 2013. "Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills." *Child Development*. Forthcoming.

Whitehurst, G. J. 2013. "Can We Be Hard-Headed about Preschool? A Look at Universal and Targeted Pre-K." *The Brown Center Chalkboard* (blog), January 23. http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/01/23-prek-whitehurst (accessed April 24, 2013).

Wong, V. C., T. D. Cook, W. S. Barnett, and K. Jung. 2008. "An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs." *Journal of Policy Analysis and Management* 27(1): 122–154.