First in the class?
Age and the education production function

Elizabeth U. Cascio (Corresponding author)

Diane Whitmore Schanzenbach

Running head: First in the Class?

April 2015

Dartmouth College, Department of Economics, 6106 Rockefeller Center, Hanover, NH, USA 03755.
Tel: +1 603 646 4096. *Email Address:* Elizabeth.U.Cascio@dartmouth.edu

Northwestern University, School of Education and Social Policy, Annenberg Hall, 2120 Campus Drive, Evanston, IL, USA 60208.
Tel: + 1 847 491 3884. *Email Address:* dws@northwestern.edu

Acknowledgments: We thank Bruce Sacerdote for helpful conversations. We are also grateful to Joshua Angrist, Sandra Black, Kristin Butcher, David Card, Damon Clark, Ethan Lewis, Jens Ludwig, Heather Royer, Douglas Staiger, two anonymous referees, and seminar participants at Case Western Reserve University, Dartmouth College, the Federal Reserve Bank of San Francisco, the University of Florida, Wellesley College, the Association for Public Policy Analysis and Management Fall Conference, and the Society of Labor Economists Annual Meeting for comments on earlier drafts. All errors are our own.

Abstract

We estimate the effects of relative age in kindergarten using data from an experiment where children of the same age were randomly assigned to different kindergarten classmates. We exploit the resulting experimental variation in relative age in conjunction with variation in expected kindergarten entry age based on birthdate to account for negative selection of some of the older school entrants. We find that, holding constant own age, having older classmates on average improves educational outcomes, increasing test scores up to eight years after kindergarten and raising the probability of taking a college-entry exam. These findings suggest that delaying kindergarten entry, or so-called academic "redshirting," does not harm other children – and may in fact benefit them – consistent with positive spillovers from higher-scoring or better-behaved peers.

<A> 1. Introduction

There is a large literature on the so-called "relative age" effect in school, with results suggesting that being the oldest in the class is beneficial to a student. The concept was popularized by Malcolm Gladwell in his 2008 book *Outliers*. Over time, more and more parents – particularly in affluent areas – have chosen to delay their child's entry into kindergarten to give him or her a strategic advantage in school (Deming and Dynarski 2008; Bassok and Reardon 2013). This practice is commonly referred to as academic "redshirting."

The academic literature that has informed this phenomenon has been dominated by estimates of the relationship between the child's age and schooling outcomes such as performance on standardized tests. Influential papers in this literature, such as Bedard and Dhuey (2006), have interpreted the resulting reduced-form coefficient on age as a "relative age" impact. But this parameter is actually a composite of three separate effects: that of the child's age at school entry, her age at the time of the test, and her age relative to her peer group. Often times, these three aspects of age move together. For example, in this literature it is typical to compare end-of-year test scores of children with birthdays near school entry cutoff birthdates. In such a comparison, the child with the birthday after the cutoff (e.g., a child with an October birthday in state with a September 1 cutoff) is not only older at school entry and at the time of the test than a child with a birthday right before the cutoff (e.g., an August birthday), but also in relation to his classmates. Recent literature has attempted to disentangle the effects of age at school entry from age at test, finding small positive effects of starting school younger that are dominated by larger positive effects of age at test (Black, Devereux, and Salvanes 2011; Carlsson et al. 2014). Less attention has been paid to the final piece: a child's age in relation to his peers.

Estimating the effect of relative age is important in its own right for better understanding the education production function. But it is also important for understanding the spillover impacts of educational policies and practices. For example, a policy change that moves the school entry cutoff date from October 1 to September 1 will not change the school entry age for most children (i.e., those born between October and August), but it will expose them to relatively older peers and thus reduce their relative age. Similarly, children who enter school on time in an area where substantial numbers of parents opt to redshirt will have their relative age shifted by the other parents' choices to delay school entry. To understand the impacts in cases like these – that is, when own age is held constant but relative age shifts – it is important to have separate estimates of the effects of age and relative age.

To generate such estimates, we use data from Tennessee's Project STAR, a randomized experiment in which kindergarten students are randomly assigned to classrooms. As a result, their classmates' age distribution is random, and their own relative age is randomly determined. We can also follow children with different relative age rankings over time to observe how their achievement trajectories compare. We find that, contrary to conventional wisdom, being relatively young improves test scores. In other words, having older peers improves achievement across children of the same age. This finding implies that having redshirted classmates has positive spillovers on children who enter on time. Further, our results suggest that the reduced-form benefit of age found in the prior literature is primarily being driven by an absolute age effect (some combination of age at entry and age at test effects), which more than offsets the negative impact of having relatively younger peers.

<A>2. Background and Previous Literature

As briefly described above, most studies to date in previous literature (e.g. Bedard and Dhuey 2006; Datar 2006; Fredriksson and Öckert 2006; Puhani and Weber 2007; McEwan and Shapiro 2008; Elder and Lubotsky 2009) have generally estimated a reduced-form equation, similar to the following:

$$y_{it} = \alpha_{0t} + \alpha_{1t} a_i + \upsilon_{it}\,, \tag{1}$$

where y_{it} represents an outcome (e.g., test score) of child i at time t, a_i is his or her school entry age, and υ_{it} represents unobserved determinants of outcomes. But as described above, when in-school outcomes like test scores are used, a_i is typically perfectly collinear with age at test (a_{it}=a_i+t). It is also highly positively correlated with relative age. Thus, estimates of the coefficient on a_i capture the combined effects of school starting age, age at test, and relative age.

It is potentially helpful to think more precisely about these three distinct effects. First, older entrants have greater preparation for school – and may therefore perform better over both the short- and longer-term – because they have had more time to develop (e.g. they have had more months of life to play, hear language, etc.). This is the "age at school entry" effect. By the same token, however, older school entrants will always have more life experience, which may independently increase their academic performance. This is the "age at test" effect. The "relative age" effect, by contrast, stems from the role of the peer composition of a child's classroom in his achievement level. Higher-achieving peers may have positive spillover effects.[1] The effects of peer ability may load onto the age distribution of the class, because calendar age and test scores are highly correlated in early life. As a result, the age distribution of a student's classmates is also expected to contribute to achievement and may vary based on the child's own age.

Distinguishing between these effects is important because they have very different implications for educational policies and practices. For example, if α_{1t} largely reflects age at test, it would imply that redshirting or changes in school entry age regulations would have very little direct impact on children. It is therefore not surprising that recent work has focused on obtaining clean estimates of age at test effects (Black, Devereux, and Salvanes 2011; Carlsson et al. 2014). Likewise, if α_{1t} largely reflects a relative age effect, then postponing a child's school entry, either by redshirting or by changing school entry laws, has negative consequences for other children. On the other hand, if the parameter

[1] A number of studies have found that exposure to higher-achieving or less-disruptive peers has benefits for a child's own achievement and behavior. On the effects of exposure to higher-achieving peers, see Hanushek et al. (2003), Ding and Lehrer (2007), Duflo, Dupas, and Kremer (2011), and Imberman, Kugler, and Sacerdote (2012). Figlio (2007) and Carrell and Hoekstra (2010) provide evidence that exposure to more disruptive peers can be harmful. See Sacerdote (2011) for a complete review of this literature.

represents absolute age (whether driven by school starting age or age at test), peer spillovers from shifts in relative age may in fact be zero or even positive.

The model of interest is therefore the following:

$$y_{it} = \beta_{0t} + \beta_{1t} a_i + \beta_{1't} a_{it} + \beta_{2t} f(a_i, \mathbf{A}_{-i}) + \varepsilon_{it} , \tag{2}$$

where the function $f(.,.)$ represents relative age at the start of school, taking as arguments own school starting age a_i and the vector of school starting ages of child *i*'s peers, $\mathbf{A}_{-i}$. Model (2) thus distinguishes between the school entry age effect (β_{1t}), the age at test effect ($\beta_{1't}$), and the relative age effect (β_{2t}), which is a function of a child's own age and the ages of his classmates. In this paper, as in most in this literature, a_i is perfectly correlated with a_{it} (that is, $a_{it} = a_i + t$). We therefore refer to the combined effect of entry age and age at test ($\beta_{1t} + \beta_{1't}$) as the "absolute age" effect. ε_{it} represents unobserved determinants of outcomes.

In this application, we specify relative age as the difference in age (in fractions of a year) between a child and the average of her classmates. That is, $f(a_i, \mathbf{A}_{-i}) = a_i - \overline{a}_{-i}$, where $\overline{a}_{-i}$ represents the average age of student *i*'s classmates at the start of school. Thus, we assume that relative age is linear, continuous, increasing in own age, and decreasing in the average age of kindergarten classmates. This specification captures the zero-sum implications of manipulating relative age: in order for a given student to become relatively old, another student must become relatively young. Relative age, so defined, thus has implications for the distribution of achievement across students, but not its level. A key advantage of our data is that classmate age $\overline{a}_{-i}$ is randomly assigned within school, so our thought experiment compares two students of the same age who by random chance fall into different spots in the classroom relative age distribution.

With regard to previous literature, our approach is similar to that of Elder and Lubotsky (2009), who first pointed out the subtlety of interpreting the reduced-form parameter on age in model (1). Like us, Elder and Lubotsky (2009) separately estimate the effects of absolute age and relative age on end-of-year test scores and several other outcomes, However, they define relative age as the difference between a student's age and the average age of her *school*mates and use data from a sample, rather than the entire population of students within a school.[2]

Perhaps more importantly, there is an important (and somewhat nuanced) difference in the variation we use to identify the relative age impact, and therefore our results speak to different situations. The variation in peer age used by Elder and Lubotsky (2009) derives primarily from policy differences across states in minimum kindergarten starting ages. As a result, their estimates may in part reflect differences across states in other determinants of the pace of learning for relatively old children, such as the rigor of the kindergarten curriculum. Relatedly, Bedard and Dhuey (2012) estimate separate effects of increases in state-mandated minimum school entry ages for individuals whose school entry ages were likely directly affected and for individuals who were only affected indirectly – for example, through reductions in relative age – based on quarter of birth. However, they hesitate to interpret the latter estimates as "relative age" effects, since the educational experience of the entire cohort might have been changed by the policy shift. By contrast, the identifying variation in relative age in the current paper is from random differences in classroom age distributions. Our estimates thus hold constant the kindergarten curriculum and other factors: not only do we compare children of different relative ages within the same state at a given point in time, but also within the same school.

To be sure, the parameters estimated by Elder and Lubotsky (2009) and Bedard and Dhuey (2012) are important, particularly for policymakers considering changing regulations governing age at school entry. However, our approach produces estimates that are more relevant for parents deciding

[2] In an attempt to reconcile our estimates with theirs, as well as to provide insight into causal mechanisms, we also considered models where relative age was defined in relation to other children in the entire entry cohort of a child's elementary school, and not just his own classroom. The estimated effect of having older schoolmates was systematically larger than that of having older kindergarten classmates, but it was impossible for us to rule out that the effects of older peers in kindergarten classrooms drive the effects of older peers at the school level. However, these estimates are not identified using experimental variation, so should be viewed with caution.

whether to redshirt their child, and also for parents who are concerned about whether the redshirting behavior of others may harm their own child. Our results are broadly consistent with Elder and Lubotsky's, in that both papers find that the absolute age effect for test scores is strongly positive while being relatively old actually lowers test scores in the early grades. Some of our results conflict, however, possibly because of the difference in the type of variation used to identify the results. In particular, they find that relative age effects dissipate as a child advances through school, whereas we do not find this pattern.

<A>3. Empirical Strategy

Our primary estimates are based on the following model:

$$y_{itk} = \gamma_{0t} + \gamma_{1t} a_i + \gamma_{2t}\left(a_i - \bar{a}_{-ik}\right) + x'_{ik}\delta_t + \varepsilon_{itk}\,, \tag{3}$$

where y_{itk} represents the outcome of individual i in year t who was randomly assigned to classroom k in kindergarten, a_i is his age at the start of kindergarten (perfectly correlated with age at test), and $a_i - \bar{a}_{-ik}$ is his relative age, where $\bar{a}_{-ik}$ represents the average age of his kindergarten classmates. x_{ik} is a vector of controls, including fixed (predetermined) characteristics of the individual, other (predetermined) characteristics of his kindergarten classmates, characteristics of his teacher, and class size. The control vector also includes indicators for school attended in kindergarten, since kindergarten classrooms were randomly assigned within schools. ε_{itk} denotes unobserved determinants of outcomes. The coefficients of interest in model (3) are the absolute age effect, γ_{1t}, and the relative age effect, γ_{2t}.

Two identification problems typically arise when estimating model (3). First, the coefficient on relative age may be biased, since children may be endogenously sorted across kindergarten classrooms on the basis of the ages of other children. We address this by using a dataset in which students were randomly assigned to classrooms, thus making peer average age randomly assigned. Below we validate the random assignment assumption by showing that relative age is not systematically related to observables.

The second identification problem stems from endogeneity in a child's observed age in the kindergarten year. Some variation in age is due to differences in month of birth relative to the school entry cutoff. For example, in Tennessee at the time of Project STAR, a child was meant to be age 5 by September 30 in the year they enroll in kindergarten. Because we observe one cohort of students who entered school in September 1985, this implies that if everyone complied with the school entry rule, we would only observe students who were born between October 1, 1979 and September 30, 1980. Among children who comply with this requirement, a child born on October 1 would be 3 months older than a child born on January 1. We seek to identify the absolute age effect from this type of variation. However, not all children comply with the entry age rule, and we observe children who are born either before or after the expected range of birthdates. Students who are older than expected were likely retained in kindergarten or redshirted, and students who are younger than expected were likely enrolled early. Kindergartners with an age not as expected may differ in unobservable ways from on-time kindergartners. In particular, children who are a year older because they delayed entry or were retained are likely negatively selected, imparting a downward bias on OLS estimates of both absolute age and relative age effects.

To isolate differences in age that arise only from differences in birth month relative to the entry cutoff, we adopt the same empirical approach as previous researchers and construct an instrument for a_i (and for the a_i component of relative age) using information on a child's birthday and Tennessee's cutoff birthdate for starting kindergarten.[3] The instrument, "expected" age at the start of kindergarten (ea_i), uses variation in the month and day, but not the year, of kindergartners' birthdays. Expected age is highly predictive of actual age, as shown in Figure 1. The solid line in Figure 1 illustrates the relationship between birthday and ea_i, under the assumption that the school year begins September 1. The figure

[3] See, for example, Bedard and Dhuey (2006) and Elder and Lubotsky (2009), among others. A similar approach can be used to construct an instrument for years of completed schooling in models of youth test scores (e.g., Gormley and Gayer 2005; Cascio and Lewis 2006) and of adult outcomes (e.g., Dobkin and Ferreira 2010; McCrary and Royer 2011).

shows clearly that actual age on September 1, 1985 (daily averages of which are represented by the hollow circles) is strongly but not perfectly related to ea_i. In particular, actual age in fourth quarter birthdays is sometimes below the predicted age, suggesting that some parents pushed their child forward to start school early. In the first through third quarters, average actual age is a little higher than predicted age, suggesting that some of these children were retained or redshirted.

Our preferred estimates are those that use expected age as an instrument for absolute age. Two-stage least squares (TSLS) estimates of γ_{1t} will be consistent if ea_i is uncorrelated with unobserved determinants of achievement.[4] While we cannot test this assumption directly, we show below that expected age at the start of kindergarten generally does not predict observed characteristics of a child that are correlated with academic performance. These estimates capture local average treatment effects for compliers – children who enter kindergarten on time.

When it comes to estimating relative age effects, one could make a case for wanting to estimate those based on either the distribution of actual age or the distribution of expected age of a child's classmates. Because of random assignment to classes, both measures are uncorrelated with own age and with the error term in model (3). When we use the distribution of actual age of a student's classmates, some of the older kindergarten classmates will have lower than average ability for their age because they were held back by their parents or teachers on the basis of their perceived school readiness. Conversely, some younger kindergarten classmates will on average be relatively high ability for their age, signaled by the fact that they started school early. For us, this is the preferred estimate because it reflects actual variation in relative age from all sources, including academic redshirting. When we instead instrument for relative age using classmates' average *expected* age, we isolate a relative age effect that abstracts from the fact that some of the observed older students have been redshirted or

[4] A sufficient condition for this assumption to be satisfied is that birthday is randomly assigned, but it is possible to identify γ_{1t} under weaker assumptions, e.g., under the assumption that expected school-entry age is randomly assigned conditional on some flexible function of birthday that is smooth through the cutoff date. We have estimated a regression-discontinuity model like this as a robustness check and found similar results to those reported below (results available on request).

retained. This parameter is more comparable to that which has been estimated in the prior literature (e.g., Elder and Lubotsky 2009; Bedard and Dhuey 2012). Both sets of estimates are therefore interesting, and so we present both in our main results table below.

<A>4. Data

<B>) Sample and Summary Statistics

Project STAR was an experiment designed to study the effects of class size on student achievement. Kindergarten students and teachers in 79 Tennessee schools were randomly assigned to three different class types – small (with target enrollment of 13-17 students), regular (with target enrollment of 22-25 students), and regular with a full-time teacher's aide – in the fall of 1985.[5] This cohort was to have maintained its class type through third grade, after which all participants were returned to regular-sized classes. Random assignment of children to class types took place within schools. Our analysis exploits the fact that most, if not all, Project STAR participants would have been randomly assigned to classrooms, not just class types, as a result of the experimental design.[6,7]

We restrict our estimation sample to participants for whom each of the available background characteristics (birthday, race, gender, free lunch status in kindergarten) and kindergarten teacher characteristics (experience, education, and race) is observed.[8] However, to construct average age and

---

[5] Children entering the experiment in grades 1 through 3, either by moving into the school or having been retained in grade the previous year, were also added to existing classes through random assignment. Each of the 79 schools had enrollment sufficient to accommodate at least one class of each type and were thus slightly larger than the state average. To ensure sufficiently large samples of poor and minority children, Project STAR schools were also disproportionately drawn from inner cities. A comparison of Project STAR schools to other Tennessee schools is provided in Schanzenbach (2007).

[6] One reason to focus on kindergarten classmates is that non-random transitions across class types (and classrooms) became more problematic over time. Using administrative data for 18 Project STAR schools, Krueger (1999) found that only five of 1581 participants did not attend their initially assigned class type in kindergarten. As the experiment continued, however, ten percent of students made transitions across class types. Anecdotally, these class switches were largely the result of student misbehavior, which might plausibly be related to age at school entry or relative age.

[7] We are not the first to exploit the random assignment of children and teachers in Project STAR to classes to gain insights into the education production function (Dee 2004; Dee and Keys 2004; Whitmore 2005; Schanzenbach 2006; Chetty et al. 2011). Others have also exploited the random assignment of students and teachers to classes of different sizes to estimate peer effects (Boozer and Cacciola 2001; Graham 2008).

[8] This results in us dropping only 75 observations. The observations dropped are not significantly predicted, individually or jointly, by the instrumental variables for age and average age of peers.

other characteristics of kindergarten classmates, we use all available data on the Project STAR kindergarten cohort, not just those individuals with non-missing outcomes data.

Table 1 gives summary statistics on these students and their teachers in our estimation sample. Consider first the statistics on age, given in Panel A. As has been found in nationally representative data, children in our sample tend to be slightly older at the start of kindergarten than expected (5.43 years old versus 5.38 years old). Approximately 6 percent of students are older than expected at the start of kindergarten, due to either retention or delay, whereas 1.3 percent are younger than expected.[9] The standard deviation of relative age exceeds that of age and implies a standard deviation in the average age of a child's kindergarten classmates of about 1.11 months. The within-school variation in kindergarten classmate average age that we use for identification is of similar magnitude.

The Project STAR sample is not nationally representative. As shown in Panel B, nearly half of Project STAR participants received free or reduced-price lunch in kindergarten, and 33 percent were black. By comparison, only 15.4 percent of five year olds in the U.S. were black in fall 1985.[10] About 17 percent of kindergarten teachers were black, 35 percent had master's degrees, and 10 percent had less than 2 years of experience (Panel D). Students were roughly equally divided across class size types (Panel E).

Our main outcomes come from tests administered to STAR participants through the end of high school. In the spring of kindergarten, STAR participants were administered the Stanford Achievement Test. For participants still attending public school in Tennessee, we have scores on the Comprehensive Test of Basic Skills (CTBS) in grades 5 through 8, regardless of year attended.[11] Both tests are multiple-

---

[9] We cannot distinguish retention from delay in our data. Note that the percent of children older than predicted in kindergarten in Project STAR is very similar to that observed in more recent cohorts of kindergartners (Bassok and Reardon 2013). However, it is lower than the fraction of children retrospectively reported by parents to have delayed school entry from the 1970s to the 1990s (West, Meek, and Hurst 2000; Lincove and Painter 2006).

[10] There are authors' calculations from the 1985 October CPS School Enrollment Supplement.

[11] This is true as long as a child attended grades 5 through 8 at some point between 1990-91 and 1996-97. Test scores were also collected in 1989-90, but are not available for a large, non-random subset of children who attended school in Memphis because the tests were not universally given there in that year.

choice standardized tests with reading and math components. We average the reading and math scale scores on each test, then standardize this average to have a mean of zero and a standard deviation of one using data on all STAR participants with non-missing test scores in a given year. Thus, coefficient estimates in the test-score models are in standard deviation units. Table 1, Panel F shows that the kindergartners in our sample score on average 0.12 standard deviation above the mean in spring 1994, when they would have been in eighth grade with normal school progression. Kindergarten attendees may have been slightly positively selected because it was not mandatory for the STAR cohort to go to kindergarten.

Our analysis focuses on test scores at the end of kindergarten (in spring 1986) and in spring 1994, when STAR participants progressing through school normally would have been completing eighth grade.[12] We focus on the spring 1994 scores because many existing studies have considered the relationship between age and test scores in eighth grade. We present estimates for the year that the cohort was expected to be in eighth grade instead of eighth-grade test scores because our sample includes individuals in the same school-entry cohort, and we wish to have our estimates be consistently interpretable across tests taken at different points in time. Because either one of the treatment variables may have affected grade progression, we also estimate separate models for whether a child was enrolled below eighth grade when tested in spring 1994. By this measure, one quarter of students tested are categorized as having been retained (Panel F).

Our final outcome measure is an indicator for whether a respondent took the ACT or SAT college-entrance exam. Researchers gathered college-entry test information on STAR participants from graduating classes through 1999 (i.e. for students who graduated early, on-time, or no later than one

[12] Scores on the Stanford Achievement Test are also available for STAR participants in grades 1 through 3 who did not leave a STAR school or repeat or skip a grade during the experiment. While we do present estimates for these grade levels (Figures 2 and 3), we do so with some caution because our instrumental variables are related to attrition during the experiment, most likely due to impacts on grade repetition (see below). In an attempt to address selective attrition from the sample in these grades, we impute test scores in grades 1 through 3 with a child's place in the test score distribution (z-score) in kindergarten.

year behind “normal” grade progression) from all high schools in the U.S.[13] Perhaps not surprisingly, individuals in our sample are less likely to have taken the ACT or SAT (47 percent) than individuals in the U.S. overall (Table 1, Panel F).[14]

<B>) Is the Variation in the Instruments Exogenous?

Before turning to the estimates, it is useful to provide evidence that our instruments for absolute and relative age are unrelated to unobserved determinants of test scores, a necessary assumption for identification. To this end, the first three columns of Table 2a give the coefficients on absolute age and relative age, ea_i and $ea_i - \bar{a}_{-ik}$, along with their joint significance, in models where observed characteristics are the dependent variables. Columns 4 through 6 do the same, but replacing the measurement of relative age with the predicted relative age alternative, $ea_i - \overline{ea}_{-ik}$. In Table 2b, we present the results from similar exercises where the dependent variables are instead indicators for whether key outcomes are observed.[15] The underlying regressions also include school fixed effects, because random assignment of children to class types took place within schools. Standard errors are consistent for heteroskedasticity and correlation of error terms among children in the same kindergarten classroom.

The estimates largely serve to validate our identification strategy and support the assumption that students were randomly assigned to classrooms.[16] As shown in the first three columns of Table 2a, in only one case are the baseline instruments individually significant predictors of an observable – having a kindergarten teacher with an MA (columns 1 and 2) – but these relationships are only marginally statistically significant, and teacher education is not a strong correlate of academic outcomes

[13] See Krueger and Whitmore (2001) for more information about how the CTBS and college entrance exam data were collected. Observation of other outcomes, like high school grades, is selected on our instruments.
[14] For example, using the National Education Longitudinal Study, Bedard and Dhuey (2006) report an ACT/SAT test-taking rate of 60 percent.
[15] We assume that ACT or SAT test-taking is observed for all Project STAR participants.
[16] The specification checks in Chetty et al. (2011), which include a wider range of background characteristics that are not available to us in this study, also support random assignment to kindergarten classrooms in Project STAR.

in these data. For our preferred set of instruments, the instruments are jointly significant for only one variable – whether a child was eligible for free or reduced-price lunch in kindergarten (*p*=0.018, column 3). Consistent with these findings, including all of these variables as controls has little impact on the TSLS point estimates, as shown below. Table 2b shows that the instruments also do not predict observation of our dependent variables, individually or jointly, using either identification strategy. This suggests that observations with missing values of dependent variables are random with respect to the identifying variation in absolute and relative age and that the estimates presented below are not biased by sample selection.

<A>5. Results

<B>) Conventional Estimates

To provide a useful benchmark to the prior reduced-form literature, Table 3 presents OLS and TSLS estimates of the coefficients on own age from a version of model (3) that omits relative age. The dependent variables include the standardized average of math and reading scores at the end of kindergarten and eight years later, when children progressing through school normally would have been in spring of eighth grade. Both the OLS and TSLS estimates are also plotted in Figure 2, along with TSLS estimates for test scores in the intervening grades.[17] Table 3 also includes estimates of age gradients in our proxy for grade retention – whether a respondent is enrolled in grade lower than expected in 1994 – and in taking a college-entrance exam. All specifications include fixed effects for school attended in kindergarten, and estimation accounts for heteroskedasticity and arbitrary correlation of error terms within kindergarten classrooms.

Consider first the estimates for test scores at the end of kindergarten, in column 1. The OLS estimates (Panel B) imply that STAR participants who were one year older at the start of kindergarten

[17] Because relatively young students are more likely to be retained, they are more likely to be missing from the test score data for grades 1 through 3. To address missing data, we carry forward the observed test score (z-score) from kindergarten to impute missing scores. Note that we omit grade 4 because of large amounts of (nonrandom) missing data that year.

scored on average about 0.24 standard deviation higher at the end of kindergarten (column 1). However, these estimates will be biased downward if, as we expect, children previously retained or delayed in entering kindergarten are negatively selected on unobservables. As described above, we confront this possibility by comparing children who *should* have entered kindergarten with a one-year difference in age, given their birthdays; if birthday is randomly assigned, these children will be on average identical in all other ways such as inherent ability. As shown in Panel A of the same column, TSLS estimates using expected age at kindergarten entry as an instrument imply that the test-score differential between two otherwise identical children who enter kindergarten with a one-year difference in age is a significantly higher 0.71 standard deviation.[18] Adding controls for predetermined individual and classroom characteristics (column 2) does not appreciably change this estimate. This estimate is comparable to those previously found in nationally representative data for the U.S.[19]

The test score differences between older and younger school entrants decline as children progress through school.[20] For example, the TSLS estimate of the age gradient in test scores is 0.22 standard deviation in spring 1994, when children progressing through school normally would have been in eighth grade (column 3). Figure 2 shows that most of the decline in the gradient occurred by the end of elementary school, as the TSLS estimates for grades 5 and 6 (with normal grade progression) are not much larger in magnitude. Older school entrants are also significantly less likely to have been retained at

[18] First-stage estimates for the specifications in Tables 3 and 4 are presented in Tables A1 and A2, respectively.

[19] For example, applying a similar identification strategy to data on a more recent kindergarten cohort, Elder and Lubotsky (2009) find that an additional year of age at school entry is associated with a 0.87 standard deviation difference in math test performance and a 0.61 standard deviation difference in reading test performance in the spring of a child's kindergarten year. When we estimate separate models by subject on our data, we arrive at TSLS estimates (with additional controls) of 0.69 standard deviation and 0.524 standard deviation for math and reading, respectively, at the end of kindergarten. Like us, Elder and Lubotsky (2009) and Bedard and Dhuey (2006) also find that OLS estimates of the coefficient on age are significantly lower than their TSLS counterparts, suggesting that students who are older by actual age (but not predicted age) are negatively selected.

[20] The test score effects of being older in kindergarten thus appear to "fade out" as a child ages. However, Cascio and Staiger (2012) point out that if a standard deviation in test scores maps into a larger difference in knowledge in later grades – i.e., if knowledge accumulates faster than it depreciates – some of this observed "fadeout" on test z-scores might be a statistical artifact. Their estimates (column 4, Table III) imply that the effect size for spring 1994 z-scores should be scaled up by a factor of roughly 1.22 to be comparable to that for spring of kindergarten. Thus, some of the observed fadeout in our context is not a statistical artifact.

any point in the eight years following kindergarten (column 4). Again, these estimates are in line with what has previously been documented for test scores and grade retention at the end of eighth grade in the U.S.[21] Unlike previous research, however, we do not find that older school entrants are more likely to take a college entrance exam (column 5).[22] In general, however, the consistency of our estimates with the existing literature suggests the possible broader applicability of inferences made from our data.

<B>) Full Model Estimates

As discussed above, the reduced-form estimates presented in Table 3 capture the combined effects of absolute age and relative age. In Table 4, we present estimates of model (3), which allow us to separately identify these effects. Our preferred TSLS estimates, in Panel A, identify own age effects from only variation in month and day of birthday – that is, they account for negative selection among students who were retained or redshirted – but measure peer effects relative to the actual age distribution. That is, in Panel A we use ea_i as an instrument for absolute age and $ea_i - \bar{a}_{-ik}$ as an instrument for relative age, are in Panel A. In Panel B, we present estimates from the alternative TSLS model keeping all else the same but identifying peer effects based on predicted age, i.e. using $ea_i - \overline{ea}_{-ik}$ as an instrument for relative age. For completeness, OLS estimates are presented in Panel C.

The first thing to notice in Table 4 is that all of the estimates for relative age effects on test scores (columns 1 to 3) are *negative* – a finding which holds true across both TSLS approaches and in OLS.[23] This suggests that being relatively old, holding own age constant, harms student outcomes.

[21] Elder and Lubotsky (2009) and Bedard and Dhuey (2006) find math test score differences of approximately 4 percentile points, or roughly 0.13 standard deviation, between eighth graders who entered school in the late 1970s with a one-year difference in age. When we re-estimate the model for actual eighth grade test scores (regardless of year attended), we find that a one-year difference in entry age is associated with a 0.12 standard deviation difference in math test performance. Elder and Lubotsky (2009) also find that a one-year increase in age at school entry lowers the likelihood of having been retained by eighth grade by 15.1 percentage points.

[22] Bedard and Dhuey (2006) find that being one year older (in eighth grade) raises the probability of taking the ACT or SAT by 11.1 percentage points. However, our sample also has a higher minority share and is poorer than the national average (Table 1), and their estimates are for individuals who were in eighth grade in the same year, not individuals who started kindergarten at the same time.

[23] We arrive at similar conclusions when we use a child's kindergarten schoolmates as a peer group, but the estimates are less precise.

Accordingly, controlling for relative age *increases* estimates of the coefficient on school entry age (compare column 1 in Tables 3 and 4). While the standard errors on our preferred TSLS estimates in Panel A do not allow us to reject a positive relative age effect for test scores at the end of kindergarten, we can reject a positive effect for test scores eight years later.

Figure 3 shows how the TSLS coefficients on age and relative age evolve across the grade levels, based on our preferred estimation strategy (as in Panel A of Table 4). The figure shows that the point estimate on relative age declines (increases in magnitude) as the cohort ages, and by sixth grade we can reject that relative age has a positive impact on test scores.[24] On the other hand, the coefficient on absolute age is positive and relatively stable in the years subsequent to kindergarten. In general, our findings for test scores are thus inconsistent with the interpretation of reduced-form age gradients in test scores, such as those presented in Table 3 or in Figure 2, as (positive) relative age effects.

Turning to magnitudes, our estimates imply that a one-year increase in relative age in kindergarten lowers test scores eight years later by about 0.43 standard deviation. However, a one-year increase in relative age is a relevant margin only for a child who delays school entry, and such a child would also be one year older in absolute terms. The net impact of redshirting is the sum of both coefficients, which is positive and consistent with the estimate of a positive net effect of age for spring 1994 test scores (Table 3, Panel A, column 3). We therefore would not conclude that delaying kindergarten entry is harmful to the academic achievement of the child experiencing the delay in the following sense: the child's performance in a given grade is predicted to be higher if the child has been redshirted. However, this is driven by the fact that the child will be a year older in absolute terms at the time of the test and not because the child is relatively old. But in fact, at any given age, the child's achievement will be harmed by being among the oldest in the class.

[24] As mentioned above, Elder and Lubotsky (2009) find the opposite pattern using a different identification strategy, with their estimates of relative age effects decreasing in magnitude across grades. When we use our alternative instrument for relative age – which is closer to their approach – we see a similar pattern to theirs, though the estimates are noisy.

Importantly, the estimates in Table 4 allow us to rule out negative effects for a given child of other families' redshirting decisions. In fact, we can conclude with some confidence that there are positive spillovers from this practice.[25] Consider how classmates' redshirting decisions change the relative age of a child with a birthday outside of the typical redshirting range, such as a child born in January in a state with a September 1 school entry cutoff. While the January-born child is not likely to be redshirted himself, i.e. his own age will be unaffected, his classmates' decisions about whether to redshirt will alter his relative age. For example, if half of classmates born in July or August opt to redshirt, then on average the January-born child's relative age declines by 1 month.[26] Other plausible redshirting "rules" – such as only boys born in summer redshirt, or only boys born in summer who will also be the youngest in the class redshirt – yield an average decline in relative age of about 0.5 to 1.2 months for a child born in January. Note that the redshirting rules we used in our simulations yielded an overall redshirting rate between 4 and 9 percent, similar to rates observed in recent years (Deming and Dynarski 2008; Bassok and Reardon 2013).

According to our estimates, holding one's own age constant, a one-month *decline* in relative age in kindergarten – the margin of variation typical in our data and the likely shift induced by others' redshirting behavior – *raises* test scores eight years later by a statistically significant 0.036 standard deviation. In other words, classmates' redshirting behavior is predicted to benefit the student. Likewise, a one-month decline in relative age in kindergarten increases the likelihood of taking the ACT or SAT by about 1.3 percentage points (significant at the 10 percent level, column 5). This is a larger effect than would be predicted on the basis of the relationship between spring 1994 test scores and college-exam

[25] Recall that the variation in relative age used to identify this parameter is relatively narrow – a matter of a few months – and it would be inappropriate to extrapolate the results to much larger differences in relative age such as a mixed-grade classroom.
[26] The exact decline in relative age depends on the distribution of birth dates. We conducted simulations in which classes of size 20 were randomly drawn from the birthday distribution, and estimated the impact on relative age of a January-born student if half of students born in July or August opt to redshirt. The average decline in relative age from this behavior was 1.04 months, with a standard deviation of 0.77 months.

taking alone, suggesting that the benefits of being relatively young continue to accumulate through high school or are working through other (non-cognitive) channels.[27]

The corresponding estimates in Panel B, which use variation in the expected entry ages of classmates instead of their actual ages, are noisier, but are broadly consistent with these findings. If anything, however, they imply larger negative impacts of relative age – and negative impacts that emerge as early as the end of kindergarten. In particular, a one-month decline in relative age in kindergarten increases test scores by a marginally significant 0.06 standard deviation (0.724/12; column 2). This finding serves to reinforce the interpretation that the mechanism for the negative relative age effects is that children benefit from having higher-achieving, better-behaved peers. In other words, while the first instrumental variables approach does not remove the negative selection of some older kindergarten classmates, the second instrumental variables approach identifies the peer component using variation from children with abilities and behavior consistent with their ages and finds larger point estimates, especially in the kindergarten year.[28]

<A>6. Heterogeneity

The likelihood of delayed school entry is more common for some subgroups in the population. For example, past research has found that boys are much more likely to be held back than girls. Children from higher-income families are also more likely to start kindergarten at an age older than expected (West, Meek, and Hurst, 2000; Bassok and Reardon, 2013). We see a similar pattern in our data. For

[27] A 1 standard deviation increase in spring 1994 test scores is associated with a 23.7 percentage point increase in the likelihood of taking the ACT or SAT, holding constant own, classmate, and kindergarten teacher and classroom observables. Thus, a 0.036σ increase in spring 1994 test scores is associated with a 0.8532 percentage point increase in the likelihood of taking the ACT or SAT – less than the 1.3 percentage point increase observed.

[28] The suggestion that negative selection of some older peers diminishes their positive spillovers is consistent with the results in Lavy, Paserman, and Schlosser (2012), who find that Israeli students are actually worse off when they have a higher share of peers in their school cohort who are overage for grade ("repeaters").

example, nearly 8 percent of boys, but only 3.7 percent of girls are older than expected in kindergarten.[29]

For boys and higher-income children, revealed preference thus suggests that being higher up in the classroom "pecking order" is more important for academic outcomes. If this were the case, relative age effects for these groups should be less negative, or even positive. At the same time, recent research on peer effects has found that girls are more responsive to higher-achieving peers than boys (Jackson 2012; Lavy, Silva, and Weinhardt 2012). This suggests that being relatively young may be more beneficial to girls. To investigate these possibilities, we estimated separate models by gender and free or reduced-price lunch status in kindergarten. We present the TSLS estimates by gender in Table 5 and by free lunch status in Table 6. We return to our baseline identification strategy, where we instrument for relative age using the difference between a student's expected age and the average actual age of her peers. Per our argument above, this approach should deliver estimates of relative age that are closer to being in the positive range. For the purposes of comparison and interpretation, each table includes subsample-specific means of the dependent variables.

We find in Table 5 that the relative age effect on test scores becomes more negative for boys as the cohort ages, and is statistically significant and negative for 8th grade test scores. Relative age is only significant for girls when looking at retention by 8th grade, and that estimate is only marginally significant. Turning to absolute age, in kindergarten its effect is similar for boys and girls.[30] After kindergarten, however, the absolute age effect drops substantially for girls and is relatively small and insignificant for taking the ACT or SAT. Among boys, the absolute age effect is stronger throughout, and is still marginally statistically significant for taking the ACT or SAT. These results suggest that boys may be more sensitive to having higher-performing peers, contrary to the existing literature.

[29] Students receiving free or reduced-price lunch in kindergarten are, however, more likely to be older than expected in kindergarten (7 percent versus 4.9 percent for children not lunch-eligible). This is likely due to the fact that we cannot distinguish in our data between kindergarten retention and delay.

[30] Recall that the reduced-form coefficient on age is approximately equal to the sum of the coefficients on absolute age and relative age, so the reader can back out approximate reduced-form coefficients for the subgroups in this section.

In Table 6, we separately examine effects by whether a student qualifies for free or reduced-price lunch. Relative age effects are stronger for low-income students, suggesting that being exposed to older, higher-scoring peers is more important for low-income students than for higher-income ones. At kindergarten entry, there is no difference in own-age effects between students who do and do not qualify for free or reduced-price lunch. Subsequently, however, the own-age impact remains strong only for low-income children, where it is a strong predictor of 8th grade test scores, grade retention, and whether a student takes a college-entrance exam. For higher-income students, the own-age effect declines and is not statistically significant in later grades. These results – stronger age effects for low-income students – stand in contrast to the observed patterns, in which higher-income children are substantially more likely to be redshirted.

<A>7. Conclusion

In this paper, we have estimated the net effects of having older peers – and being young relative to one's peers – using data from an experiment where children of the same age were randomly assigned to different classrooms in kindergarten. We find that, holding own age constant, children who were young relative to their kindergarten classmates generally performed no worse on achievement tests, were no more likely to be retained, and were no less likely to take the ACT or SAT. In fact, though some of our estimates are imprecise, having older classmates appears to have made the average child better off. And while our analysis of heterogeneity suggests that absolute age effects might be more present for boys – consistent with the higher incidence of delay for them – on net, boys and higher-income children still appear to benefit from having older kindergarten classmates.

These findings suggest the reduced-form age effects that have frequently been estimated in the literature on school entry age are better interpreted as reflecting the positive impact of being older in an absolute sense, rather than in a relative one. In other words, when we separate the reduced-form age parameter into absolute age and relative age, we find that, after holding own age constant, being old relative to one's peers is predicted to reduce a student's academic achievement. When a student is redshirted, the negative effect of being relatively old is outweighed by the positive effect of being absolutely older, yielding a positive overall (reduced-form) effect.

Furthermore, our findings suggest that the tendency for some parents to opt to hold back their child's kindergarten entry, so-called "redshirting", does not harm the student's classmates by making them relatively younger in the classroom age distribution. In fact, the spillover effects are predicted to be positive yet modestly sized. Likewise, our estimates imply that the consequences of increasing minimum school entry age laws for children whose own entry ages are not directly affected – for whom relative age declines but absolute age remains unchanged – are likely to be positive. However, our estimates have held school entry age regulations constant, and thus some caveats apply when generalizing to this scenario.

All in all, our findings are consistent with the broader peer effects literature documenting positive spillovers from having higher-quality peers. They add another dimension to the evidence on how kindergarten classrooms matter over the long term (Chetty et al. 2011): The ages of a child's kindergarten classmates – a feature of the kindergarten experience than can be easily manipulated by school principals – appear to have a lasting impact on his or her own school outcomes.

References

Bassok, Daphna, and Sean Reardon. 2013. 'Academic redshirting' in kindergarten: Prevalence, patterns, and implications. *Educational Evaluation and Policy Analysis* 35(3): 283-297.

Bedard, Kelly, and Elizabeth Dhuey. 2006. The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics* 121(4): 1437-1472.

Bedard, Kelly, and Elizabeth Dhuey. 2012. School entry policies and skill accumulation across directly and indirectly affected men. *Journal of Human Resources* 47(3): 643-683.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2011. Too young to leave the nest? The effects of school starting age. *Review of Economics and Statistics* 93(2): 455-467.

Boozer, Michael A., and Stephen E. Cacciola. 2001. Inside the 'black Box' of Project STAR: Estimation of peer effects using experimental data. Yale University Economic Growth Center Working Paper 832.

Carrell, Scott, and Mark Hoekstra. 2010. Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics* 2(1): 211-228.

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth. 2014. The effect of schooling on cognitive skills. *Review of Economics and Statistics*, forthcoming.

Cascio, Elizabeth U., and Ethan G. Lewis. 2006. Schooling and the armed forces qualifying test: Evidence from school entry laws. *The Journal of Human Resources* 41(2): 294-318.

Cascio, Elizabeth U., and Douglas O. Staiger. 2012. Knowledge, tests, and fadeout in educational interventions. NBER Working Paper No. 18038.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane W. Schanzenbach, and Danny Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics* 126(4): 1593-1660.

Datar, Ashlesha. 2006. Does delaying kindergarten entrance give children a head start? *Economics of Education Review* 25(1): 43-62.

Dee, Thomas S. 2004. Teachers, race, and student Achievement in a randomized experiment. *The Review of Economics and Statistics* 86(1): 195-210.

Dee, Thomas S., and Benjamin J. Keys. 2004. Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management* 23(3): 471-488.

Deming, David, and Susan Dyarski. 2008. The lengthening of childhood. *Journal of Economic Perspectives* 22(3): 71-92.

Ding, Weili, and Steven F. Lehrer. 2007. Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics* 89(2): 300-312.

Dobkin, Carlos, and Fernando Ferreira. 2010. Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review* 29(1): 40-54.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* 101(5): 1739-1774.

Elder, Todd E., and Darren H. Lubotsky. 2009. Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers. *The Journal of Human Resources* 44(3): 641-683.

Figlio, David. 2007. Boys named Sue: Disruptive students and their peers. *Education Finance and Policy* 2(4): 376-394.

Fredriksson, Peter, and Björn Öckert. 2006. Is early learning really more productive? The effect of school starting age on school and labor market performance. IFAU Working Paper No. 2006:12.

Gladwell, Malcolm. 2008. *Outliers: The story of success*. New York, NY: Little, Brown and Company.

Gormley, William T., and Ted Gayer. 2005. Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *The Journal of Human Resources* 40(3): 533-558.

Graham, Bryan S. 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* 76(3): 643-660.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18(5): 522-544.

Jackson, C. Kirabo, 2012. Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics* 96(1): 173-187.

Krueger, Alan B. 1999. Experimental estimates of education production functions. *The Quarterly Journal of Economics* 114(2): 497-532.

Krueger, Alan B., and Diane M. Whitmore. 2001. The effect of attending a small class in the early grades on college test-taking and middle school test results: Evidence from Project STAR. *The Economic Journal* 111(468): 1-28.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review* 102(5): 2048-2082.

Lavy, Victor, Olmo Silva, and Felix Weinhardt. 2012. The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics* 30(2): 367-414.

Lavy, Victor, M. Daniele Paserman, and Analia Schlosser. 2012. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal* 122(559): 208-237.

Lincove, Jane A., and Gary Painter. 2006. Does the age that children start kindergarten matter? Evidence of long-term educational and social outcomes. *Educational Evaluation and Policy Analysis* 28(2): 153-179.

McCrary, Justin, and Heather Royer. 2011. The effect of maternal education on fertility and infant health: Evidence from school entry policies using exact date of birth. *American Economic Review* 101(1): 158-195.

McEwan, Patrick J., and Joseph S. Shapiro. 2008. The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates. *The Journal of Human Resources* 43(1): 1-29.

Puhani, Patrick A., and Andrea M. Weber. 2007. Does the early bird catch the worm? Instrumental variable estimates of educational effects of age at school entry in Germany. *Empirical Economics* 32(2): 359-86.

Sacerdote, Bruce. 2011. Peer effects in education: How might they work, how big are they, and how much do we know thus far? In *Handbook of the Economics of Education* 1(3), edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessman, pp. 249-277. North Holland: Elsevier.

Schanzenbach, Diane W. 2006. Classroom gender composition and student achievement: Evidence from a randomized experiment. Mimeo, University of Chicago.

Schanzenbach, Diane W. 2007. What have researchers learned from Project STAR? *Brookings Papers on Education Policy* 2006: 205-228.

West, Jerry, Anne Meek, and David Hurst. 2000. *Children who enter kindergarten late or repeat kindergarten: Their characteristics and later school performance.* NCES 2000-039. Washington, DC: U.S. Department of Education.

Whitmore, Diane M. 2005. Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review* 95(2): 199-203.

Table 1. Descriptive statistics for Project STAR kindergarten cohort

	Mean (1)	Standard deviation (2)	N (3)
Panel A. Age variables			
Expected age at the start of kindergarten	5.38	0.28	6248
Age in at the start of kindergarten	5.43	0.35	6248
Above expected age at the start of kindergarten	0.059	-	6248
Below expected age at the start of kindergarten	0.013	-	6248
Relative age: age - average age of peers	0.000	0.35	6248
Relative age: expected age - average age of peers	-0.051	0.29	6248
Relative age: expected age - average expected age of peers	0.00	0.28	6248
Panel B. Demographic/SES variables			
Female	0.49	-	6248
Black	0.33	-	6248
Free/reduced-price lunch (in kindergarten)	0.49	-	6248
Panel C. Other characteristics of kindergarten classmates			
Fraction female	0.49	0.12	6248
Fraction black	0.33	0.41	6248
Fraction free/reduced-price lunch (in K)	0.49	0.28	6248
Panel D. Kindergarten teacher characteristics			
Black	0.17	-	6248
Has MA	0.35	-	6248
Has <2 years of experience	0.10	-	6248
Panel E. Kindergarten class characteristics			
Small	0.30	-	6248
Regular with side	0.35	-	6248
Panel F. Outcome variables			
Standardized test score, end of kindergarten	0.00	1.00	5719
Standardized test score, 8 years later	0.12	0.99	4436
Ever retained, 8 years later	0.25	-	4508
Took ACT/SAT	0.47	-	6248

Notes: Sample includes individuals with non-missing demographic/SES variables, kindergarten classmate characteristics, kindergarten teacher characteristics, and kindergarten class characteristics. See text for more details.

Table 2a. Predictive power of instrumental variables for background characteristics

	Coefficient (standard error):			**Coefficient (standard error):**		
Dependent variable:	**Expected age at start of K (1)**	**Expected age at start of K - average age of K classmates (2)**	***P* on joint significance (3)**	**Expected age at start of K (4)**	**Expected age at start of K - expected average age of K classmates (5)**	***P* on joint significance (6)**
	Demographic/SES Variables					
Female	0.069	-0.055	0.653	0.001	0.013	0.826
	(0.084)	(0.081)		(0.096)	(0.095)	
Black	0.000	0.006	0.809	0.026	-0.019	0.732
	(0.038)	(0.036)		(0.042)	(0.041)	
Free/reduced-price lunch (in K)	-0.101	0.048	0.018	-0.131	0.077	0.018
	(0.074)	(0.072)		(0.096)	(0.092)	
	Other characteristics of kindergarten classmates					
Fraction female	-0.078	0.076	0.602	-0.026	0.024	0.878
	(0.082)	(0.079)		(0.096)	(0.091)	
Fraction black	-0.011	0.012	0.666	0.021	-0.020	0.885
	(0.035)	(0.033)		(0.042)	(0.040)	
Fraction free/reduced-price lunch (in K)	0.032	-0.034	0.428	-0.131	0.127	0.028
	(0.072)	(0.069)		(0.095)	(0.091)	
	Kindergarten teacher characteristics					
Black	-0.217	0.201	0.279	-0.334	0.315	0.096
	(0.169)	(0.162)		(0.232)	(0.220)	
Has MA	-0.514*	0.498*	0.180	-0.319	0.303	0.616
	(0.289)	(0.277)		(0.371)	(0.352)	
Has <2 years of experience	-0.045	0.042	0.974	-0.050	0.047	0.781
	(0.220)	(0.210)		(0.288)	(0.274)	
	Kindergarten class characteristics					
Small	0.433	-0.404	0.379	0.594	-0.562	0.226

	(0.360)	(0.345)		(0.464)	(0.440)	
Regular with aide	0.137	-0.128	0.920	0.136	-0.127	0.296
	(0.387)	(0.371)		(0.459)	(0.435)	

Notes: Each row in each of columns (1) to (3) and (4) to (6) represents a different regression. All regressions are based on 6248 observations and include school fixed effects. Standard errors are clustered on kindergarten classroom. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2b. Predictive power of instrumental variables for observation of dependent variables

	Coefficient (standard error) on:			**Coefficient (standard error) on:**		
Dependent variable is dummy=1 if observed:	**Expected age at start of K (1)**	**Expected age at start of K - average age of K classmates (2)**	***P* on joint significance (3)**	**Expected age at start of K (4)**	**Expected age at start of K - expected average age of K classmates (5)**	***P* on joint significance (6)**
Standardized test score, end of kindergarten	-0.003	0.011	0.762	0.083	-0.073	0.376
	(0.052)	(0.049)		(0.062)	(0.061)	
Standardized test score, 8 years later	0.045	-0.014	0.281	-0.002	0.033	0.293
	(0.065)	(0.063)		(0.076)	(0.075)	
Ever retained, 8 years later	0.041	-0.010	0.281	0.012	0.019	0.297
	(0.065)	(0.063)		(0.076)	(0.075)	

Notes: Each row in each of columns (1) to (3) and (4) to (6) represents a different regression. All regressions are based on 6248 observations and include school fixed effects. Standard errors are clustered on kindergarten classroom. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3. Conventional estimates of the effect of school entry age

	Dependent variable:				
			Eight years later:		
	Average test score, end of K		Average test score	Ever retained	Took ACT or SAT
	(1)	(2)	(3)	(4)	(5)
	Panel A. Two-stage least squares[a]				
Age at start of kindergarten	0.707***	0.674***	0.215***	-0.189***	0.002
	(0.057)	(0.054)	(0.065)	(0.026)	(0.024)
Root MSE	0.900	0.855	0.890	0.404	0.465
First stage partial *F* on instrument	2916	2972	2717	2837	3219
	Panel B. Ordinary Least Squares				
Age at start of kindergarten	0.242***	0.282***	-0.079*	-0.151***	-0.064***
	(0.038)	(0.036)	(0.046)	(0.019)	(0.015)
Root MSE	0.885	0.844	0.885	0.404	0.465
R-squared	0.230	0.302	0.219	0.146	0.147
Observations	5,719	5,719	4,436	4,508	6,248
Additional controls[b]	N	Y	Y	Y	Y
School fixed effects	Y	Y	Y	Y	Y

Notes: Each column and panel of the table presents estimates from a different regression. Test scores are standardized to have a mean of zero and a standard deviation of one using all available data. Standard errors (in parentheses) are clustered on kindergarten classroom.

[a] The instrument for a child's age at the start of kindergarten is his expected age given his birthday and the September 30th kindergarten entry cutoff birthdate in Tennessee.

[b] Dummies for whether child is female, black, or received free/reduced-price lunch in K; fractions of K classmates with these characteristics; whether the kindergarten teacher is black, has an MA, or has 0 to 1 years of experience; and dummies for whether kindergarten class is small or regular sized with teacher's aide.

***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4. Estimates of the effect of school entry age and relative age

	Dependent variable:				
			Eight years later:		
	Average test score, end of K		Average test score	Ever retained	Took ACT or SAT
	(1)	(2)	(3)	(4)	(5)
	Panel A. Two-stage least squares - not accounting for lower ability of older peers [a]				
Age at start of K	1.020***	0.910***	0.650***	-0.156*	0.155*
	(0.282)	(0.265)	(0.181)	(0.080)	(0.092)
Relative age in K	-0.310	-0.233	-0.429**	-0.033	-0.151*
classroom	(0.265)	(0.250)	(0.167)	(0.075)	(0.086)
Root MSE	0.900	0.855	0.890	0.405	0.465
	Panel B. Two-stage least squares - accounting for lower ability of older peers [b]				
Age at start of K	1.564***	1.407***	0.651***	-0.327***	0.175
	(0.429)	(0.394)	(0.251)	(0.118)	(0.107)
Relative age in K	-0.847**	-0.724*	-0.430*	0.136	-0.172
classroom	(0.407)	(0.374)	(0.248)	(0.115)	(0.106)
Root MSE	0.901	0.856	0.890	0.405	0.465
	Panel C. Ordinary Least Squares				
Age at start of K	0.423	0.396	0.271	-0.107	0.069
	(0.278)	(0.263)	(0.173)	(0.079)	(0.088)
Relative age in K	-0.178	-0.112	-0.345**	-0.043	-0.132
classroom	(0.267)	(0.251)	(0.167)	(0.075)	(0.085)
Root MSE	0.885	0.844	0.885	0.404	0.464
R-squared	0.230	0.302	0.219	0.146	0.147
Observations	5,719	5,719	4,436	4,508	6,248
Additional controls[c]	N	Y	Y	Y	Y
School fixed effects	Y	Y	Y	Y	Y

Notes: Each column and panel of the table presents estimates from a different regression. Test scores are standardized to have a mean of zero and a standard deviation of one using all available data. Standard errors (in parentheses) are clustered on kindergarten classroom.

[a] The instrument for a child's age at the start of kindergarten is his expected age given his birthday and the September 30th kindergarten entry cutoff birthdate in Tennessee. The instrument for a child's relative age is his expected age – average age of his kindergarten classmates.

[b] The instrument for a child's age at the start of kindergarten is his expected age given his birthday and the September 30th kindergarten entry cutoff birthdate in Tennessee. The instrument for a child's relative age is his expected age – average *expected* age of his kindergarten classmates.

[c] Dummies for whether child is female, black, or received free/reduced-price lunch in K; fractions of K classmates with these characteristics; whether the kindergarten teacher is black, has an MA, or has 0 to 1 years of experience; and dummies for whether kindergarten class is small or regular sized with teacher's aide.

***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5. TSLS estimates of the impacts of relative age, by gender

	Dependent variable:			
		Eight years later:		
	Average Score End of K	**Average score**	**Retained in any grade**	**Took ACT or SAT**
	(1)	**(2)**	**(3)**	**(4)**
	Panel A. Boys			
Mean of dependent variable	-0.089	0.021	0.327	0.401
TSLS Coefficients:				
Age at start of kindergarten	0.809**	0.971***	-0.397***	0.209*
	(0.329)	(0.289)	(0.124)	(0.111)
Relative age in K classroom	-0.171	-0.622**	0.118	-0.147
	(0.303)	(0.254)	(0.117)	(0.101)
Root MSE	0.842	0.978	0.445	0.463
Observations	2,936	2,191	2,234	3,213
	Panel B. Girls			
Mean of dependent variable	0.098	0.221	0.175	0.549
TSLS Coefficients:				
Age at start of kindergarten	0.992***	0.428**	0.049	0.100
	(0.303)	(0.209)	(0.103)	(0.122)
Relative age in K classroom	-0.288	-0.310	-0.162*	-0.151
	(0.288)	(0.195)	(0.095)	(0.119)
Root MSE	0.867	0.798	0.361	0.468
Observations	2,783	2,245	2,274	3,035
P for age: male = female	0.291	0.011	0.002	0.147
P for rel. age: male = female	0.920	0.071	0.120	0.477

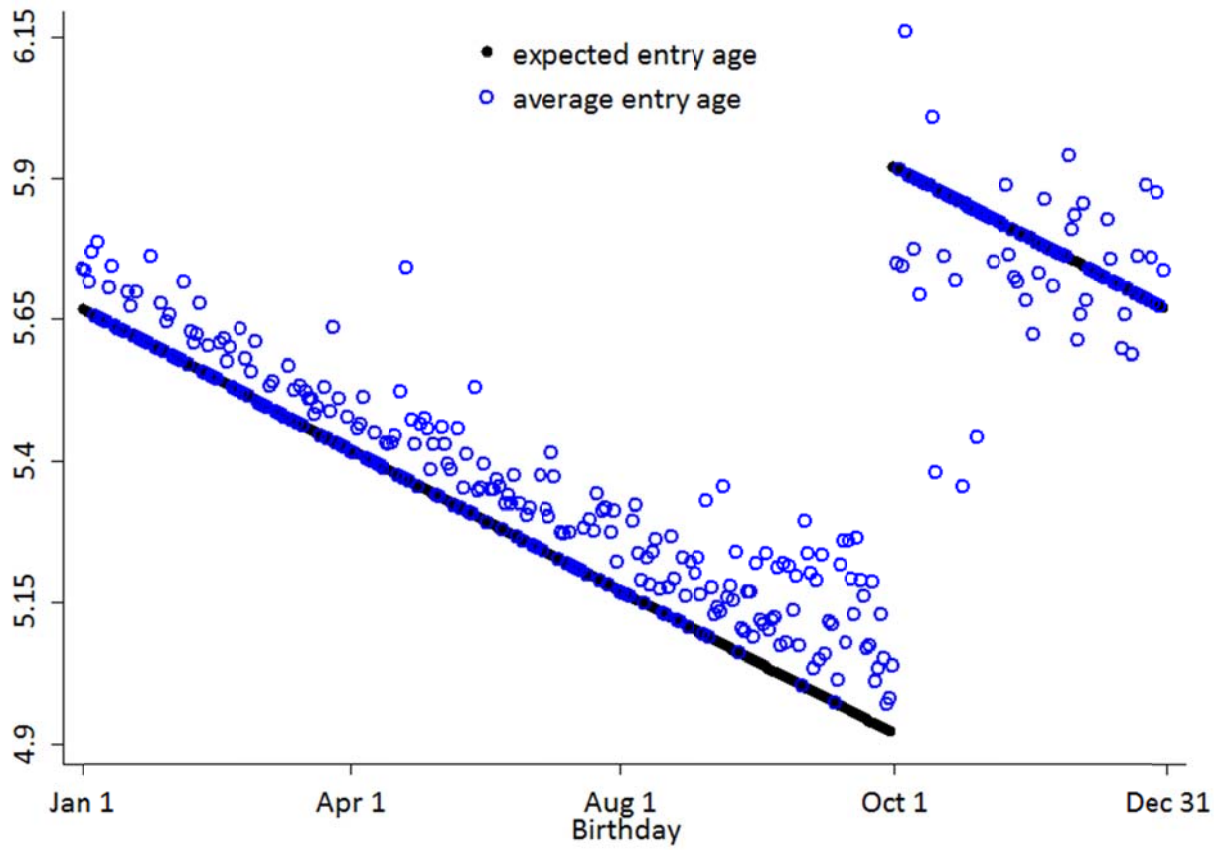
Notes: Each column and panel of the table presents estimates from a different regression. All models are estimated using two-stage least squares and are based on the specification presented in the final four columns of Table 4, Panel A. Thus, the models include the full set of controls; age at the start of kindergarten is instrumented with expected age at the start of kindergarten; and relative age is instrumented with own expected age – average age of kindergarten classmates. Standard errors (in parentheses) are clustered on kindergarten classroom.
***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6. TSLS estimates of the impacts of relative age, by free/reduced-price lunch status

	Dependent variable:			
		Eight years later:		
	Average Score End of K (1)	Average score (2)	Retained in any grade (3)	Took ACT or SAT (4)
	Panel A. Not free/reduced-price lunch in K			
Mean of dependent variable	0.268	0.467	0.155	0.623
TSLS Coefficients:				
Age at start of kindergarten	0.892***	0.397	-0.060	0.088
	(0.324)	(0.243)	(0.100)	(0.130)
Relative age in K classroom	-0.124	-0.166	-0.127	-0.058
	(0.309)	(0.222)	(0.090)	(0.118)
Root MSE	0.891	0.850	0.347	0.473
Observations	2,951	2,398	2,420	3,217
	Panel B. Free/reduced-price lunch in K			
Mean of dependent variable	-0.281	-0.284	0.360	0.313
TSLS Coefficients:				
Age at start of kindergarten	0.911***	0.856***	-0.270**	0.240*
	(0.339)	(0.295)	(0.134)	(0.134)
Relative age in K classroom	-0.319	-0.665**	0.082	-0.242*
	(0.315)	(0.270)	(0.132)	(0.127)
Root MSE	0.804	0.925	0.460	0.454
Observations	2,768	2,038	2,088	3,031
P for age: low SES = high SES	0.161	0.042	0.089	0.183
P for rel. age: low SES = high SES	0.401	0.047	0.290	0.120

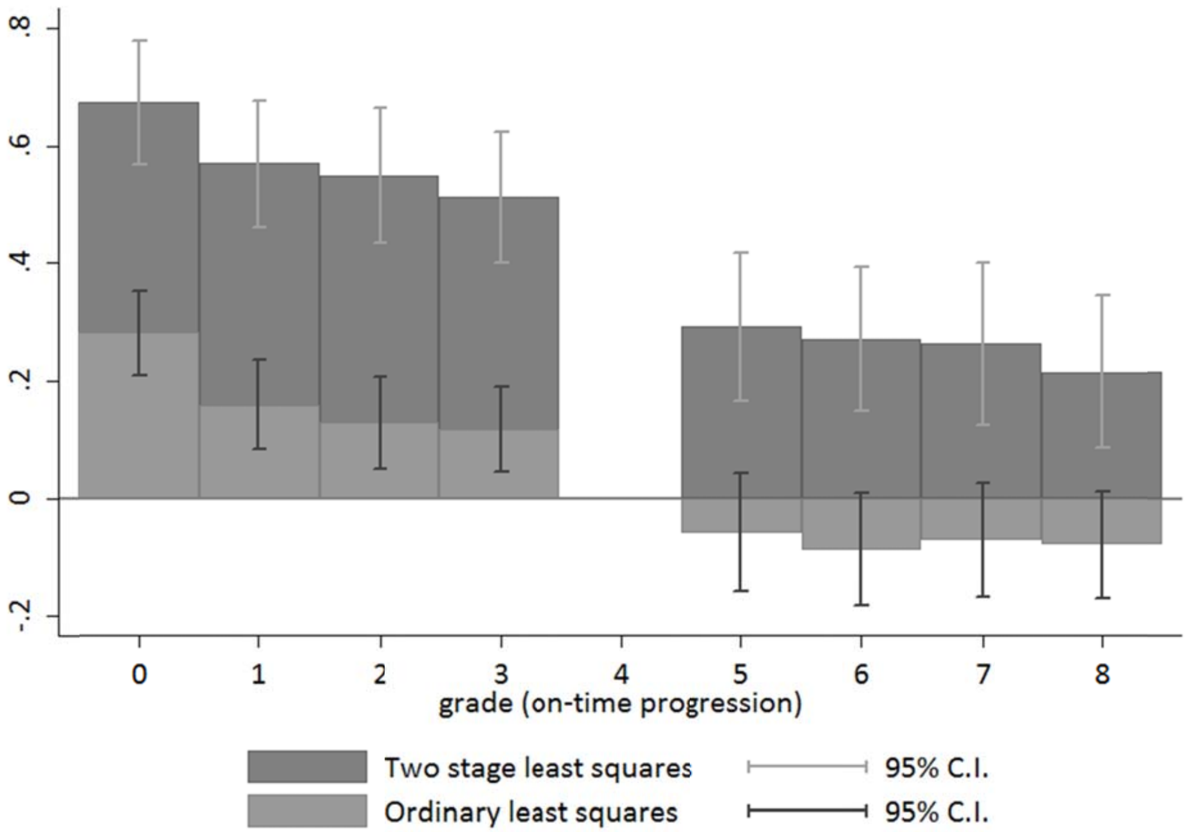
Notes: Each column and panel of the table presents estimates from a different regression. All models are estimated using two-stage least squares and are based on the specification presented in the final four columns of Table 4, Panel A. Thus, the models include the full set of controls; age at the start of kindergarten is instrumented with expected age at the start of kindergarten; and relative age is instrumented with own expected age – average age of kindergarten classmates. Standard errors (in parentheses) are clustered on kindergarten classroom. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Figure 1. Expected and Actual Kindergarten Entry Age by Birthday: Project STAR Kindergarten Cohort



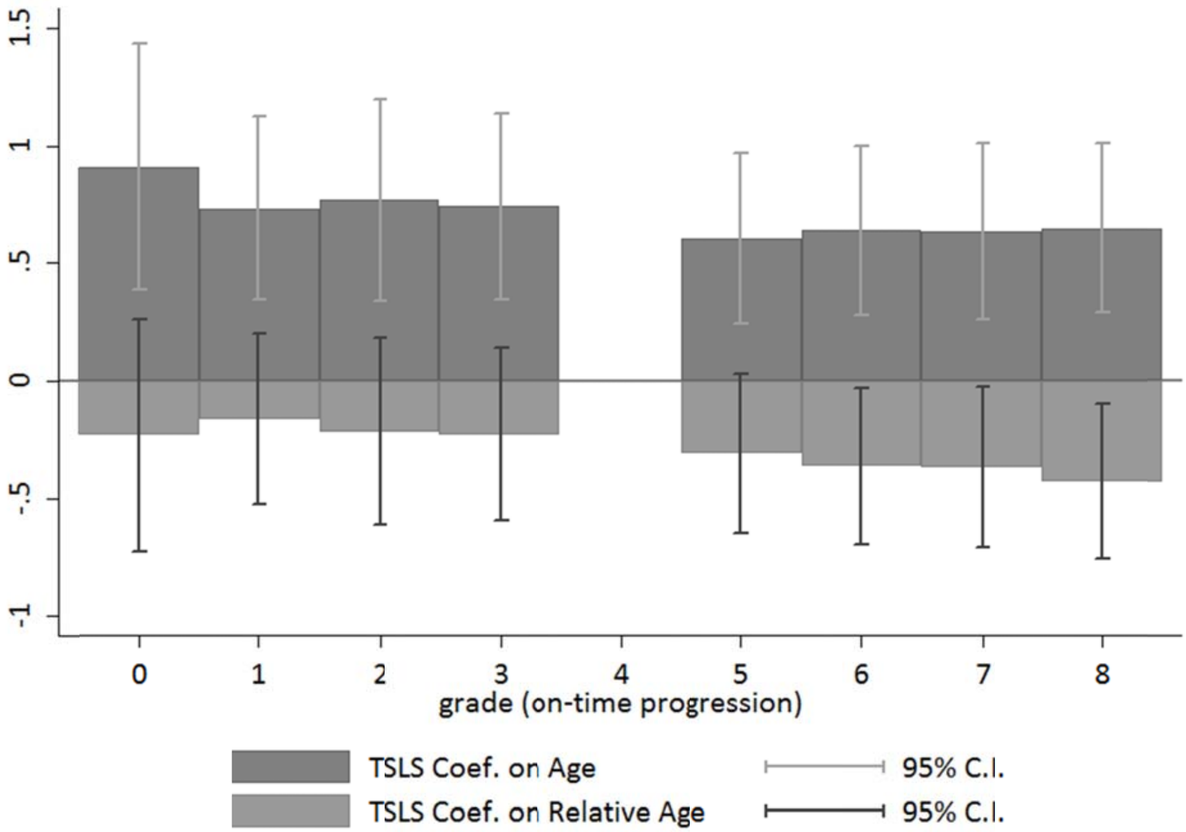
Notes: The darkened circles represent the age at which Project STAR participants born on each day of the calendar year would have been expected to enter kindergarten (assuming the school year starts September 1), given Tennessee's regulation that entering kindergartners must be age five by September 30. The hollow circles represent the average actual age of Project STAR participants on September 1, 1985, separately by birthday.

Figure 2. Reduced-Form Age Effects on Test Scores by Grade



Notes: The instrument for age at kindergarten entry is expected age at kindergarten entry based on exact birthday and the September 30th kindergarten entry cutoff birthdate in Tennessee. Grade 0 represents kindergarten. Test score data are available for grades 1 through 3 only for individuals who remain in a Project STAR school and progressed through school on time. Because observation of these individuals is predicted by the instrument, we impute test scores for missing observations with end-of-kindergarten z-scores. Test score data for grade 4 is missing for a large, non-random sample of schools, so results are not reported for this grade. Test score data used for grades 5 through 8 actually represent test scores from spring 1991 through spring 1994, respectively, regardless of grade of enrollment.

Figure 3. TSLS Estimates of Absolute and Relative Age Effects, by Grade



Notes: The instrument for age at kindergarten entry (absolute age) is expected age at kindergarten entry based on exact birthday and the September 30th kindergarten entry cutoff birthdate in Tennessee. The instrument for relative age is expected age at kindergarten entry minus the average age of a child's kindergarten classmates at kindergarten entry. Grade 0 represents kindergarten. Test score data are available for grades 1 through 3 only for individuals who remain in a Project STAR school and progressed through school on time. Because observation of these individuals is predicted by the instrument, we impute test scores for missing observations with end-of-kindergarten z-scores. Test score data for grade 4 is missing for a large, non-random sample of schools, so results are not reported for this grade. Test score data used for grades 5 through 8 actually represent test scores from spring 1991 through spring 1994, respectively, regardless of grade of enrollment.

Table A1. First-stage estimates for Table 3 TSLS models

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Age at start of K				
Expected age at start of K	0.812	0.815	0.831	0.829	0.815
	(0.015)	(0.015)	(0.016)	(0.016)	(0.014)
Partial *F*-stat on instrument	2916	2972	2717	2837	3219
Observations	5,719	5,719	4,436	4,508	6,248
Additional controls[a]	N	Y	Y	Y	Y
School fixed effects	Y	Y	Y	Y	Y

Notes: Standard errors (in parentheses) are clustered on kindergarten classroom. Each column corresponds to the same numbered column in Table 3.

[a] Dummies for whether child is female, black, or received free/reduced-price lunch in K; fractions of K classmates with these characteristics; whether the kindergarten teacher is black, has an MA, or has 0 to 1 years of experience; and dummies for whether kindergarten class is small or regular sized with teacher's aide.

Table A2. First-stage estimates for Table 4 TSLS models

	(1)	(2)	(3)	(4)	(5)
	Panel A. First Stage Estimates for Table 4, Panel A				
	Dependent variable: Age at start of K				
Expected age at start of K	0.642	0.634	0.675	0.672	0.625
	(0.064)	(0.066)	(0.057)	(0.060)	(0.072)
Expected age at start of K	0.169	0.179	0.155	0.155	0.189
- average starting age of K classmates	(0.059)	(0.061)	(0.051)	(0.053)	(0.067)
Partial *F*-stat on instruments	1560	1583	1460	1549	1731
	Dependent variable: Relative age at start of K				
Expected age at start of K	-0.358	-0.366	-0.325	-0.328	-0.375
	(0.064)	(0.066)	(0.057)	(0.060)	(0.072)
Expected age at start of K	1.169	1.179	1.155	1.155	1.189
- average starting age of K classmates	(0.059)	(0.061)	(0.051)	(0.053)	(0.067)
Partial *F*-stat on instruments	1991	1993	1972	2075	2126
Observations	5,719	5,719	4,436	4,508	6,248
	Panel B. First Stage Estimates for Table 4, Panel B				
	Dependent variable: Age at start of K				
Expected age at start of K	0.880	0.879	0.936	0.938	0.871
	(0.053)	(0.052)	(0.048)	(0.048)	(0.052)
Expected age at start of K	-0.067	-0.062	-0.103	-0.108	-0.055
- average expected starting age of K classmates	(0.052)	(0.051)	(0.048)	(0.048)	(0.051)
Partial *F*-stat on instruments	1506	1523	1449	1513	1633
	Dependent variable: Relative age at start of K				
Expected age at start of K	0.025	0.023	0.072	0.077	0.008
	(0.021)	(0.021)	(0.042)	(0.038)	(0.007)
Expected age at start of K	0.788	0.792	0.759	0.752	0.806
- average expected starting age of K classmates	(0.026)	(0.026)	(0.042)	(0.040)	(0.017)
Partial *F*-stat on instruments	1377	1419	1268	1321	1541
Observations	5,719	5,719	4,436	4,508	6,248
Additional controls[a]	N	Y	Y	Y	Y
School fixed effects	Y	Y	Y	Y	Y

Notes: Each column and panel of the table presents estimates from a different regression. Standard errors (in parentheses) are clustered on kindergarten classroom. Each column corresponds to the same numbered column in Table 4.

[a] Dummies for whether child is female, black, or received free/reduced-price lunch in K; fractions of K classmates with these characteristics; whether the kindergarten teacher is black, has an MA, or has 0 to 1 years of experience; and dummies for whether kindergarten class is small or regular sized with teacher's aide.