The effect of Catholic schooling on math and reading development in kindergarten through fifth grade

Sean F. Reardon
Stanford University

Jacob E. Cheadle
University of Nebraska-Lincoln

Joseph P. Robinson
Stanford University

March, 2008

Direct correspondence to Sean F. Reardon, 520 Galvez Mall, #526, Stanford University, Stanford, CA 94305, 650.736.8517, (sean.reardon@stanford.edu). This draft was prepared for the Annual Meeting of the Society for Research on Educational Effectiveness, March 2-4, 2008. Data for this paper were collected by the National Center for Education Statistics (NCES) and provided to Reardon under an NCES Restricted Data License (license #04012902). Because some of the data we use are only available to researchers with an NCES restricted data license, the data are not available from the authors, but may be obtained from NCES. The research and opinions here are ours, and do not necessarily reflect those of the NCES. Any errors are ours as well.

The effect of Catholic schooling on math and reading development in kindergarten through fifth grade

---

Abstract

Prior research estimating the effect of Catholic schooling has focused on high school, where evidence suggests a positive effect of Catholic versus public schooling. In this paper, we estimate the effect of attending a Catholic elementary school rather than a public school on the math and reading skills of children in kindergarten through fifth grade. We use nationally representative data and a set of matching estimators to estimate the average effect of Catholic schooling and the extent to which the effect varies across educational markets. When we use public school students nationwide to provide a counterfactual estimate of how Catholic school students would have performed in public schools, we find that strong evidence indicating that Catholic elementary schools are less successful at teaching math skills than public schools (Catholic school students are 3-4 months behind public school students by third and fifth grade), but no more or less successful at teaching reading skills. When we compare Catholic students to matched public school students attending schools in the same county, however, we obtain estimated math effects that are generally somewhere between the (negative) national estimates and zero (but statistically indistinguishable from either). We again find no evidence of a positive or negative Catholic schooling effect on reading skills.

The effect of Catholic schooling on math and reading development

in kindergarten through fifth grade

In this paper, we estimate the effect of attending a Catholic school rather than a public school on the math and reading skills of children in kindergarten through fifth grade. While there has been considerable prior research on the effects of attending a Catholic high school, there is very little existing research on the effects of attending a Catholic elementary school.

The paper proceeds as follows. In section 1, we discuss the importance of understanding the effects of Catholic schooling. In section 2, we review prior research on Catholic schooling. In section 3, we describe the data and methods used for our analyses. Section 4 describes our results. In section 5 we discuss the results at length, attending in particular to the reasons why the estimates from different models differ. Section 6 concludes.

1. Introduction

The question of whether Catholic schools provide better education than public schools is of interest for several reasons. First, the question has a rich history in the sociology of education literature. In 1982, Coleman and colleagues reported that Catholic schooling provided substantial positive effects on high school students' math and reading skills (Coleman, Hoffer, and Kilgore 1982a, 1982b, 1982c). This research prompted considerable debate and subsequent research on the effects of Catholic versus public schooling, and the reasons for such effects (see, for example, Altonji, Elder, and Taber 2002; Bryk, Lee, and Holland 1993; Grogger and Neal 2000; Hoffer, Greeley, and Coleman 1985; Morgan 2001; Neal 1997).

Second, understanding the effects of Catholic schools on student achievement is important on the basis of numbers alone. Catholic schools enroll a larger number of students than any other type of private school. Of the 5.1 million K-12 students (10% of all U.S. K-12 students) enrolled in private school, almost half (2.3 million in 2003-04) attend Catholic schools (Broughman and Swaim 2006; Reardon and Yun 2002). If Catholic schools produce better average student outcomes than public schools for this large number of students (as suggested in some of the Catholic high school effects literature), then it would be useful to understand the mechanisms through which Catholic schools produce such effects. And if Catholic schools produce worse average outcomes, then the large number of students affected is a cause for concern.

Third, policy proposals to provide private school vouchers to families hinge on claims that students would learn more if they attended private schools than if they attended their local public schools. Because Catholic schools make up by far the largest share of the private school sector in the U.S., and because Catholic school tuition is much lower, on average, than tuition in other private schools,[1] voucher recipients (and low-income voucher recipients in particular) are more likely to attend Catholic schools than other types of private schools. Thus, unless voucher programs were to create large changes in the structure of the private schooling sector (which they may do, but not in the short term), most students who take advantage of any proposed private school voucher program will likely attend Catholic schools. Estimating the causal effects of Catholic schooling on achievement thus provides information relevant to evaluating the likely effects of voucher programs.

Finally, information about the effects of Catholic schooling may shed light on the potential of competition among schools to lead to improvements in school quality. Catholic school enrollment, after reaching a peak in 1965 (when 5.6 million students were enrolled in 13,500 Catholic schools), has declined steadily over the last 40 years (to 2.3 million students in 7,500

[1] Average elementary school annual tuition in 2003-04 was $3,533 in Catholic schools; $5,398 in other religious private schools; and $12,169 in non-sectarian private schools (Digest of Educational Statistics, 2006, Table 56; at Hhttp://nces.ed.gov/programs/digest/d06/tables/dt06_056.aspH).

schools in 2006-7).[2] This decline has occurred largely as a result of rising tuitions caused by increasing financial pressure on the church and the movement of many Catholics to the suburbs, where the Catholic Church has traditionally operated few schools (Baker 1999; Baker and Riordan 1998; Bryk, Lee, and Holland 1993). A market model of private schooling would predict competition among Catholic schools to attract and retain remaining Catholic students and to attract non-Catholics away from public schools. This competition should (according to a market competition theory) lead to improvements in Catholic schooling and to the closing of the weakest Catholic schools. As a result, we might expect Catholic schools to be better than public schools because the same declining enrollment pressures have not operated on public schools. Of course, a variety of other pressures operate on both Catholic and public schooling, so any differences we find in the effects of Catholic and public schooling should not be read as unambiguous evidence regarding the effects of competition.

2. Prior research on the effect of Catholic schooling

We have been able to find only one study using a large-scale data set to estimate the effects of Catholic elementary schooling on achievement outcomes in the elementary grades. Nonetheless, there are two bodies of research which are relevant to consider. First, there is a considerable body of research on the effects of attending a Catholic *high school*. Second, there is some research on the effects of attending a *private* (not necessarily Catholic) school in the elementary grades. Much of this latter research stems from research estimating the effect of school vouchers. We review each of these research strands in turn, attending in particular to the methods used to identify the causal effect in each case.

---

[2] Note that this decline continues through the present. In 2006-7, 212 Catholic schools closed or consolidated, while 36 new Catholic schools opened. Since 2001-02, the total number of Catholic schools has declined from 8,146 to 7,498 and enrollment declined from 2.6 million to 2.3 million (Digest of Educational Statistics 2003, Table 62). See Hhttp://www.ncea.org/news/AnnualDataReport.aspH (accessed February 19, 2008) and Digest of Educational Statistics 2003, Table 62.

Prior research on the effect of attending Catholic high school

Most studies on the effects of Catholic high schools conclude that attending a Catholic high school is has a positive effect on achievement for at least some group of students (Coleman, Hoffer, and Kilgore 1982a, 1982b, 1982c; Morgan 2001; Murnane, Newstead, and Olsen 1985; Neal 1997). None of these studies are based on experimental designs, of course (it would be hard to randomly assign some students to attend Catholic schools and others to attend public schools); rather, each relies on observing the test scores and graduation rates of students in both Catholic and public schools and inferring causation using a statistical model. As a result, the validity of their conclusions rests on the extent to which their assumptions of their models are plausible. Researchers have relied on three types of models to address the issue of selection bias: 1) covariate adjustment (regression) models; 2) instrumental variables models; and 3) matching estimators.

Covariate adjustment (regression) models. The seminal research on the effect of Catholic high schools was done by Coleman, Hoffer, and Kilgore (1982b; 1982c). Using the High School and Beyond (HSB) dataset and regression models that control for race and socioeconomic background, they found a substantial positive effect of Catholic schooling on vocabulary and mathematics scores. However, some have criticized the Coleman, Hoffer, and Kilgore results on the grounds that the regression models do not adequately control for initial achievement and unobserved selection mechanisms, resulting in biased estimates on Catholic schooling (Goldberger and Cain 1982). Principal among the criticisms of the Coleman et al. paper is that proper steps were not taken to ensure the comparability of the Catholic and public school students at entry into high school because no assessments were given prior to the sophomore assessments. Moreover, Murnane, Newstead, & Olsen (1985) illustrate how modeling the school-choice selection process—and, in particular, doing so separately for whites, blacks, and Hispanics—can lead to a very different outcome: Blacks and Hispanics experience significant and substantial benefits in both reading and mathematics from attending Catholic schools, whereas whites exhibit no significant evidence of an effect. Similar results are found in other studies using covariate adjustment (Grogger and Neal 2000).

Instrumental variable models. Rather than attempting to eliminate selection bias through modeling, another solution for removing selection bias is to use an instrumental variable or variables—variables that predict Catholic school enrollment but are otherwise uncorrelated with student outcomes. Neal (1997), for example, considers several potential instrumental variables: 1) whether the student is Catholic; 2) the availability of Catholic secondary schools in the county; and 3) the proportion of the county that is Catholic. Using these three instruments and a bivariate probit model estimated on data from the National Longitudinal Survey of Youth (NLSY-79), Neal finds positive effects of Catholic schooling, with the effects being small for suburban students and larger for urban minority students. For urban students, Neal's results show a positive Catholic school effect on academic achievement, graduation rates, college attendance, and future wages. Neal suggests that urban minority students benefit more than suburban students and white urban students because of the poor quality of the public schools serving urban minority children. Thus, Catholic schooling is a greater contrast to public schooling for urban minorities, which contributes to the large estimated effect for this group. Evans and Schwab (1995) also use a variable indicating whether a student is Catholic as an instrument, and find a strong positive effect of Catholic schooling on graduation and college enrollment.

Of course, instrumental variables models rely on a set of assumptions as well. Altonji, Elder, and Taber (2002) evaluate the extent to which the instruments used by Neal and Evans and Schwab satisfy the assumptions needed to produce unbiased estimates. Using data from national several national surveys and several tests of the instrumental variables assumptions, Altonji, Elder, and Taber conclude that none of the three potential instruments are valid, calling into question the results that rely on them to estimate Catholic school effects. Instead of an instrumental variables approach, Altonji, Elder, and Taber (2005) propose a method for estimating lower and upper bounds on the Catholic school effects by using the observable selection processes as a basis for bounding the possible bias due to unobservable selection mechanisms. Using the NELS dataset, they obtain lower and upper bounds on the Catholic school effect, which suggest Catholic high schools have a positive effect on graduation rates but not necessarily on achievement.

Matching estimators. The use of matching estimators is another approach to addressing selection bias. Matching estimators, including propensity score matching, rely on the assumption that Catholic and public school students who are similar on a set of observed characteristics can stand as valid counterfactuals for one another. Matching estimators depend on the same assumption that there are no omitted confounder variables as does covariate adjustment, but have several advantages. First, matching estimators do not rely on assumptions about the functional form of the model. Second, they estimate average effects only for the subpopulation of students who have matches in the opposite treatment condition. And third, they facilitate the investigation of treatment effect heterogeneity.

Morgan (Morgan 2001) uses a propensity score matching estimator to estimate the effect of attending a Catholic school on math and reading achievement. Using NELS data, he finds a positive effect of Catholic schooling on high school math and reading scores.

Heterogeneity of treatment effects

There are a number of reasons to expect that Catholic schooling effects may vary among subsets of the population. There are three types of variation of particular interest. First, a number of studies have found Catholic schooling to have larger effects for minority (typically Black) students and urban students (Altonji, Elder, and Taber 2005; Evans and Schwab 1995; Neal 1997). Second, Neal argues that the effect of Catholic schooling may vary among locations—where public schools are lower in quality, Catholic schools may be better by comparison (Neal 1997). Third, the likelihood of attending a Catholic school may be associated with the expected effect of Catholic schooling. We would expect this to occur because of positive selection—those most likely to benefit from Catholic schooling are likely to be those most likely to attend Catholic schools. Contrary to this expectation, however, research has suggested students who are least likely to attend Catholic school, including minority students, have the greatest benefits (Altonji, Elder, and Taber 2005; Evans and Schwab 1995; Morgan 2001; Neal 1997). Morgan (2001) suggests that this might occur for a variety of reasons—Catholic schools may in fact be better for such students; such students may work hard in Catholic schools because the financial sacrifice of their parents to pay for private schooling is more salient to them than to higher-income students; Catholic schools may be better for such students only relative to the lower-quality public schools available to them; or because the enrollment of low income students in Catholic schools is more strongly conditioned on their parents' expectation of the effectiveness of Catholic schooling for their child than for higher-income students. Morgan is unable to distinguish among these competing hypotheses in his analysis, however.

Effects of attending a private elementary school

As described above, most of the studies of high school Catholic school effects report positive effects of Catholic schooling. One exception is a recent study that uses a multilevel regression model and NAEP data to estimate the effect of private schools (including Catholic schools) on 4th and 8th grade math achievement (Lubienski and Lubienski 2006). Notably, this is also the only study to specifically estimate Catholic elementary school effects as opposed to high school effects. Lubienski and Lubienski find that Catholic school students score roughly .50 standard deviations lower than observationally similar public school students in 4th grade and .20 standard deviations lower in 8th grade. Nonetheless, although their regression model controls for basic student demographic characteristics (e.g., race, gender, free-lunch status), it does not include an extensive set of control variables, relies on functional form assumptions, and assumes the Catholic school effect is constant across locations.

Studies of voucher effects may provide information on the effects of Catholic schools. Because Catholics are more likely to apply for vouchers and use those vouchers for Catholic school attendance (Campbell, West, and Peterson 2005), voucher studies—at least those with substantial Catholic school samples—may shed light on Catholic school effects.[3] For example, voucher experiments in New York City, Dayton, OH, and Washington, DC, provide estimates of the effects of vouchers (thus private schools, including Catholic schools) on students in low-income (all three cities) and moderate-income families (Dayton, OH, and Washington, DC). Each of the cities used a randomized controlled experimental design, implemented through a lottery system. The dollar amount of the vouchers varied across the programs, but the voucher covered the majority of the tuition for students attending religious schools, which tend to charge lower tuition than secular private schools. Of the voucher participants, 68% in NYC, 59% in Dayton, and 49% in DC, began the experiment by using the voucher to attend Catholic schools; however, the percentages attending Catholic schools dropped monotonically throughout the programs, with voucher participants returning to public schools (Howell and Peterson 2006). Howell & Peterson's analyses of these programs found no consistent evidence of a positive effect of moving from a public to private school. The only exceptions to this pattern of null effects were the finding (based on post-hoc race-specific analyses) that black students benefitted significantly more in the second year of the program

[3] Although research on the publicly-funded Milwaukee voucher program, for example, finds a small benefit of private schools on math achievement, the public funds from this voucher program could not be spent on religious-school tuition, and thus these results cannot be generalized to Catholic schools (Rouse 1998).

(but not first or third years) in Washington, DC, and that vouchers had a positive effect on achievement by fifth grade in the NYC program. Of course, although a large proportion of voucher users attended Catholic schools, these experimental results should only be used as suggestive of what we might find as a Catholic school effect.

One additional voucher study provides estimates of Catholic school effects, albeit not in the U.S. context. McEwan and Carnoy's (2000) study of Chilean schools found Catholic schools to be slightly more effective than public schools in Spanish and math achievement levels for fourth-grade students, after controlling for family socioeconomic status. However, they note that Chilean Catholic schools spend more than public schools, and thus Catholic schools are no more efficient at producing high achievement.

In sum, the research on Catholic high school effects generally has found positive effects of attending a Catholic school, particularly for minority students and those least likely to attend Catholic schools. None of this research is without potential bias, but the consistency of the results suggests there may be a positive effect of Catholic high school enrollment. Evidence of Catholic elementary schooling, however, is much sparser, and does not suggest any positive effect. In this paper, we estimate Catholic elementary schooling effects in order to determine whether the pattern of positive Catholic high school effects is found also in elementary school.

3. Data and Methods

The data for this study come from the Early Childhood Longitudinal Study–Kindergarten Class of 1998-1999 (ECLS-K), sponsored by the National Center for Education Statistics (NCES). The ECLS-K contains data on a nationally-representative sample of 21,260 students from the kindergarten class of 1998-99 (thus, representing a cohort born in roughly 1992-93). Students in the sample were assessed in reading and mathematics skills at six time points during the years 1998-2004 (fall 1998, spring 1999, fall 1999, spring 2000, spring 2002, and spring 2004). In addition to these cognitive developmental measures, the ECLS-K data include information gathered from surveys of parents, teachers, and school administrators regarding family, school, community, and student characteristics.

We use math and reading scores on the ECLS-K direct cognitive assessments as measures of children's math and reading skills. The ECLS-K direct cognitive assessments are individually-administered, oral, untimed, adaptive tests of math and reading skills. The content areas of the tests are based on the NAEP 4th grade content areas, adapted to be age appropriate at each assessment. The assessments were administered by trained ECLS-K assessors, and were scored using a 3-parameter Item Response Theory (IRT) model. Details of the assessments are provided in the ECLS-K psychometric reports (Pollack, Narajian, Rock, Atkins-Burnett, and Hausken 2005; Pollack, Rock, Weiss, Atkins-Burnett, Tourangeau et al. 2005; Rock and Pollack 2002). We use the *T*-scores (a version of the test score that is standardized within each assessment wave to a mean of 50 and a standard deviation of 10) reported by NCES for all analyses in this paper. These scores are suitable for repeated cross-sectional analyses (as we use here), though not for longitudinal growth models.

At each wave, students were only administered the ECLS-K math assessment if they were proficient in oral English or oral Spanish, and were only administered the ECLS-K reading assessment if they were proficient in oral English.[4] In the early waves of the ECLS-K data collection, many Hispanic and Asian students (29% of Hispanic students; 22% of Asian students) were not fluent enough in oral English to be assessed in reading (in English). The 22% of non-English proficient Asian students were also not assessed in math. Because the proportion of oral

[4] Students were administered the math and reading assessments orally in English if either the school reported the student was from a home where English was the primary language or if the student passed the English Oral Language Development Scale (OLDS) assessment given by ECLS-K assessors. If the student was not proficient in English, but passed the Spanish OLDS, then he or she was administered the math test orally in Spanish, but was not administered the reading test. Students not proficient in oral English or Spanish (most of whom were Asian-origin students) at a given wave were not given the math or reading tests. These students were readministered the English OLDS assessment at subsequent waves, and given the math and reading tests once they passed the English OLDS.

English-proficient Hispanic and Asian students grows over time (to 99% by the spring of third grade), and because students not proficient in English certainly have lower average reading skills in English than students proficient in oral English, trends in the mean reading and math scores of those students with reading scores at a given wave are confounded by changes in the population of students represented in the sample of students with test scores at that wave. The same bias may be present in the patterns of math scores, but it will likely be much smaller than for reading, because the number of students excluded from the math test for language reasons is much smaller. In order to avoid bias in our estimates because of changes in the sample of students with test scores at each wave, we estimate the effects of Catholic schooling on math and reading achievement at each wave only for the subpopulations of students who had valid math and reading test scores, respectively, at the start of kindergarten. Thus, our estimates of the average effect of Catholic schooling on math skills may not generalize to children who are not proficient in English or Spanish at the start of kindergarten. Likewise, our estimates of the average effect of Catholic schooling on reading skills may not generalize to children who are not proficient in English at the start of kindergarten.

Data restrictions

In addition to restricting our sample to those students with valid math and reading scores in the fall of kindergarten, we restrict our analyses to a subsample of students who meet several other criteria. First, we restrict the sample to first-time kindergarten students attending either a Catholic or public kindergarten in the fall of 1998 (second-time kindergarteners (n=2,641) and students attending non-Catholic private schools (n=2,318) are excluded). These restrictions yield a sample of 16,301 students (1,979 Catholic and 14,322 public school students). Third, as noted above, we restrict the sample to students present in the sample in each of waves 1, 2, 4, 5, and 6 and with valid reading and math scores at wave 1. These restrictions ensure that our estimates at each wave are based on the same sample of students, and so are not biased by sample attrition or the missing test scores of non-English or-Spanish proficient students in the early waves of the study. Moreover, the restriction to students present in each wave of the study is necessary so that we can use the ECLS-K panel weight (variable *c1_6fc0*).[5] This reduced the sample to 7,614 (1,090 Catholic and 6,524 public school students). Finally, we excluded 136 students missing data on race/ethnicity (8 students) or key family variables (128 students) needed for the propensity score models. The final analytic sample size is N=7,480 (1,078 students in 97 Catholic schools, and 6,402 students in 692 public schools). Although this is much smaller than the full ECLS-K study sample, most of the sample restrictions are by design rather than missing data. Of the sample of 7,614 students who were English- or Spanish- proficient first-time kindergarten students in the Fall of 1998 attending Catholic or public schools, only 2% are excluded from the sample due to missing data.

The ECLS-K sample was drawn by first selecting a random sample of counties or groups of counties in the U.S., sampling schools within these counties, and then sampling students within these schools. As a result, there are many Catholic schools in the sample in counties in which there is also at least one public school in the sample. This feature of the sample design of ECLS-K allows us, in some of our analyses, to compare the outcomes of Catholic and public school students attending schools within the same educational market—schools which might meaningfully be considered as viable educational alternatives for the same families. For these analyses, we further restrict the main analytic sample to include only students who attended school in one of 60 counties in the U.S. in which there was at least one Catholic and one public school in the ECLS-K study. This sample, which we refer to as the "60 market sample" includes 1,001 Catholic and 2,554 public school students (in 89 Catholic and 285 public schools).

[5] By design, not all students who transferred out of their original school were followed in the ECLS-K study. The panel sample weights are used to reweight the sample remaining the study to be representative of the fall kindergarten population.

Sample description

On average, students attending Catholic school in kindergarten differ substantially from those attending public school (see Tables 1 and A1). Catholic school students, on average, come from more advantaged families—they have higher incomes ($23,000 higher, on average), higher parental education levels, older mothers, were less likely to receive food stamps or be below the poverty line, and were more likely to attend center-based child care before kindergarten (among other differences—see Appendix Table A1 for detail). Not surprisingly, Catholic school students also have much higher math and reading scores (.49 standard deviations higher in math and .42 standard deviations higher in reading)[6] in the fall of kindergarten than public school students.

Averaging over the full set of covariates included in Table A1, Catholic school students differ from public school students, on average, by roughly three-tenths of a standard deviation. The patterns in the 60 market sample are very close to those in the full analytic sample, suggesting that the 60 market sample is roughly representative of the larger population of interest.

Analytic strategy

Given that Catholic and public school students differ in many ways likely to affect their cognitive skills, a simple comparison of their mean test scores at subsequent waves is likely to yield substantially biased estimates of the effect of Catholic schooling. In this paper, we rely primarily on a set of propensity score matching estimators to purge bias from our estimates of the effect of Catholic schooling.

To estimate the effect of Catholic versus public schooling, we use several different

---

[6] The ECLS-K *T*-scores are designed to have mean 50 and standard deviation 10 in the population. The values of .49 and .42 are the fall kindergarten Catholic-public differences in *T*-scores in reading and math, respectively, divided by the standard deviation 10. The 'standardized differences' reported in Tables 1 and 3 are computed by dividing the Catholic-public differences in *T*-scores by the pooled standard deviation, which is slightly smaller than the population standard deviation. As a result the standardized differences in Tables 1 and 3 are slightly larger than the population differences.

estimators, including linear regression, linear regression including market fixed effects, a standard propensity score matching estimator, a propensity score model that includes market fixed effects, and two within-market matching estimators. For each estimator, we estimate a series of cross-sectional regression models, estimating the Catholic-public student difference in test scores, conditional on the model, at each wave. Although it would also be desirable to fit growth models, taking advantage of the longitudinal nature of the ECLS-K design, we do not do so. Growth models rely on the assumption that test scores are measured in an interval-scaled metric, meaning that a difference of one point has the same substantive meaning at any point on the test score scale. Reardon (2007), however, shows that the scale scores provide by ECLS-K are not measured in a metric that is meaningfully interval-scaled (a one-point difference on the scale corresponds to one additional item correct on the test, but because there are many more 'difficult' items on the ECLS-K tests than 'easy' items, equal gains in skill at different initial skill levels will yield larger score gains for students within higher initial skills than those with lower initial skills). As a result, growth models using ECLS-K data may yield biased estimates of Catholic school effects.

WLS regression estimates

Initially, we estimate the effect of Catholic schooling using simple weighted least squares regression. We begin by estimating a simple model (Model 1) with no controls, weighting observations by their ECLS-K panel weight *c1_6fc0*. These estimates provide a description of the observed difference in average test scores between Catholic and public school students at each wave. In model 2, we add to Model 1 a long list of covariates (those included in Table A1, plus a set of interaction and higher order terms—see footnote 8) as control variables. Because these models may not fully capture differences in prior cognitive skill between Catholic and public school students, we add the wave 1 test score as a control variable in Model 3 (thus, we cannot estimate model 3 for wave 1 outcomes). While it is possible that this model may overcontrol for prior cognitive skill, because the wave 1 score was measured 1-2 months into the school year (and so may be affected by Catholic schooling), this would have the effect of biasing the Catholic school estimates toward zero.

WLS market fixed effects estimates

Although the WLS regression models control for differences in student characteristics, the estimates from these models may be biased by the non-random distribution of Catholic school attendance in relation to public school quality. If, for example, the likelihood of attending Catholic school and achievement levels, conditional on the type of school attended) are correlated, then models that do not account for these between-market differences will yield biased estimates. For example, if students are more likely to attend Catholic schools in educational markets with low-quality public schools, conditional on the observed covariates in the regression models, the regression estimates will be biased toward zero. To eliminate this potential bias, we include county fixed effects (as a proxy for educational market fixed effects) in Models 4 and 5. These models have the same form as Models 2 and 3, save for the inclusion of the fixed effects and the necessary restriction of the sample to those students in the 60 counties with at least one Catholic and one public school in the ECLS-K sample. The estimates from these models can be interpreted as the average adjusted difference in test scores between students attending Catholic and public schools in the same county.

National propensity score estimates

Models 1-5 use linear regression to estimate the effect of Catholic schooling. The linear regression model depends not only on the assumption that there are no omitted confounding variables that affect both Catholic school attendance and subsequent academic performance, but also on the assumption that the linear model is correctly specified. One alternative to regression is propensity score matching (Rosenbaum 1983), an approach used by Morgan in the estimation of the effect of Catholic schooling (Morgan 2001). While propensity score matching estimation relies on the same assumption that there are no omitted confounders as does the regression model, it relaxes the modeling assumption, and enables us to estimate the effect of Catholic schooling without relying on the linearity assumption or the extrapolation of the model to regions where the density of public or Catholic school students is low. Moreover, propensity score matching allows us to explicitly identify the region of common support—the region of data in which there are both public and Catholic school students who can be matched and used to estimate the effect of Catholic versus public schooling.

In practice, there are many methods of matching, including stratification on the propensity score, 1-1 (or *M*-1) propensity score matching, caliper matching on the propensity score, nearest neighbor matching, exact matching on some covariates, and matching with and without replacement (Abadie and Imbens 2002; Imbens 2004). The choice among these methods typically involves some tradeoff between precision and potential bias, though it is beyond the scope of this paper to investigate or employ each of these methods. Matching each Catholic school student to many public school students (as we would, for example, with stratification on the propensity score, M-1 matching, or caliper matching with a wide caliper) would yield a large matched public school sample, leading to greater precision in our treatment effect estimates, but it would do so at the risk of introducing (additional) potential bias into the estimates, since the more matches we use, the less similar, on average, they may be to their Catholic school counterparts. We opt here for a fairly narrow caliper and match with replacement in order to ensure that each Catholic school student is either matched to one or more very similar public school students or is dropped from the estimation.

Specifically, we implement the propensity score matching estimator as follows. First, using the full analytic sample and the full set of student covariates described in Table A1, we fit a logit model predicting enrollment in Catholic versus public school in the fall of kindergarten. From this model, we obtain the predicted probability $\hat{p}_i$ of each student's enrollment in Catholic school, given his or her vector of observed covariates; this is the estimated propensity score. Second, for each Catholic school student, we select as matches all public school students with estimated propensity scores within one percentage point of that of the Catholic school student (if there are more than 10 public school students within one percentage point of a given Catholic school student, we select the closest 10 as matches). We match with replacement, so that a given public school student may be used as a match for more than one Catholic school student. Third, each matched public school student is given a weight w_i, defined as

$$w_i = \sum_c v_c \left(\frac{m_{ic} v_i}{\sum_p m_{pc} v_p} \right) \qquad [1]$$

where c indexes Catholic school students and p indexes public school students, v_j is the ECLS-K longitudinal sample weight (ECLS-K variable *c1_6fc0*) for student j, and m_{jc}=1 if public school student j is used as a match for Catholic school student c (i.e., if $|\hat{p}_j - \hat{p}_c| \leq .01$) and m_{jc}=0 otherwise. Each Catholic school student is assigned a weight $w_i = v_i$. The sums of these weights are equal for the matched samples of Catholic and public school students. Importantly, the construction of weights in this fashion reweights the matched public school sample to be similar to the distribution of Catholic school students, meaning that our estimates should be interpreted as estimates of the average effect of Catholic schooling on (the type of) students enrolled in Catholic schools. We return to this point later. Fourth, we assess the balance of the matched samples—the extent to which the matching has eliminated or reduced the correlation between Catholic/public school status and the observed covariates—by using the matched samples and fitting, for each covariate X included in the propensity score model, a regression model (weighted by w_i) of the form

$$X_i = \beta_0 + \beta_1(C_i) + \varepsilon_i, \quad [2]$$

where C_i is an indicator for whether a student attends Catholic or public school. The estimated coefficient $\hat{\beta}_1$ indicates the average difference in the covariate X between the weighted matched public and Catholic student samples; a coefficient of zero indicates there is no average difference between the two samples. We assess the overall quality of the matching by computing the average of the absolute values of the $\hat{\beta}_1$'s across all the covariates. To optimize the matching, we fit variations of the propensity score model using higher-order terms and polynomial terms and use the model that minimizes this average absolute standardized difference between the Catholic and public school samples.[7]

Finally, we estimate the effect of Catholic versus public school on math and reading test scores at each wave t by fitting regression models of the form

$$Y_{it} = \gamma_t + \delta_t(C_i) + \varepsilon_{it}. \quad [3]$$

The estimated coefficient $\hat{\delta}_t$ will be an unbiased estimate of the effect of Catholic schooling on test score Y at wave t under the assumption that matching has removed all potential bias—that, on average, for students with the same estimated propensity score, the selection of a Catholic or public school is uncorrelated a with student's potential test scores. As above, we fit two versions of these models. The first (Model 6) includes no covariates; the second (Model 7) includes the propensity

[7] Here we follow the suggestions made by Imai, King, and Stuart and by Morgan and Todd for assessing balance rather than using the more conventional approach of assessing balance on the basis of whether we fail to reject the null hypotheses that each $\beta_1 = 0$. They remind readers of the well-known (but sometimes ignored) facts that tests of null hypothesis are dependent on sample size, and that interpretation of a failure to reject the null hypothesis does not necessarily affirm the null is true. In small samples, in particular, we may fail to reject the null hypothesis even when there is a large difference in the covariates between the matched samples. (Imai, King, and Stuart forthcoming; Morgan and Todd 2007). Based on this approach, we use a final model that includes the full set of covariates listed in table A1 as well quadratic and cubic terms for household income, a quadratic term for the number of months the child has lived in the current home, interaction terms of race with household income, and interaction terms of mother's and father's occupation with a variable indicating if the student was black.

score as a covariate in the model to adjust for any residual bias due to the observed covariates that remains because of imperfect matching (Ho, Imai, King, and Stuart 2007; Imai, King, and Stuart forthcoming). In addition, Model 7 includes the wave 1 test score as a covariate in the wave 2-6 models both because the matching and propensity score adjustment may not capture all differences in pre-kindergarten cognitive skill, and because including the wave 1 score will likely increase the precision of our estimates.

Propensity score estimates with market fixed effects

Although the propensity score matching estimator removes bias associated with the observed covariates included in the matching model, the estimates from these models may still be biased by the non-random distribution of Catholic school attendance in relation to public school quality, as we noted above. To assess this, we restrict the sample to the 60 market sample, and construct a matched sample of public school students to the Catholic school students in these markets, using the propensity scores model estimated above. As above, we use 10-1 matching with a .01 caliper and replacement. We then fit another pair of models (Models 8 and 9) that include county fixed effects, using the matched sample within the 60 markets. The estimates from these models are purged of bias associated with the covariates included in the matching model, as well as any residual bias associated with the correlation of Catholic school enrollment with characteristics of counties.

Within-market propensity score matching

We include one further refinement to the propensity score model. The matching algorithm used in Model 6-9 matches students on observed student characteristics, but allows matched students to be drawn from anywhere in the country. This means, for example, that a low-income

Catholic school student in Boston may be matched to a low-income public school student in Birmingham. Moreover, because Catholic schooling is more common in certain parts of the country and in urban rather than suburban areas, it is likely that the strategy of matching students without regard for their location may yield samples that are matched on observable student characteristics but not matched on factors that vary across locations. Recent research on the conditions under which matching estimators are most likely to reduce bias, however, suggests that matching performs best when matches are local—that is, when matched samples come from the same local market.[8] In order to ensure that our matched samples are matched on individual as well as contextual factors, we use a within-market matching estimator, similar to that described by Reardon (2006).

The within-market matching estimator follows a similar algorithm to the standard matching procedure described above, with a few key changes. We use the sample of students in the 60 market sample, and fit a multilevel logit model (Raudenbush and Bryk 2002; Raudenbush, Bryk, Cheong, and Congdon 2005) to estimate students' propensity scores, including within-market centered values of each of the covariates, as well as the same set of interaction and higher-order terms as we use in the national propensity score matching model (see footnote 8). We include a market-level random effect on the intercept. Initially, we constrain each of the coefficients on the covariates (which indicate the average within-market association between the covariate and enrollment in Catholic school) to be constant across markets. This may be an inappropriate constraint for some covariates, however. For example, the association between family income and the probability of Catholic school enrollment may vary across sites because of tuition differences across sites. We test the hypotheses that the coefficients on key covariates are constant across sites, and relax these

[8] Much of the work on the performance of matching estimators comes from the job training and employment literature. Given that employment outcomes depend on local labor market conditions, the need to match samples within local markets has some intuitive justification. The same logic likely applies in education, where schooling outcomes may be affected by local conditions (teacher quality may be affected by local labor market conditions; school resources may be affected by local socioeconomic conditions and school finance laws, and so on).

assumptions in cases where the hypothesis is rejected and where allowing the coefficient to vary across sites improves the quality of our matches.[9] In other words, we fit a model to estimate the within-market propensity score, where the association between some covariates and Catholic school enrollment is allowed to vary across markets, while others are fixed (allowing the model to 'borrow strength" from observations in multiple markets). From this model, we obtain an estimated propensity score for each student.

For each Catholic school student, we select as matches all public school students who are enrolled in school within the same market and who have estimated propensity scores within five percentage points of the target as the target Catholic school student. We use a caliper of .05 here rather than .01 as above in order to obtain a larger sample of matched students. Even so, in 16 of the 60 markets, there are 5 or fewer public school students with propensity scores within .05 of any Catholic school student (despite the fact that there are an average of 42 public school students per market in the 60 markets). We drop these 16 markets from the sample because they can provide very little information on within-market Catholic school effects given their small regions of common support. Within each of the remaining 44 markets, we drop from the sample public and Catholic students who have no matches (no students in the opposite type of school and with propensity score within .05 of theirs). We construct weights for the matched students using the same procedure as described above. To assess the balance of the matched samples, we estimate the average within-market difference between the matched samples for each covariate X (where we first standardize each X) using a random-coefficient model of the form

[9] We tested models allowing the within-market association between Catholic school enrollment and each of a set of family covariates to vary across markets. These covariates were family income, mother's age at the birth of her first child, child's date of birth, birth weight, mother's occupation score, father's occupational score, and number of places child has lived for more than 4 months. Some of these were chosen because local features of the educational system may affect their association (e.g., tuition differences across markets may affect the effect of income on Catholic school enrollment; differences in enrollment age cutoff policies in both Catholic and public kindergartens across markets may affect the association between age and Catholic school enrollment). Others were chosen because balance checks suggested that within-market balance might be improved by allowing the associations to vary across markets. Our final model includes random effects on three of these variables: family income, mother's age at first birth, and date of birth.

$$X_{mi} = (\gamma_0 + \nu_{0m}) + (\gamma_1 + \nu_{1m})(C_{mi} - \bar{C}_m) + \varepsilon_{mi}, \qquad [4]$$

where *m* indexes markets, and where the market-level error terms ν_{0m} and ν_{1m} are assumed multivariate normal with means 0 and an unconstrained variance-covariance matrix. For categorical covariates, we use a similar model with the appropriate non-linear link function. If the treatment and control samples are matched on *X* within each site, we will have $\gamma_1 = 0$ and ${\sigma^2}_{\nu_{1m}} = Var(\nu_{1m}) = 0$. We seek a matching model that minimizes the (or produces a very small) average of the absolute γ_1's and average of the $\sigma_{\nu_{1m}}$'s. As above, there is no analytic solution to this minimization; rather, we fit a number of versions of the multilevel propensity score model—adding higher-order terms, interactions, and/or additional random effects—and match the sample based on each set of estimated propensity scores, seeking the matching model specification that yields the minimum imbalance.

To estimate the average within-market effect of Catholic versus public schooling, we fit two types of models. First, we fit fixed effects models similar to those in models 8 and 9, except that they are based on the within-county matched sample. These models provide estimates of the average within-county effect of Catholic schooling. Unlike models 8 and 9, however, these models do not rely on the additivity assumption of the linear fixed effects model, since they do not rely on between-county information to estimate average within-county effects.

Second, we fit a pair of three-level random-coefficient models of the form

$$y_{ti} = \gamma_t + (\delta_t + \nu_{tm1})\left(C_{mj} - \bar{C}_m\right) + \left(\nu_{tm0} + u_{tmj} + \varepsilon_{tmji}\right). \qquad [5]$$

Here C_{mj} is an indicator of whether school *j* in market *m* is a Catholic school, and $\bar{C}_m$ is the precision-weighed mean value of C_{mj} in market *m*.[10] The coefficient δ_t indicates the average within-

[10] For these models, we construct the weights slightly differently than in the prior models. Specifically, we construct weights as described in Equation 1, save that we do not include the ECLS-K sample weights here (essentially, we constrain each v_i=1), since these weights are not appropriate for use in multi-level models. The ECLS-K panel weights are constructed using county, school, and student level sampling information, and so are not appropriate for use in

market difference in test scores at time *t* between students in Catholic and public schools in the within-market matched samples: it is the average Catholic effect at wave *t*. The error term ν_{tm1} is assumed normally distributed with mean 0 and variance to be estimated; it denotes the difference in the Catholic school effect between market *m* and the overall average within-market Catholic school effect. The remaining three error terms ν_{tm0}, u_{tmj}, and ε_{tmji} are the market-, school-, and student-level error terms, respectively, each assumed independent of the others and normally distributed with mean 0.

The key parameters of the model are δ_t and $Var(\nu_{tm1})$. We test the null hypothesis that the Catholic school effect is constant across markets by testing $\mathrm{H}_0\colon Var(\nu_{tm1}) = 0$. We test the null hypothesis that the average Catholic school effect is 0 by testing $\mathrm{H}_0\colon \delta_t = 0$. As above, we fit two versions of the within-market models, one as described in [5] (Model 12), and one that includes the propensity score and the wave-1 test score as covariates (Model 13)[11] to improve precision and purge the estimates of bias due to imperfect matching.

Heterogeneity of Catholic school effects

Prior research has suggested that Catholic schooling has more positive effects for black and Hispanic students and for poor students than for white and non-poor students. Although we wanted to assess these hypotheses using the ECLS-K data, we were unable to, given the small

---

multilevel models, which require separate sample weights at each level of the model. Given that the multilevel models include county and school information implicitly, as well as a range of information on student characteristics, the only bias that may remain in the estimation is bias due to differential attrition from the ECLS-K sample. Catholic school students remaining in the sample through fifth grade may be different, on average, than those who transferred out of their school. If students who leave Catholic schools have lower average gains than students who stay in Catholic schools (a plausible scenario, because families may remove their children from Catholic schools if they think they are ineffective), then our omission of weights to adjust for sample attrition may lead to overestimation of the effect of Catholic schooling. However, when we fit Models 1-11 with and without the panel weights included, we obtain very similar results in each case, suggesting that the sample attrition is unrelated to potential outcomes and so does not lead to substantial bias.

[11] Specifically, we include the propensity score and wave 1 test score centered around their school means and we include the school-mean propensity score and wave 1 score centered around their precision-weighted market means. In each case, the school means are constructed using the matching weights.

sample sizes of some groups. Table 2 describes the sample sizes, by school type, race/ethnicity, and poverty status, in the 60 market matched sample. Given the small number of poor and black students, in particular, we are unable to obtain precise estimates of Catholic school effects by race/ethnicity and poverty status.

Table 2 here

Comparison of analytic strategies

Each of the analytic strategies we use relies on a different set of assumptions and estimates the effect of Catholic schooling over different populations. A brief summary of these differences follows.

First, it is important to note that regression models (e.g., models 1-5) and propensity score models do not typically estimate the average treatment effect over the same population. The WLS and fixed effects models produces average treatment effect (ATE) estimates—estimates of the average Catholic school effect over the total population represented by the ECLS-K analytic sample (students who were first-time kindergarteners in Catholic or public schools in the Fall of 1998 and who English proficient at the start of kindergarten). The fixed effects models (models 4 and 5) limit the generalizability of the estimates to a population represented in the 60 market sample, but this sample is very similar to that in the full analytic sample (see Table 1), suggesting that these estimates likely generalize to the full analytic population. The propensity score models (models 6-13), in contrast, produce estimates of the average Catholic school effect over the population of Catholic school students for whom public school matches can be found in the data. In other words, these are average-treatment-effect-on-treated (ATT) estimates—estimates of the average Catholic school effect among the type of students who attend Catholic schools (they tell us the extent to which the students in Catholic schools benefit or suffer as a result of attending Catholic school; they do not tell us the extent of benefit/harm that public school students would have experienced had they attended

Catholic schools). If families for whom Catholic schooling would be likely to have a larger effect are families that are most likely to select Catholic schooling (i.e., if the propensity score is related to the expected treatment effect), then we would expect that the average of Catholic schooling will be larger among those with higher propensity to attend Catholic school, meaning that the ATT estimates will likely be more positive than the ATE estimates (assuming the estimates from each model are unbiased).

Second, it is important to note that each model contains different potential sources of bias. As noted above, the regression models may contain bias resulting from both omitted covariates and from incorrect functional form specification. The propensity score models may contain bias resulting from omitted covariates or from imperfect matching (though the models containing the propensity score and wave 1 score as covariates should reduce much of the latter bias). The difference between the within-market matching and the national matching with county fixed effects in the outcome model is that the former matches on student level characteristics, on average, and matches exactly on market characteristics, while the latter matches on student-level covariates, on average, and adjusts for market characteristics using market fixed effects. Because we use a wider caliper (.05 rather than .01) in the within-market matching, the within-market matched samples may differ more on observed covariates than the national matched samples. In contrast, the within-market samples will match exactly on market characteristics, while the national matched samples likely will not, given the unequal distribution of Catholic schooling around the country. The tradeoff in potential bias between models 9 and both of 11 and 13, then, is that model 9 includes better matching on observable student covariates, but relies on regression adjustment to control for market differences, while models 11 and 13 rely on somewhat weaker matching on observable covariates, but ensure exact matching on market/county factors.

Third, it is important to note that the different sample sizes across models means that each of our estimators will have different precision. While we may prefer the models based on within-market matching (models 10-13) from the perspective of bias reduction, these models will have relatively low statistical power—they will not provide very precise estimates of average Catholic schooling effects—because they rely on much smaller samples than the other models. For this reason, it is important to attend to the confidence intervals for each estimate and not just their point estimates.

4. Results

Matching

We begin by describing the balance achieved by the propensity score matching models. Table 3 describes the average values of selected covariates in the Catholic and public matched samples. In general the differences between the Catholic school students and their matched public school counterparts are much smaller than in the full data (the average absolute differences .045 and .077 standard deviations in the national and within-market matched samples, respectively). Importantly, the differences in wave 1 math and reading scores are small and non-significant (approximately .01 standard deviations in math and .10 standard deviations in reading) in the matched samples, despite the fact that we did not use the wave 1 scores in the matching algorithm. This suggests that the matching models are successful at capturing differences between public and Catholic school students in factors that may affect their cognitive development when they enter school.

Table 3 here

Effects on reading scores

Table 4 presents the estimated effects of Catholic schooling on students' reading scores obtained from models 1-13. Model 1 reports unadjusted differences in reading scores between Catholic and public school students, and shows that Catholic school students score better than public school students at each wave in kindergarten through fifth grade. Nonetheless, all but one of the other models indicates that these differences are fully accounted for by observable differences between Catholic and public school students. In fact, the most obvious result from Table 4 is that we find no evidence of even a modest Catholic school effect (positive or negative) on students' reading scores at any wave. Moreover, the largest estimates are from the models predicting wave 1 Catholic school effects. The consistently positive sign on these (and the significance of the wave 1 effect in Model 4) suggest that the covariates and/or matching may not have fully accounted for selection, since it is unlikely that Catholic schooling would have had a large effect so soon after students started kindergarten. Because of this pattern, we focus on the results from models that include the wave 1 score as a covariate (Models 3, 5, 7, 9, 11, and 13).

Table 4 here

Given the size of the estimated standard errors in these models, the models have 80% power to detect reading effects on the order of 1.00 to 1.75 points (.100-.175 standard deviations) in magnitude.[12] Most of the point estimates are negative, but none approach statistical significance.

In addition to providing estimates of the average within-market effect of Catholic schooling, Model 12 and 13 allow us to estimate the variance of the effect across markets. In both models 12 and 13, we cannot reject the null hypothesis that the effect of Catholic schooling is constant across markets (all *p*-values are >.50). We therefore report results for Models 12 and 13 in Table 4 from versions of these models that constrain the effect of Catholic schooling to be constant across markets.

[12] We compute the minimum detectable effect size as 2.80 times the standard error (Bloom 1995).

Effects on math scores

Table 5 presents the estimated effects of Catholic schooling on students' math scores. Here the pattern of estimates is quite different than in the reading models. First, note that covariate adjustment and/or matching generally does a much better job of accounting for the observed Catholic-public school difference in fall kindergarten scores (see Models 2, 4, 6, 8, 10, 12), with the exception of model 4, where the difference is still moderately large. Because of this, we do not observe large differences in the estimated effects at later waves between the models that include the wave 1 score and those that do not. Nonetheless, we focus on the models that include the wave 1 score in each case, since these models have increased precision (smaller standard errors) due to the inclusion of the wave 1 score.

Table 5 here

The regression models (model 3) suggest that Catholic schooling has a negative effect on math scores, with the effect beginning as early as the Spring of kindergarten (δ=-.75) and growing to -2.0 by the spring of third grade. None of the other models find a significant effect this early, however. Nonetheless, each of models 5, 7, and 9 estimates that indicate Catholic schooling has had a negative effect on students' math skills by the spring of third grade (and persisting through fifth grade). These effects are moderate in size, ranging from one-tenth to more than one-fifth of a standard deviation of math scores. Unlike the regression and national propensity score models, the models based on the within-market matching yield estimates that are less negative (closer to zero, or positive) than the other models, and that are not statistically distinguishable from zero.

As above, Models 12 and 13 allow us to test the hypothesis that the effect of Catholic schooling on math scores varies across markets. Again, however, in both models 12 and 13, we cannot reject the null hypothesis that the effect of Catholic schooling is constant across markets (p>.5 in all models). We therefore report results for Models 12 and 13 in Table 5 from versions of these models that constrain the effect of Catholic schooling to be constant across markets.

5. Discussion

Main findings

Perhaps the most striking findings from our analyses are that 1) there is no evidence in any of our models that Catholic schooling has a positive effect on reading or math skills in kindergarten through fifth grade; and 2) Catholic schooling appears, in some of our models, to have a relatively sizeable *negative* effect on Catholic school students' math skills during the period from kindergarten through fifth grade. This negative effect emerges during the period from first to third grade in each of models 3, 5, 7, and 9. This result is consistent with the only other study of the effects of Catholic schooling in elementary school, which found that Catholic school students score, on average, one-half a standard deviation below their observationally similar public school counterparts on NAEP 4^{th} grade math tests (Lubienski and Lubienski 2006). Our estimates are smaller than this (.22 standard deviations in third grade in model 7), but in the same direction. Given the richness of the ECLS-K dataset, our estimates are conditioned on a much more extensive set of covariates (via matching in our case) than the Lubienski and Lubienski NAEP estimates. Lubienski and Lubienski did not estimate Catholic school effects on reading scores, so we cannot compare our (null) findings regarding reading effects with theirs.

Figures 1 & 2 here

Interpreting the estimates from different models

Not all of our models find significant negative effects of Catholic schooling in third and fifth grade, however. Models 7 and 9 yield negative and statistically significant estimates of Catholic school effects on math skills by third grade. Models 11 and 13, however, yield estimates that are likewise negative, but whose standard errors are sufficiently large that we cannot reject the null hypothesis that the effect of Catholic schooling is zero. Given the difference in estimates across models (or at least the differences in whether we can reject the null hypothesis of no effect of Catholic schooling), what can we conclude regarding the effect of Catholic schooling on math skills?

First, consider the different interpretations of the estimates in models 7, 9, 11, and 13. We focus on these models because the matching provides arguably better (or at least no worse) adjustment for confounding that does the covariate adjustment in models 3 and 5. In addition, models 7, 9, 11, and 13 each provide estimates of the same estimand—the average Catholic schooling effect on students enrolled in Catholic schools (ATT estimates)—while models 3 and 5 estimate average Catholic schooling effects over the population (ATE estimates). Model 7 estimates the average effect of attending the Catholic school attended by the average Catholic school student in the nation relative to attending the average public school attended by public school students nationwide who are comparable to Catholic school students. The propensity score estimates simulate an experiment where we assigned Catholic school students to attend public schools all over the country. This is, of course, not a practical experiment, though the estimates may be nonetheless informative. As a national average, they indicate that students enrolled in Catholic schools learn substantially less (.21-.22 standard deviations less in math by 3rd and 5th grade, s.e.=0.05, $p<.001$) than similar students enrolled in public schools. This translates into a difference of roughly 3 months of schooling by third grade and 4 or more months of schooling by fifth grade.[13]

From a policy or parental decision perspective, however, these estimates are not exactly what

[13] A back-of-the envelope calculation provides this translation. Using the IRT theta scores for the ECLS-K tests, we compute that the average student gains roughly .75 on the theta math metric from spring first grade through spring third grade, and gains roughly .42 in the theta math metric from spring third grade through spring fifth grade. The standard deviation of theta scores in third and fifth grade is roughly .40. This means that students gain, on average, roughly one standard deviation per year from first to third grade, and roughly one-half a standard deviation per year from third to fifth. Thus, the fifth of a standard deviation effect size we estimate in third grade translates to either a fifth of a year (2 academic months) difference in first-third grade learning or two-fifths of a year (4 academic months) difference in third-fifth grade learning. Assuming a linear decline in learning rates, this suggests Catholic school students are roughly 3 months behind their public school counterparts in third grade, and 4 or more months behind in fifth grade.

we want. Given the choice to enroll in a local public or Catholic school, what should a parent choose? Models 9, 11, and 13 each attempt to estimate the Catholic school effect relative to local public schooling. This is a more relevant estimand from the point of view of a parent. The estimates from each of these models are uniformly less negative than those from model 7, suggesting that the average Catholic school student lives in a county where the public schools are of lower quality (in terms of teaching math skills) than the average public school nationally. In other words, given their observed characteristics, students appear more likely to be enrolled in Catholic schools in counties where the public schools attended by similar students are of lower quality. This makes sense, as it suggests that the choice of Catholic versus public schooling is based in part on some knowledge of the quality of local public schooling.

Nonetheless, model 9 indicates that Catholic schools are less successful at teaching math than are local public schools (-0.16 standard deviations by fifth grade, s.e.=0.06, $p<.05$), while models 11 and 13 are inconclusive on this point. How are we to resolve the differences between model 9 and models 11 and 13? Although each of models 9, 11, and 13 estimate the within-market Catholic school effect, they do so on the basis of different assumptions and with different statistical power. One the one hand, it may be that the Catholic school effect estimates from model 9 are biased by the lack of within-market matching. The matched samples in model 9 include very few matched students within the same county (only 2% of the Catholic students have matches in model 9 who are in the same county), meaning that the adjustment for county differences using fixed effects relies heavily on the additivity assumption of the linear fixed effects model.

From the perspective of potential bias reduction, our preferred models are 11 and 13. These models eliminate bias due to differences between Catholic and public school students in factors that affect math skill development prior to kindergarten (this is evident in the wave 1 difference in math scores, which is less than .01 standard deviations in the within-market matched sample). They also eliminate bias due to between-county differences between Catholic and public school students without relying on the additivity assumption of the fixed effects model (this is accomplished by the fact that they rely on within-market matching). However, the potential bias reduction of models 11 and 13 comes at the cost of a substantial loss of precision. The estimates from models 11 and 13 have much larger standard errors than our other models, a result of the fact that they rely on a much smaller sample size—the sample of Catholic and public school students who have matches within their same county in the ECLS-K sample. As a result, minimum detectable effect sizes in models 11 and 13 are roughly .25 standard deviations, and the confidence intervals generally span values from -0.25 to +0.10 in third and fifth grade. These estimates are consistent with the possibility that Catholic schools have no effect or even a small effect, but they are also consistent with the possibility that Catholic schools have a moderately-sized negative effect.

Between models 11 and 13, the key difference is the fact that model 11 estimates the *student*-level effect of attending a Catholic school. We can think of it as indicating the average difference in math skills we would observe if students were randomly assigned to attend either Catholic or public schooling, but we let their parents choose which school of their assigned type they enrolled in. Model 13, in contrast, estimates the *school*-level difference between Catholic and public schools. We can think of it as indicating the average difference in math skills we would observe if students were randomly assigned to attend randomly chosen local Catholic schools or a randomly chosen local public school. If parents are good at choosing the better Catholic school among available options, but have no choice regarding public school, then we would expect model 11 to yield more positive comparisons among Catholic schools and local public schools. In math, this is what we observe—the point estimates of Catholic school effects in model 11 are generally more positive (by .05-.10 standard deviations) than the corresponding estimates in model 13. Given the size of the standard errors in these models, however, these differences between models are not statistically significant.

In reading, however, the model 11 estimates are generally more negative than those in model 13, so we caution against making too much of these differences.

If, for example, Catholic schools differ in quality, and our sample sizes are smaller in low-quality public schools, then the between-school differences will differ from the between-student differences, since in the latter each *student* gets equal weight, while in the former each *school* gets equal weight. The fact that the estimates are more positive in the student-level model (11) suggests that our weighted samples are smaller in lower-quality Catholic schools than in higher-quality Catholic schools and/or smaller in higher-quality public schools than in lower-quality public schools. The former would be expected if students are more likely to enroll and stay in higher-quality Catholic schools than lower-quality schools.

In sum, students in Catholic elementary schools, on average, lose ground in math skills and stay even in reading skills by third and fifth grade relative to similar students in public schools nationwide. However, given that Catholic school enrollment is more likely in places where the public schools are less successful at teaching math, our estimates suggest that Catholic students lose somewhat less ground (and maybe lose no ground) in math relative to what they would have learned had they attended their local public schools instead. Our estimates of these local effects are relatively imprecise, however. We cannot rule out the possibility that Catholic schools are equally good as their local public schools; nor can we rule out the possibility that they are significantly worse.

How well do our estimates do at eliminating selection bias?

An issue to consider is the extent to which our models are successful at eliminating selection bias. There are several reasons to believe we that we have been successful at eliminating most, if not, all selection bias through our matching strategies (and less successful in the fixed effects models). First, although we did not match on students' fall kindergarten test scores, the matching models generally show relatively small differences between Catholic and public school students' test scores in the fall of kindergarten. Had we failed to match on some factors that were related to Catholic school enrollment and to students' cognitive development, we would expect to see differences in scores between matched Catholic and public school students as early as fall kindergarten. Because this is too early for Catholic schooling to have had a substantial effect on student scores, evidence of Catholic school "effects" in wave 1 would suggest a failure to match on all covariates relevant to Catholic schooling and cognitive skill. Our matching at wave 1 is particularly good in the math outcome matching models (in models 6, 8, 10, and 12, the fall kindergarten "effects" are all within .01 standard deviations of zero). In the reading models, the Catholic-public differences between matched students are positive and larger (as large as .10 standard deviations) at wave 1, though not significant. Had there been any evidence of reading effects, we would have been skeptical, given the less than perfect matching.

Second, we expect that any remaining selection bias in our within-market estimates would likely tend to bias our estimates upward, so these estimates might be seen as upper bounds on the within-county effects of Catholic schooling (Altonji, Elder, and Taber 2005). Setting aside the uncertainty of the estimates for the moment, the negative coefficients on math scores are difficult to explain away as a result of selection bias—we would have to postulate some unobserved factor(s) that, within a given county and net of the long list of covariates including in the matching, are positively correlated with enrollment in Catholic school, uncorrelated with fall kindergarten test scores, but negatively correlated with math scores in third and fifth grade. In other words, we would have to believe that the students who attend Catholic schools would have had similar math skills in kindergarten but lower math skills by third and fifth grade, had they enrolled in their local public school, than their observationally similar counterparts who were enrolled in public schools in the same county. We think this implausible.

Heterogeneity of Catholic school effects

Given existing research suggesting that Catholic high schools and voucher programs may have larger benefits for minority students than white students, we wanted to investigate whether there was any evidence of hetereogeneity of effects across subgroups. However, small sample sizes limited our ability to use the propensity score matching estimator to estimate effects by subgroups with any useful degree of statistical power. We did, however, test for variation in effects across schooling markets (counties). We found no evidence of variation in Catholic school effects across counties. Finally, because some prior research has found that Catholic high school effects are largest for students with the lowest propensity to attend Catholic schools (Morgan 2001), we conducted some additional analyses (not shown) to investigate whether a similar pattern holds in our data. We found no evidence that the effects vary by propensity score within markets. In all, our results suggest that Catholic schooling exerts a negative effect on math skill development, and this effect is stable (within our power to detect) across the population and across locations.

Limitations and suggestions

Although the within-market matching estimator was successful at eliminating mean differences between Catholic school students and their local public school counterparts, our ability to obtain precise within-market estimates here was limited by the relatively small samples within each county. Although the stratified random sampling of ECLS-K results in a number of counties with sampled Catholic and public schools, a prospective design to estimate such effects would be much stronger if we had larger within-county samples. Better still, we would have liked within-district (rather than within-county) samples, so that we could be sure the Catholic and public school students are actually in the same public school market. Likewise, our ability to estimate effects by subgroup would be improved by oversampling minority and poor students in Catholic schools.

6. Conclusion

We have focused in this paper on obtaining unbiased estimates of the effects of Catholic schooling relative to public schooling. When we use public school students nationwide to provide a counterfactual estimate of how Catholic school students would have performed in public schools, we find that strong evidence indicating that Catholic elementary schools are less successful at teaching math skills than public schools, but no more or less successful at teaching reading skills. However, this conclusion is unambiguous only when we compare Catholic schooling to the national average of public schooling. When we compare Catholic schooling to local public schooling, we obtain estimated math effects that are generally somewhere between the (negative) national estimates and zero (but statistically indistinguishable from either).

These results are somewhat surprising, we think. Certainly nothing in the Catholic high school effects literature hints at negative effects of Catholic school enrollment, even though all prior estimates are like our national estimates—the debate in that literature is between positive effects or no effects. Nonetheless, NAEP data do suggest a negative effect of Catholic schooling on math skills in 4th and 8th grades (Lubienski and Lubienski 2006; Lubienski, Lubienski, and Crane 2007), so our results are not completely without precedent.

With regard to the national estimates, the obvious question is "*why*?" What features of Catholic elementary schooling differ from public schooling that might lead to lower math skills (but no differences in reading skills) among students enrolled in Catholic schools relative to their similar peers in public schools nationwide? Perhaps the most likely explanations are differences in teachers, instruction, and/or curricula. Lubienski, Lubienski, and Crane (2007), for example, find that

Catholic 4th grade schools have fewer certified teachers, spend less time on teacher professional development, and focus their math curricula less on NCTM-recommended topics such geometry, measurement, probability, and algebra than do public 4th grade schools. Moreover, they find that these differences together account for roughly 20% of the adjusted difference in 4th grade NAEP math scores between Catholic and public schools. It is, however, beyond the score of this paper to investigate the mechanism by which Catholic schooling appears to produce lower math skills for students enrolled in Catholic schooling.

Another possible explanation for some of the patterns we observe is that Catholic schools operate in a market where they compete with the public schools for students. In some ways, our estimates are consistent with the idea that parents choose Catholic schooling in part as a response to low-quality public schools, though they are far from conclusive on this point. They may simply choose for reasons that are correlated with public school quality (safety, discipline, etc.). However, we find no evidence here that the Catholic schools parents choose are systematically better than their local public schools, so it is clear that a well-informed comparison of school quality does not solely determine their decisions. It may be that parents have better information about the quality of public schools than of Catholic schools; it may be that parents choose for reasons other than school quality in regard to math teaching (e.g., religious preference, discipline, school safety, etc.). Beyond speculation, however, we can say no more on the basis of the evidence presented here.

References

Abadie, Alberto and Guido W. Imbens. 2002. "Simple and bias-corrected matching estimators for average treatment effects." National Bureau of Economic Research.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2002. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools." National Bureau of Economic Research, Cambridge, Massachusetts.

—. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy* 113:151-184.

Baker, David P. 1999. "Schooling all the masses: Reconsidering the origins of American schooling in the postbellum era." *Sociology of Education* 72:197-215.

Baker, David P. and Cornelius Riordan. 1998. "The 'eliting' of the common American Catholic school and the national education crisis." *Phi Delta Kappan* 80:16-23.

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Power of Experimental Designs." *Evaluation Review* 19:547-556.

Broughman, Stephen P. and Nancy L. Swaim. 2006. "Characteristics of Private Schools in the United States: Results From the 2003–2004 Private School Universe Survey." U.S. Department of Education. National Center for Education Statistics., Washington, DC.

Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland. 1993. *Catholic Schools and the Common Good.* New York: Basic Books.

Campbell, David E., Martin R. West, and Paul E. Peterson. 2005. "Participation in a national, means-tested school voucher program." *Journal of Policy Analysis and Management* 24:523-541.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982a. "Achievement and segregation in secondary schools: A further look at public and private school differences." *Sociology of Education* 55:162-182.

—. 1982b. "Cognitive outcomes in public and private schools." *Sociology of Education* 55:65-76.

—. 1982c. *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic Books.

Evans, William N. and Robert M. Schwab. 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *The Quarterly Journal of Economics* 110:941-974.

Goldberger, Arthur S. and Glen G. Cain. 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report." *Sociology of Education* 55:103-122.

Grogger, Jeffrey and Derek Neal. 2000. "Further Evidence on the Effects of Catholic Secondary Schooling." *Brookings-Wharton Papers on Urban Affairs*:151-201.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15:199-236.

Hoffer, Thomas, Andrew M. Greeley, and James S. Coleman. 1985. "Achievement Growth in Public and Catholic Schools." *Sociology of Education* 58:74-97.

Howell, William G. and Paul E. Peterson. 2006. *The education gap: Vouchers and urban schools*. Washington, DC: Brookings Instutution.

Imai, Kosuke, Gary King, and Elizabeth A. Stuart. forthcoming. "Misunderstandings among experimentalists and observationalists about causal inference." *Journal of the Royal Statistical Society, Series A*.

Imbens, Guido W. 2004. "Nonparametric estimation of average treatment effects under exogeneity: A review." *The Review of Economics and Statistics* 86:4-29.

Lubienski, Christopher and Sarah Thule Lubienski. 2006. "Charter, Private, Public Schools and Academic Achievement: New Evidence From NAEP Mathematics Data." National Center for the Study of Privatization in Education, Teachers College, Columbia University, New York.

Lubienski, Sarah Thule, Christopher Lubienski, and Corrinna Crawford Crane. 2007. "Understanding Achievement Differences in Public and Private Schools: The Roles of School Climate, Teacher Certification, and Teaching Practices." in *Annual Meeting of the American Education Research Association*. Chicago, IL.

McEwan, Patrick J. and Martin Carnoy. 2000. "The effectiveness and efficiency of private schools in Chile's voucher system." *Educational Evaluation and Policy Analysis* 22:213-239.

Morgan, Stephen L. 2001. "Counterfactuals, causal effect heterogeneity, and the Catholic school effect on learning." *Sociology of Education* 74:341-373.

Morgan, Stephen L. and Jennifer J. Todd. 2007. "A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects: With a Demonstration from School Effects Research." Cornell University.

Murnane, Richard J., S. Newstead, and R. J. Olsen. 1985. "Comparing public and private schools: The puzzling role of selection bias." *Journal of Business and Economic Statistics* 3:23-35.

Neal, Derek A. 1997. "The Effects of Catholic Secondary Schooling on Educational Achievement." *Journal of Labor Economics* 15:98-123.

Pollack, Judith M., Michelle Narajian, Donald A. Rock, Sally Atkins-Burnett, and Elvira Germino Hausken. 2005. "Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for the Fifth Grade." U.S. Department of Education, National Center for Education Statistics, Washington, DC.

Pollack, Judith M., Donald A. Rock, Michael J. Weiss, Sally Atkins-Burnett, Karen Tourangeau, Jerry West, and Elvira Germino Hausken. 2005. "Early Childhood Longitudinal Study-

Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for the Third Grade." U.S. Department of Education, National Center for Education Statistics, Washington, DC.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

Raudenbush, Stephen W., Anthony S. Bryk, Yuk Fai Cheong, and Richard Congdon. 2005. *Hierarchical Linear and Non-linear Modeling, Version 6: Users' Guide and Software Program*. Chicago: Scientific Software International.

Reardon, Sean F. 2006. "Estimating Effects of Educational Practice and Policy Across Multiple Sites Using a Multisite Matching Estimator." in *Annual Meeting of the American Educational Research Association*. San Francisco, CA.

—. 2007. "Thirteen ways of looking at the black-white test score gap."

Reardon, Sean F. and John T. Yun. 2002. "Private school racial enrollments and segregation." Civil Rights Project, Harvard University, Cambridge, MA.

Rock, Donald A. and Judith M. Pollack. 2002. "Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade." U.S. Department of Education, National Center for Education Statistics, Washington, DC.

Rosenbaum, Paul R., and Rubin, Donald B. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70:41-55.

Rouse, Cecilia E. 1998. "Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program." *The Quarterly Journal of Economics* 113:553-602.

TABLE 1: Selected Characteristics of Catholic and Public School Students in Analytic Samples

	Full Analytic Sample				60 Market Sample			
Variable description	Catholic Mean	Public-Catholic Difference		Standardized Difference	Catholic Mean	Public-Catholic Difference		Standardized Difference
Wave-1 math t-score	54.505	-4.195	**	-0.448	55.258	-4.718	**	-0.504
Wave-1 reading t-score	54.466	-4.949	**	-0.518	55.106	-5.494	**	-0.575
Family income (in $10,000)	7.074	-2.340	**	-0.533	7.114	-2.055	**	-0.468
Mother's occupational score	33.958	-5.805	**	-0.263	34.731	-6.142	**	-0.278
Mother's age at child's birth	31.564	-7.320	**	-0.332	32.303	-5.703	**	-0.259
Child ever attended center care	0.845	-0.183	**	-0.401	0.855	-0.173	**	-0.379
Ever receive food stamps	0.033	0.081	**	0.234	0.029	0.079	**	0.227
Mother's age at first child birth	25.840	-3.237	**	-0.548	26.140	-3.210	**	-0.554
Family at or below poverty line	0.028	0.132	**	0.357	0.023	0.123	**	0.332
Average Absolute Standardized Difference				0.308				0.306
Sample Size	1,078	6,402			1,001	2,554		
School Count	97	692			89	285		

Notes: Sample means are weighted by ECLS-K panel weight *c1_6fc0*. **p<.01. Standardized differences are computed by dividing the public-Catholic difference by the pooled standard deviations of the Catholic and public school student samples in the full analytic sample. For Catholic-public sample means on the full list of covariates included in the analyses, see Appendix Table A1. Average absolute standardized difference is the average of the absolute value of the standardized public-Catholic difference in each of the variables listed in the top panel of Table A1 (which includes more covariates than those listed here).

TABLE 2: Available Analytic Sample Sizes, by Poverty, Race/Ethnicity, and School Sector

	Public	Catholic	Total
Poverty status			
Not in poverty	1,838	968	2,806
In poverty	149	27	176
Race			
White	1,382	744	2,126
Black	80	38	118
Hispanic	323	121	444
Asian	134	55	189
Other	68	37	105
Total	1,987	995	2,982

Sample based on model 8 and 9 sample (nationally matched sample restricted to 60 markets).

TABLE 3: Selected Characteristics of Catholic and Public School Students in Matched Samples

Variable description	National Matched Sample: Catholic Mean	National Matched Sample: Public-Catholic Difference	National Matched Sample: Standardized Difference	Within-Market Matched Sample: Catholic Mean	Within-Market Matched Sample: Public-Catholic Difference	Within-Market Matched Sample: Standardized Difference
Wave-1 math t-score	54.515	0.109	0.012	54.492	-0.059	-0.006
Wave-1 reading t-score	54.480	-0.631	-0.066	54.373	-1.183	-0.124
Family income (in $10,000)	7.066	0.151	0.034	6.849	0.556	0.127
Mother's occupational score	33.919	-0.548	-0.025	33.890	1.474	0.067
Mother's age at child's birth	31.570	0.214	0.010	31.556	1.814	0.082
Child ever attended center care	0.870	-0.037	-0.062	0.879	-0.034	-0.074
Ever receive food stamps	0.048	-0.008	-0.024	0.051	-0.042 **	-0.121
Mother's age at first child birth	25.829	-0.004	-0.001	25.573	0.661	0.112
Family at or below poverty line	0.035	-0.008	-0.021	0.037	-0.021	-0.056
Average Absolute Standardized Difference			0.045			0.077
Sample Size	1,073	3,744		744	478	
School Count	96	359		61	50	

Notes: Catholic school student samples are weighted by ECLS-K panel weight *c1_6fc0*. Public school student samples are weighted their matching weights (see text for details). Within-market differences are estimated by a model like model 10 (see text). Except as noted (** p<.10), all public-Catholic differences are statistically insignificant (p>.05). Standardized differences are computed by dividing the public-Catholic difference by the pooled standard deviations of the Catholic and public school student samples in the full analytic sample. For Catholic-public sample means on the full list of covariates included in the analyses, see Appendix Table A2. Average absolute standardized difference is the average of the absolute value of the standardized public-Catholic difference in each of the variables listed in the top panel of Table A2 (which includes more covariates than those listed here), excluding the propensity score.

TABLE 4: Estimated Effects of Catholic Schooling on Reading Scores, by Wave and Model

	WLS			WLS with market fixed effects		National propensity score matching				Within-market propensity score matching			
Wave	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Fall Kindergarten	4.949**	0.919		1.589*		0.631		0.570		1.183		0.719	
	(0.726)	(0.701)		(0.753)		(0.809)		(0.850)		(0.911)		(0.902)	
Spring Kindergarten	3.663**	0.277	-0.392	1.173	-0.009	0.124	-0.320	0.526	0.057	0.543	-0.293	0.480	0.400
	(0.711)	(0.657)	(0.395)	(0.804)	(0.496)	(0.766)	(0.450)	(0.770)	(0.493)	(0.982)	(0.559)	(0.783)	(0.742)
Spring 1st Grade	3.156**	0.214	-0.311	0.879	-0.052	-0.096	-0.430	0.317	0.096	-0.168	-0.805	-0.242	-0.382
	(0.644)	(0.573)	(0.415)	(0.739)	(0.553)	(0.632)	(0.470)	(0.735)	(0.580)	(0.844)	(0.583)	(0.706)	(0.687)
Spring 3rd Grade	3.573**	0.162	-0.300	0.878	0.060	-0.200	-0.495	0.151	-0.231	0.226	-0.225	-0.016	-0.006
	(0.632)	(0.502)	(0.375)	(0.626)	(0.500)	(0.619)	(0.482)	(0.606)	(0.477)	(0.829)	(0.686)	(0.682)	(0.666)
Spring 5th Grade	3.845**	0.052	-0.390	0.778	0.028	-0.262	-0.556	0.324	-0.104	0.872	0.459	0.551	0.556
	(0.616)	(0.485)	(0.361)	(0.606)	(0.483)	(0.621)	(0.487)	(0.631)	(0.537)	(0.995)	(0.883)	(0.732)	(0.727)
Included covariates													
Covariates included		*Yes*	*Yes*	*Yes*	*Yes*								
Propensity score							*Yes*		*Yes*		*Yes*		*Yes*
Wave-1 test score			*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*
Market fixed effects				*Yes*	*Yes*			*Yes*	*Yes*	*Yes*	*Yes*		
Sample													
Matched sample						*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
60 market sample				*Yes*	*Yes*			*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Within-market match										*Yes*	*Yes*	*Yes*	*Yes*
N(catholic)	1,078	1,078	1,078	1,001	1,001	1,073	1,073	995	995	744	744	744	744
N(public)	6,402	6,402	6,402	2,554	2,554	3,744	3,744	1,987	1,987	478	478	478	478

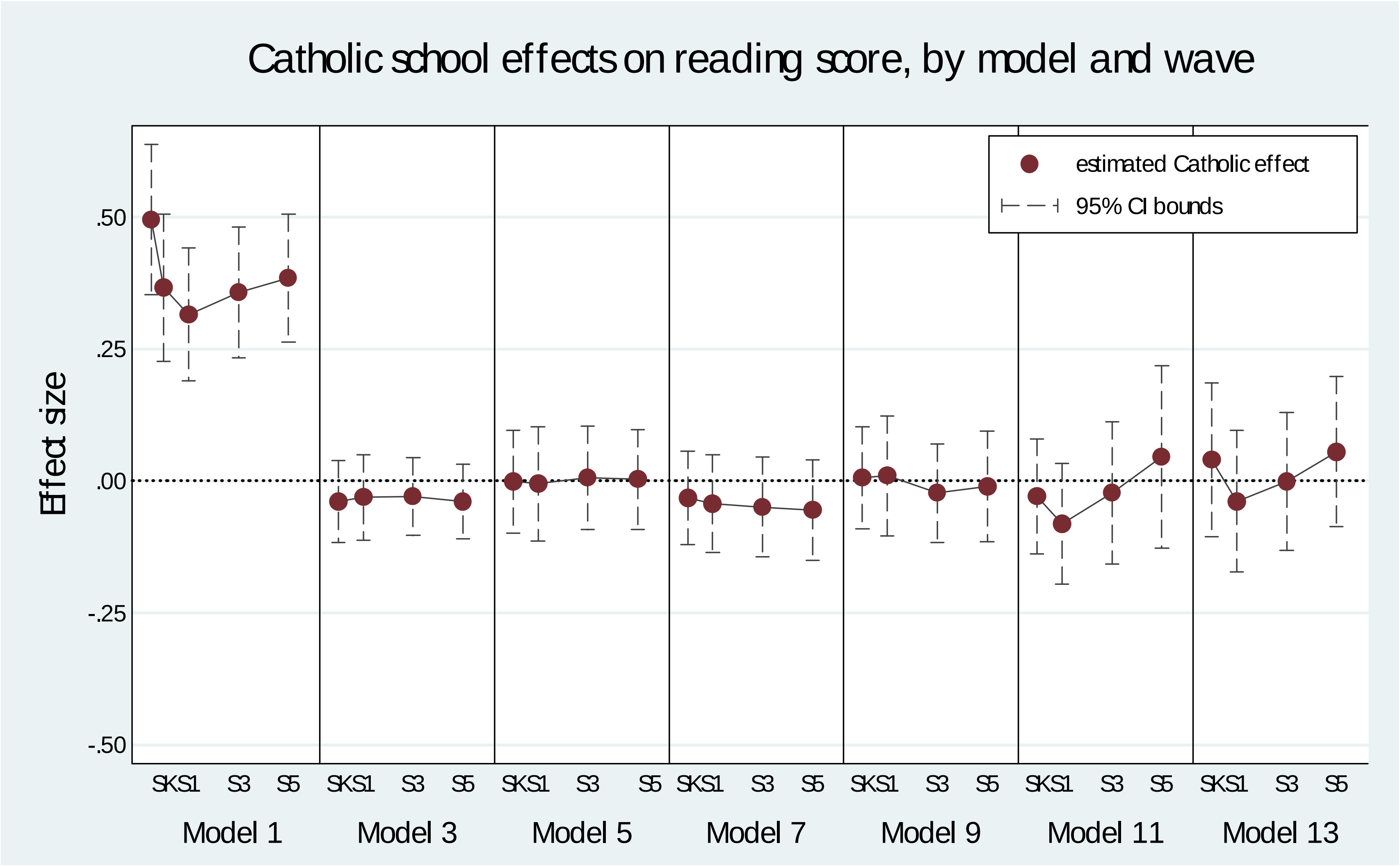
Models 1-5 include ECLS-K sampling weights (*c1_6fc0*). Models 6-11 include matching weights based on Equation 1. Models 12-13 include matching weights constructed without ECLS-K sampling weights. Models 12 and 13 do not include a random effect on Catholic school. Models 10-13 are based on within-market matched sample within 44 counties with at least 5 matched public school students. Standard errors (in parentheses) are corrected for clustering at the county level. See text for details on samples, models, and construction of matching weights.

TABLE 5. Estimated Effects of Catholic Schooling on Math Scores, by Wave and Model

	WLS			WLS with market fixed effects		National propensity score matching				Within-market propensity score matching			
Wave	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Fall Kindergarten	4.195**	0.360		0.936		-0.109		0.097		0.059		-0.018	
	(0.550)	(0.479)		(0.612)		(0.536)		(0.637)		(0.607)		(0.847)	
Spring Kindergarten	3.290**	-0.474	-0.748*	0.341	-0.394	-0.671	-0.590	-0.139	-0.303	0.122	0.106	0.313	0.503
	(0.631)	(0.570)	(0.352)	(0.695)	(0.395)	(0.662)	(0.444)	(0.753)	(0.492)	(0.803)	(0.617)	(0.807)	(0.771)
Spring 1st Grade	2.618**	-0.681	-0.891**	0.449	-0.090	-0.916~	-0.853*	-0.324	-0.511	0.690	0.720	-0.357	-0.295
	(0.552)	(0.461)	(0.340)	(0.636)	(0.465)	(0.545)	(0.388)	(0.671)	(0.481)	(0.973)	(0.832)	(0.804)	(0.780)
Spring 3rd Grade	1.784**	-1.780**	-2.025**	-0.604	-1.243**	-2.230**	-2.168**	-1.412*	-1.567**	-0.869	-0.883	-1.209	-1.324
	(0.556)	(0.502)	(0.403)	(0.610)	(0.477)	(0.588)	(0.475)	(0.689)	(0.545)	(1.095)	(0.947)	(0.920)	(0.924)
Spring 5th Grade	2.118**	-1.710**	-1.935**	-0.390	-0.970*	-2.152**	-2.088**	-1.381~	-1.565*	-0.359	-0.347	-0.814	-0.895
	(0.570)	(0.499)	(0.399)	(0.629)	(0.492)	(0.647)	(0.526)	(0.757)	(0.609)	(0.998)	(0.880)	(0.872)	(0.890)
Included covariates													
Covariates included		*Yes*	*Yes*	*Yes*	*Yes*								
Propensity score							*Yes*		*Yes*		*Yes*		*Yes*
Wave-1 test score			*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*
Market fixed effects				*Yes*	*Yes*			*Yes*	*Yes*	*Yes*	*Yes*		
Sample													
Matched sample						*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
60 market sample				*Yes*	*Yes*			*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Within-market match										*Yes*	*Yes*	*Yes*	*Yes*
N(catholic)	1,078	1,078	1,078	1,001	1,001	1,073	1,073	995	995	744	744	744	744
N(public)	6,402	6,402	6,402	2,554	2,554	3,744	3,744	1,987	1,987	478	478	478	478

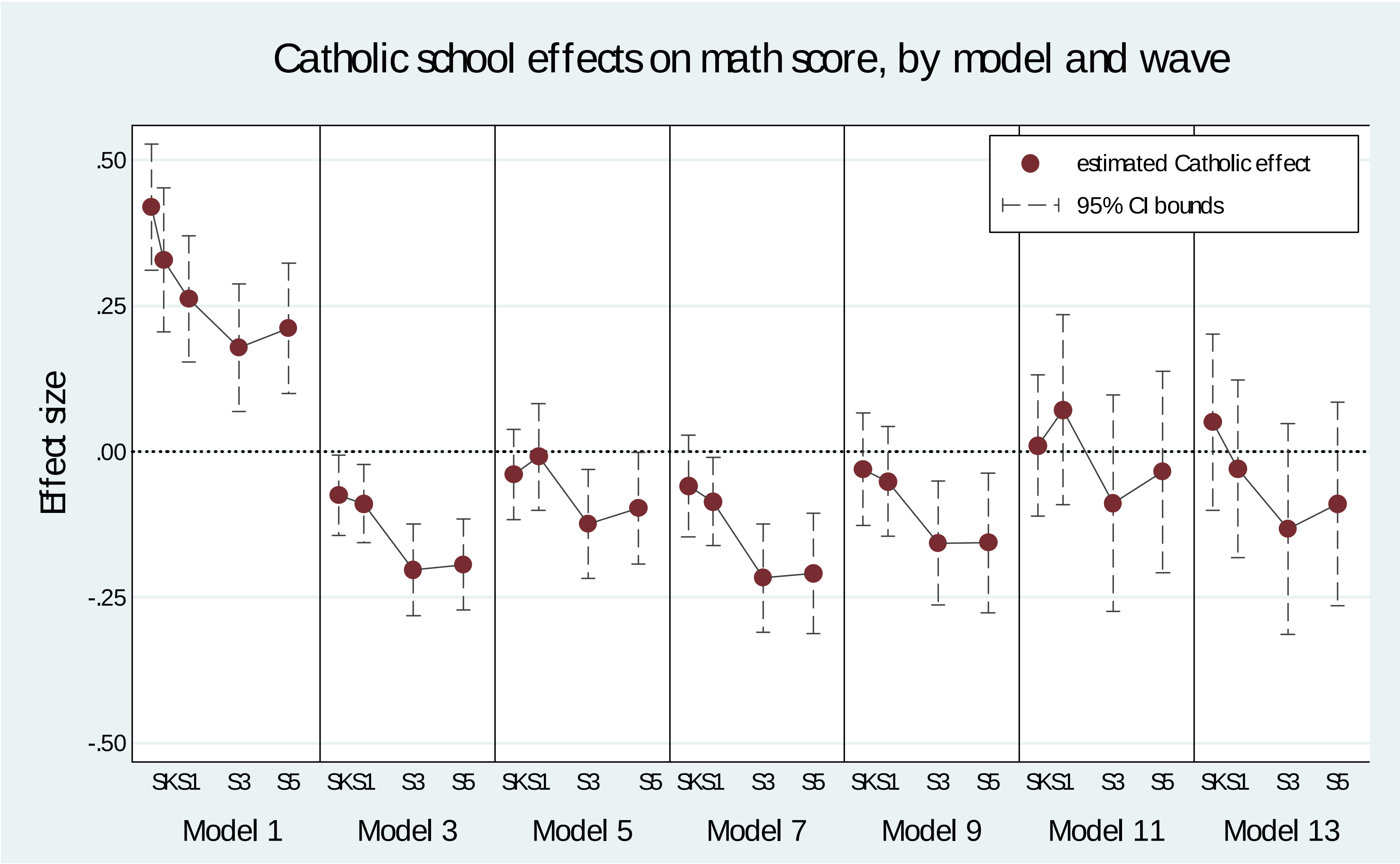
Models 1-5 include ECLS-K sampling weights (*c1_6fc0*). Models 6-11 include matching weights based on Equation 1. Models 12-13 include matching weights constructed without ECLS-K sampling weights. Models 12 and 13 do not include a random effect on Catholic school. Models 10-13 are based on within-market matched sample within 44 counties with at least 5 matched public school students. Standard errors (in parentheses) are corrected for clustering at the county level. See text for details on samples, models, and construction of matching weights.

Figure 1



Catholic school effects on reading score, by model and wave
estimated Catholic effect
95% CI bounds
Effect size
.50
.25
.00
-.25
-.50
SKS1
S3
S5
Model 1
Model 3
Model 5
Model 7
Model 9
Model 11
Model 13

Figure 2



Catholic school effects on math score, by model and wave
Effect size
.50
.25
.00
-.25
-.50
estimated Catholic effect
95% CI bounds
SKS1 S3 S5
Model 1
Model 3
Model 5
Model 7
Model 9
Model 11
Model 13

Appendix Table A1

	Full sample				Sample restricted to 60 markets			
	Catholic	Public-Catholic	p-value	Standardized difference	Catholic	Public-Catholic	p-value	Standardized difference
Variable description								
Propensity score	0.223	-0.114	**	-1.084	0.231	-0.114	**	-1.080
Wave-1 math t-score	54.505	-4.195	**	-0.448	55.258	-4.718	**	-0.504
Wave-1 reading t-score	54.466	-4.949	**	-0.518	55.106	-5.494	**	-0.575
Date of birth (measured in days)	80.642	-3.422		-0.028	71.204	-0.981		-0.008
Father's occupational score	39.817	-9.748	**	-0.475	40.437	-9.661	**	-0.471
Family income (in $10,000)	7.074	-2.340	**	-0.533	7.114	-2.055	**	-0.468
Entered kindergarten late (0/1)	0.082	-0.017		-0.070	0.090	-0.024	*	-0.099
Mother's occupational score	33.958	-5.805	**	-0.263	34.731	-6.142	**	-0.278
Number of places child has lived for at least 4 months	1.754	0.481	**	0.359	1.733	0.498	**	0.371
Mother's age at child's birth (in years)	31.564	-7.320	**	-0.332	32.303	-5.703	**	-0.259
Child ever attended center care (0/1)	0.845	-0.183	**	-0.401	0.855	-0.173	**	-0.379
Child ever attended pre-k care (0/1)	0.789	-0.214	**	-0.442	0.804	-0.201	**	-0.415
Child in center care now (0/1)	0.248	-0.070	*	-0.175	0.250	-0.079	**	-0.197
Ever receive food stamps (0/1)	0.033	0.081	**	0.234	0.029	0.079	**	0.227
Father's current age (in years)	36.788	-4.146	**	-0.388	37.095	-4.055	**	-0.380
Mother's age at first child birth (in years)	25.840	-3.237	**	-0.548	26.140	-3.210	**	-0.544
Mother's current age (in years)	34.757	-2.236	**	-0.330	34.916	-1.867	**	-0.275
Child was in non-relative care on a regular basis the year before entering kindergarten (0/1)	0.181	-0.017		-0.047	0.187	-0.016		-0.045
Child's age at first non-relative care	7.313	0.293		0.018	6.400	1.440		0.089
Child ever received non-relative care on a regular basis (0/1)	0.383	-0.045	*	-0.094	0.389	-0.043	*	-0.090
Child now in non-relative care (not center) (0/1)	0.117	0.019		0.056	0.118	0.026	*	0.077
Number of non-relative care arrangements year before kindergarten	0.159	0.032		0.068	0.161	0.044		0.093
Number of months of serious financial problems since child's birth	0.448	0.430	**	0.236	0.425	0.475	**	0.260
Number of months received TANF/AFDC	0.023	0.382	**	0.360	-0.008	0.376	**	0.355
Number of months received food stamps	0.055	0.634	**	0.478	-0.022	0.689	**	0.520
Received WIC benefits for child (0/1)	0.119	0.301	**	0.626	0.105	0.266	**	0.553
Received WIC benefits when pregnant (0/1)	0.103	0.264	**	0.554	0.096	0.228	**	0.479
Family at or below poverty line (0/1)	0.028	0.132	**	0.357	0.023	0.123	**	0.332
Number of months child lived in current home	49.247	-10.494	**	-0.425	49.247	-10.305	**	-0.418
Child birthweight (in ounces)	119.788	-1.797		-0.085	120.557	-2.223		-0.105
Average standardized difference				**0.308**				**0.306**

Primary type of center care			**			**
day care	0.107	0.000		0.099	0.004	
pre-school	0.499	-0.238		0.508	-0.213	
pre-kindergarten	0.184	-0.026		0.187	-0.037	
Father's education			**			**
less than high school	0.034	0.070		0.036	0.049	
high school diploma or vo/tech	0.254	0.075		0.260	0.077	
at least some college	0.615	-0.257		0.630	-0.263	
Mother's education			**			**
less than high school	0.008	0.115		-0.004	0.146	
high school diploma	0.177	0.150		0.168	0.134	
vo/tech	0.047	0.000		0.048	-0.005	
some college	0.329	-0.044		0.336	-0.067	
bachelor's degree or higher	0.427	-0.224		0.438	-0.208	
Parent type			**			**
biological mother & other father	0.056	0.030		0.056	0.040	
biological mother only	0.098	0.136		0.077	0.154	
biological father, but not biological mother	0.014	0.012		0.018	0.008	
other	0.021	0.016		0.021	0.003	
Race			**			**
Black, not Hispanic	0.058	0.125		0.039	0.093	
Hispanic	0.108	0.037		0.146	0.023	
Asian	0.024	0.002		0.026	0.003	
Other, not Hispanic	0.052	-0.016		0.048	-0.010	

Appendix Table A2

	National Matched Sample				Within-Market Matched Sample			
	Catholic	Public-Catholic	p-value	Standardized difference	Catholic	Public-Catholic	p-value	Standardized difference
Variable description								
Propensity score	0.223	0.000		0.000	0.211	0.006		0.059
Wave-1 math t-score	54.515	0.109		0.012	54.492	-0.059		-0.006
Wave-1 reading t-score	54.480	-0.631		-0.066	54.373	-1.183		-0.124
Date of birth (measured in days)	80.281	-10.851		-0.090	78.613	-11.400		-0.094
Father's occupational score	39.820	0.042		0.002	39.809	-0.418		-0.020
Family income (in $10,000)	7.066	0.151		0.034	6.849	0.556		0.127
Entered kindergarten late (0/1)	0.070	0.005		0.020	0.073	0.023		0.096
Mother's occupational score	33.919	-0.548		-0.025	33.890	1.474		0.067
Number of places child has lived for at least 4 months	1.756	-0.031		-0.023	1.744	-0.114		-0.085
Mother's age at child's birth (in years)	31.570	0.214		0.010	33.849	-4.568	*	-0.207
Child ever attended center care (0/1)	0.870	-0.037		-0.080	0.879	-0.034		-0.074
Child ever attended pre-k care (0/1)	0.804	-0.018		-0.038	0.821	-0.030		-0.061
Child in center care now (0/1)	0.258	-0.027		-0.068	0.247	0.055		0.137
Ever receive food stamps (0/1)	0.048	-0.008		-0.024	0.051	-0.042	**	-0.121
Father's current age (in years)	36.777	0.318		0.030	36.643	0.355		0.033
Mother's age at first child birth (in years)	25.829	-0.004		-0.001	25.573	0.661		0.112
Mother's current age (in years)	34.751	0.196		0.029	34.636	0.164		0.024
Child was in non-relative care on a regular basis the year before entering kindergarten (0/1)	0.139	0.049	*	0.136	0.153	0.029		0.079
Child's age at first non-relative care	7.330	-1.056		-0.065	8.062	-2.582		-0.159
Child ever received non-relative care on a regular basis (0/1)	0.408	-0.017		-0.035	0.433	-0.028		-0.059
Child now in non-relative care (not center) (0/1)	0.109	0.018		0.054	0.126	0.002		0.007
Number of non-relative care arrangements year before kindergarten	0.159	0.074	**	0.157	0.176	0.046		0.096
Number of months of serious financial problems since child's birth	0.449	0.023		0.012	0.496	-0.157		-0.086
Number of months received TANF/AFDC	0.023	0.131		0.124	0.015	0.030		0.028
Number of months received food stamps	0.055	0.129		0.097	0.035	0.120	*	0.091
Received WIC benefits for child (0/1)	0.141	-0.009		-0.018	0.133	-0.022		-0.045
Received WIC benefits when pregnant (0/1)	0.117	-0.013		-0.028	0.112	-0.020		-0.043
Family at or below poverty line (0/1)	0.035	-0.008		-0.022	0.037	-0.021		-0.056
Number of months child lived in current home	49.222	-0.170		-0.007	49.310	2.494		0.101
Child birthweight (in ounces)	119.757	-0.105		-0.005	119.672	-0.157		-0.007
Average standardized difference				**0.045**				**0.077**

Primary type of center care				
day care	0.105	-0.023	0.095	0.055
pre-school	0.498	0.008	0.542	-0.063
pre-kindergarten	0.185	-0.013	0.165	-0.019
Father's education				
less than high school	0.035	-0.014	0.021	-0.009
high school diploma or vo/tech	0.252	0.004	0.287	0.031
at least some college	0.619	0.027	0.601	-0.022
Mother's education				
less than high school	0.008	0.001	0.007	-0.003
high school diploma	0.175	0.022	0.166	0.102
vo/tech	0.049	-0.009	0.057	0.001
some college	0.331	-0.078	0.355	-0.109
bachelor's degree or higher	0.424	0.073	0.399	0.020
Parent type				
biological mother & other father	0.059	-0.022	0.051	-0.027
biological mother only	0.096	-0.017	0.093	-0.003
biological father, but not biological mother	0.015	-0.010	0.019	-0.011
other	0.022	-0.008	0.024	-0.014
Race				
Black, not Hispanic	0.058	-0.028	0.048	0.004
Hispanic	0.107	0.049	0.106	0.001
Asian	0.024	0.030	0.028	0.037
Other, not Hispanic	0.051	-0.014	0.046	-0.014