Program on Education Policy and Governance Working Papers Series

PEPG 15-04

Assessing Tradeoffs between Observational and Experimental Designs for Charter School Research[1]

Matthew Ackerman
Anna J. Egalite

Harvard Kennedy School
79 JFK Street, Taubman 304
Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg

Abstract

When lotteries are infeasible, researchers must rely on observational methods to estimate charter effectiveness at raising student test scores. Considerable attention has been paid to observational studies by the Stanford Center for Research on Education Outcomes (CREDO), which have analyzed charter performance in 27 states. However, the sensitivity of CREDO's findings to its methodology has not been subject to a thorough, independent evaluation. Our analysis of the CREDO estimates of charter effectiveness in Florida between 2001 and 2009 reveals four main findings that increase confidence in its results. 1) Its use of multiple observations when matching students does not materially affect results. 2) Student participation in programs such as special education and free/reduced price lunch is inconsistently measured across sectors, but matching on these variables only modestly affects CREDO's estimates. 3) Exogenous instrumental variable estimates of charter effectiveness produce qualitatively similar results to CREDO's observational estimates. 4) Impact estimates differ for oversubscribed and undersubscribed charters, which helps explain why lottery-based studies tend to find larger charter impacts than CREDO does.

[1] The authors would like to thank the Florida Department of Education and Florida Charter Alliance for providing us with the necessary data for this analysis. We are also deeply appreciative of Marc Piopiunik for his guidance and for helpful comments on an earlier draft of this paper.

1. Introduction

Since the first charter school was established in 1991, the number of these publicly funded but independently operated schools has been steadily increasing. By 2014, there were over 6,000 charter schools nationwide, serving over 2.5 million students (National Alliance for Public Charter Schools, 2015). In keeping with the nationwide trend, the state of Florida has seen a significant rise in the number of charter schools. By the 2014-15 school year, Florida had approximately 646 charter schools, serving more than 251,000 students (Florida Department of Education, 2015).

As independently-operated institutions, charter schools are not required to adhere to the same accountability guidelines to which public schools are subject, agreeing to an alternative set of provisions instead. This provides charter school administrations with the flexibility to implement unique curricula and school organizational structures. Given the large rise in the charter school movement and the use of public funds for these privately-managed schools, it is important to establish the implications of this accountability-for-autonomy trade-off by evaluating the effectiveness of charter schools.

While many studies of charter school impacts have been conducted, findings vary, often depending on whether they rely upon an experimental, quasi-experimental, or observational design. Lottery-based experimental studies generally find large, positive impacts on student test score performance (Abdulkadiroğlu, Angrist, Cohodes, Dynarski, Fullerton, Kane, & Pathak, 2009; Angrist, Cohodes, Dynarski, Fullerton, Kane, Pathak, & Walters, 2011; Angrist, Dynarski, Kane, Pathak, & Walters, 2010; Angrist, Pathak & Walters, 2013; Dobbie & Fryer, 2011; Gleason, Clark, Tuttle, & Dwoyer, 2010; Hastings, Neilson & Zimmerman, 2012; Hoxby, Murarka & Kang, 2009; Hoxby & Rockoff, 2004; Tuttle, Gill, Gleason, Knechtel, Nichols-

Barrer, & Resch, 2013). Meanwhile, impact estimates from observational studies, including those from the Center for Research on Education Outcomes (CREDO, 2009a, 2013a, 2015), tend to point in the opposite direction, finding very small positive overall effects and negative impacts for charters in a substantial number of states (Abdulkadiroğlu et al., 2009; Bettinger, 2005; Bifulco & Ladd, 2006; Booker, Gilpatric, Gronberg & Janson, 2007; Chingos & West, 2015; Davis & Raymond, 2012; Hanushek, Kain, Rivkin & Branch, 2007; Imberman, 2011; Mills, 2013; Sass, 2006; Witte, Wolf, Carlson & Dean, 2012; Zimmer & Buddin, 2006; Zimmer, Gill, Booker, Lavertu & Witte, 2012).

As tools for evaluating charters, there are important tradeoffs involved in choosing one empirical approach over another. Experimental studies typically compare students who have won and lost admissions lotteries at popular charter schools. This identification strategy can yield unbiased estimates of the local average treatment effect for oversubscribed schools, yet such studies can potentially suffer from problems of external validity, as oversubscribed charter schools might differ from undersubscribed charters. Similarly, quasi-experimental studies, such as an instrumental variable approach, approximate an experiment's robust identification strategy. However, reliable and valid instruments are hard to find, particularly in large-scale evaluations of charters across multiple locales. Observational studies, meanwhile, permit the researcher to include most charter students in a region, thereby enhancing their external validity. Yet these studies may have weak internal validity because of the challenges associated with establishing an appropriate counterfactual. In this paper, we evaluate the strengths and limitations of a particularly influential observational design—a synthetic matching method used by the Stanford Center for Research on Education Outcomes (CREDO)—to assess the sensitivity of its findings with respect to the assumptions underlying the model.

The CREDO reports of charter school effectiveness have had a significant impact on education policy in recent years. CREDO (2013a) generates a synthetic control group against which to judge charter effectiveness, analyzing test scores for millions of students across 27 states to produce the largest and most influential charter school evaluation to date. The CREDO reports have been highlighted by national media outlets including the *Wall Street Journal, New York Times,* and *USA Today*, as well as by numerous federal and state-level policymakers. Nonetheless, the model has been heartily criticized by experimental researchers and others who doubt the validity of their substantive findings (e.g., Hoxby, 2009).

We assess the strengths and limitations of the CREDO analytical model by using student-level data for the years 2001 to 2009 from the State of Florida to replicate, to the extent possible, the results from the CREDO study for that state (CREDO, 2013b). We then employ alternative models to assess the sensitivity of the CREDO approach. The purpose of this study is not to generate contemporary estimates of charter school effectiveness in Florida, as a dataset spanning 2001 through 2009 would be outdated. Rather, it is to evaluate the robustness of the CREDO results to alternative specifications. We examine: 1) whether CREDO's estimates are sensitive to its use of "virtual control records" (VCR's) that are generated from the records of up to seven public school "virtual twins", instead of one-to-one matching. 2) Whether CREDO's estimates are biased by its use of inconsistent program variables, such as student participation in the National School Lunch Program, as matching and control variables. 3) The internal validity of the CREDO model by comparing its estimates to those produced by an instrumental variable approach. 4) The external validity of experimental charter studies, which we judge by comparing estimates of the relative effectiveness of oversubscribed and undersubscribed charter schools in Florida.

Four main results arise from this analysis. First, we find no difference in estimates of charter effectiveness that come from models that match up to seven public school students to each charter student compared to models that match just one public student to each charter student. Second, CREDO's estimates are only modestly affected by its use of indicators for student participation in the subsidized lunch, limited English proficiency (LEP), and special education programs, even though students who qualify for these programs are more likely to participate if they are attending traditional public schools than if they are attending charter schools. Third, we find that an instrumental variables (IV) approach produces estimates that are similar in both magnitude and direction to those produced by observational models when used to estimate effects for the same sample of eighth-grade charter school students. Finally, when applying a modified CREDO model to estimate separate effects for different types of charter schools, we find that oversubscribed charter schools significantly outperform undersubscribed charter schools.

The rest of this paper proceeds as follows. Section two reviews the prior literature on within-study comparisons that assess the validity of observational methods for judging charter effectiveness, as well as providing a detailed description of the CREDO matching methodology. Section three describes our data. Section four outlines the three specific methodological approaches employed in our critique of the CREDO model. Section five presents the findings, and section six discusses the results.

2. Prior Literature

2.1 Within-Study Comparisons of Methodological Approaches

An experimental approach is considered the gold standard methodology for evaluating charter schools that hold admission lotteries because the treatment and control groups are generated by chance. However, lottery-based studies can only occur when the number of applicants exceeds the number of available seats. As a result, it is unsurprising that the vast majority of experimental studies have been conducted in population-dense urban centers such as Chicago, IL (Hoxby & Rockoff, 2004), New York City (Dobbie & Fryer, 2011; Hoxby, Murarka, & Kang 2009), and Boston, MA (Abdulkadiroğlu et al., 2009). For researchers wishing to study a more representative sample of charter schools, including those in more sparsely populated regions, a non-experimental design may be preferable (e.g., Chingos & West, 2015).

There is some evidence that well-implemented observational methods can produce unbiased estimates of charter effects, even in the absence of random assignment. Indeed, in a literature review of charter school research, Betts and Tang (2011, p. 51) explain that “the choice of method may not be as important as generally believed, as long as value-added methods are being used.” Within-study comparisons by Angrist et al. (2011), Abdulkadiroğlu et al. (2009), and Fortson, Verbitsky-Savitz, Kopa, and Gleason (2012) compare impact estimates generated from experimental data to those from alternative, non-experimental methods to judge how close the estimates produced by an observational approach come to replicating the unbiased experimental estimates. All three studies report promising findings supporting the validity of observational methods for estimating charter effectiveness.

Abdulkadiroğlu et al. (2009) compare lottery-based and observational estimates of the effectiveness of Boston’s charter, pilot, and traditional public schools. Although they find similarly large and positive effects associated with attending a charter school using both strategies, their observational approach is weakened by its failure to restrict the comparison data to “feeder schools,” that is, the traditional public schools and grades previously attended by the charter school students. This restriction is generally recommended in the standards for within-study comparisons outlined by Cook, Shadish and Wong (2008) and Shadish, Clark, and Steiner (2008) as a way to ensure that treatment and comparison students come from similar neighborhoods and communities so that any observed differences between the experimental and observational results are not driven by differences across students' geographic localities.

Angrist et al. (2013) also use both lottery and observational approaches to estimate charter school effects in Massachusetts. By matching on baseline school, year, sex, and race, the authors produce highly comparable effect sizes that validate the use of an observational research design in the Massachusetts charter context.[2]

Matching methods in particular may provide a promising non-experimental alternative for judging charter effects in the absence of random assignment and Fortson et al. (2012) is an extremely well designed study to test the validity of matching methods for this purpose. The dataset is rich and geographically diverse, incorporating 15 charter schools from a variety of locales in six different states. The methods are clearly explained and well-documented, with the authors comparing test score impacts from an experimental approach to ordinary least squares (OLS) regression, exact matching, propensity score matching, and fixed effects methods. By documenting which non-experimental method comes closest to replicating the experimental results, the findings are informative for researchers designing charter impact evaluations in contexts in which an RCT is infeasible. Unfortunately, the population of charter schools studied by Fortson et al. (2012) is not representative of all charters in any of the six states, with the primary limitation being that the charters studied only serve the middle grades. Further, the charters included in the dataset are all oversubscribed and had to agree to participate in the study, which introduces an element of self-selection bias that is verified by the summary statistics

[2] Appendix Table A5 of Angrist et al. (2013) presents the comparable lottery and observational estimates of charter effectiveness, broken out by urban/rural locale, school level, and subject.

presented in Gleason, Clark, Tuttle, and Dwoyer (2010). Charters in this study serve more advantaged students than other charter middle schools nationally. Specifically, they serve fewer students who are eligible for free and reduced price lunch, fewer minorities, and fewer students who scored below their state's math proficiency cut-off at the time of application.

The matching approach tested by Fortson et al. (2012) is also not an exact replication of the CREDO model, as it differs in two important ways. First, the Fortson et al. sample is composed of students who attended a TPS in the baseline year of that study, whereas the CREDO approach allows the researchers to include charter students who have never attended a TPS. Second, the Fortson et al. (2012) study matches on baseline tests in both math and reading, whereas CREDO matches on just baseline math scores for the math analysis and vice versa.

One other important note on this study is that while Fortson et al. (2012) do run a few specification checks to test the sensitivity of the CREDO model, that is not the primary focus of their study, so they do not present exhaustive sensitivity analyses of the CREDO model in particular. For instance, they test if the impact estimates vary depending on whether a 0.05 or 0.10 bandwidth for the test score match is used, but they don't show if results change based on the number of comparison students matched to each treatment student or what specific variables are included in the matching process, such as indicators for student participation in the National School Lunch Program, even though there is reason to believe that students are inconsistently labeled depending on which school sector they attend (Wolf, Witte & Fleming, 2012).

That leaves only one within-study comparison that exactly imitates the CREDO model and compares its results to those of an alternative observational design in an attempt to assess its validity. That study was conducted by two CREDO researchers, Davis and Raymond (2012), who demonstrate that a student fixed effects approach produces estimates of charter school effectiveness that are highly similar to the CREDO estimates.

Within-study comparisons provide an important replication tool to validate non-experimental approaches to estimating charter effects when a lottery is not possible. To date, however, there have been no independent, longitudinal, within-study comparisons of the CREDO model that incorporate all charter schools in a state or all grades in which students are tested annually. Further, there have been an insufficient number of studies that match charter students with public school counterparts that attend public schools in similar geographic areas. Given that the largest and most widely publicized observational studies of charter school impacts rely on the CREDO model (CREDO, 2009a, 2013a), a study that tests the internal validity of this particular approach would make an important contribution.

2.2 The CREDO Model

The observational method assessed in this study is the one used by CREDO. Its estimation strategy starts by identifying each charter school's "feeder schools," which is the set of all traditional public schools previously attended by any student currently enrolled at that charter school. For all charter students, "virtual twins" are then selected from these feeder schools. "Virtual twins" are traditional public school students who exactly match the charter student with respect to grade level, year, gender, race/ethnicity, free or reduced-price lunch eligibility, limited English proficiency (LEP) status, and special education status. The matched public school students must also perform within 0.10 standard deviations of the charter student's performance on standardized exams in his/her first year at a charter. "Virtual twins" are identified separately for reading and math.[3] The CREDO algorithm next generates a "Virtual Control Record" (VCR) for each charter student, which is an amalgamation of the test score performance of up to seven “virtual twins.” CREDO then estimates charter effectiveness by comparing growth scores of charter students and their VCRs, separately for math and reading.

Various aspects of the CREDO approach have been criticized. For instance, Hoxby (2009) highlights the unequal distribution of measurement error in test scores across the treatment and control groups resulting from this amalgamation of up to seven "virtual twins" in the generation of the each charter student's VCR. By averaging a group of traditional public school students to create a single comparison record against which to judge a charter school student’s performance, this approach results in much smaller measurement error for the controls than for the charter school students, which Hoxby claims results in a substantial negative bias.

Of special interest is CREDO’s reliance on program variables that are influenced by subjective administrative decisions. While the omission of predictive variables such as the limited English proficiency indicator could potentially introduce omitted-variables bias, the inclusion of these particular variables is potentially more problematic than their exclusion. That is because there is strong reason to believe that program participation varies depending on sector-specific incentives for student classification (Wolf, Witte, & Fleming, 2012). Estimations of charter effects will be biased if charter schools and traditional public schools systematically differ in the way they classify their students for the LEP, special education, and free or reduced-price lunch programs (Hoxby, 2009). The CREDO matching methodology may be especially vulnerable to this bias as it categorizes all special education students as one homogeneous group, regardless of their specific disability. LEP status is similarly treated as a uniform category,

[3] It is possible to impose a further restriction which eliminates any feeder school students who are currently attending a traditional public school but go on to attend a charter school in later years. We run all analyses both ways (with and without imposing this restriction) and find that this additional restriction makes little difference.

despite significant differences in students' English language acquisition. This will introduce a negative bias if the populations of charter students who are classified as special education or LEP are more likely to have profound disabilities or serious language deficiencies. Responding to this criticism, CREDO (2009b) contends that variation in program variables "must be systematic across the sample" in order to affect results (p. 6). In Appendix B, we provide an explanation for why these program variables may indeed vary systematically between charter and traditional public schools.

Many scholars have noted that an observational model cannot account for impacts of unobserved characteristics. As in all research that attempts to evaluate charter impacts at scale, CREDO is unable to exploit randomization procedures to generate its treatment and control groups. CREDO's key identifying assumption is that the set of factors on which CREDO matches accounts for all potential differences between the treatment and control groups so that there are no remaining differences that are correlated with test scores except for, of course, attending a charter school. But if two students who are identical based on observable characteristics differ in unobservable dimensions that are correlated with the outcome, matching may either overestimate or underestimate the true charter effect, depending on the type of selection bias at play. For example, if students who are highly intrinsically motivated are more likely to enroll in charters, then this selection bias will lead to an overestimation of charter effectiveness. It is also possible that the bias goes in the other direction. For instance, the parents of under-performing students may be more likely to seek an alternative to traditional public schooling, resulting in an underestimation of charter effectiveness. For any matching strategy to be valid, therefore, we must assume that the researchers have matched on a sufficient number of other characteristics that are correlated with important unobserved characteristics of students so that there is no difference left between the two groups.

Finally, it could be the case that the CREDO method is accurate at estimating charter effects only when the true effects are close to zero, but biased otherwise. The only way to test this hypothesis would be to compare impact estimates for a subset of particularly high- or low-performing charters generated by models relying on the CREDO approach to those from an experimental or rigorous quasi-experimental analysis. If the observed effects from both models are similar in magnitude and direction, we can be confident that the CREDO method is capable of consistently producing unbiased estimates.

3. Data

To assess the CREDO model, this paper draws upon three data sources. First, we rely on an extensive administrative dataset provided by the Florida Department of Education that includes student achievement scores on the math and reading portions of the Florida Comprehensive Assessment Test (FCAT) for all students in third through tenth grade in the Florida public school system between 2001 and 2009. FCAT test scores are converted to z-scores by standardizing within subject, grade, and year. This administrative dataset also includes information on each student's demographic characteristics, including gender, race, and age, as well as information on the number of days' attended each school year and student participation in the free/ reduced lunch, LEP, and special education programs. Second, longitude and latitude coordinates for each school come from the National Center for Education Statistics' Elementary and Secondary Information System. Finally, the Florida Charter School Alliance provided us with data on whether or not each charter school self-reported as being oversubscribed for the 2005-06 to 2008-09 school years. We incorporate this information by adding an indicator variable that equals one in years that a charter school is oversubscribed and is otherwise equal to zero.

These data allow for an estimation of the sensitivity of the CREDO methodology because they overlap with data used by CREDO to estimate charter school effectiveness in Florida: CREDO (2009a), which spans the period from 2001 to 2008, and Davis and Raymond (2012), which studies the period from 2005 through 2008.

4. Methodology

4.1 Unequal Measurement Error in VCRs

A unique feature of the CREDO methodology is its use of a composite Virtual Control Record as a synthetic comparison record against which charter achievement is judged instead of a one-to-one, charter-to-TPS student match. Because the VCR is actually an amalgamation of up to seven public school students' test scores, it is measured with more precision than that of the single charter student it is being compared to, which could potentially lead to inconsistency in measurement error across sectors that could bias estimated charter effects. The one-to-seven matching method has been defended on the grounds that it is a way to reduce study attrition, but there is a trade-off being made between reducing attrition and possibly introducing bias. We evaluate whether the VCR method introduces a significant bias by using just one control student for each treated student.

4.2 Observational Studies' Use of Inconsistent Program Variables

We examine the sensitivity of CREDO's estimates to the inclusion of potentially troublesome program variables by first replicating CREDO's matching process as outlined in CREDO (2013c) and applying their model to generate estimates of charter effectiveness, then repeating the process with minor modifications that exclude the program variables. By comparing how much of the variation in the dependent variable is accounted for by the inclusion of these variables, it is possible to estimate the potential tradeoff in bias achieved by including or excluding them.

We match students separately in math and reading, using the same matching methodology reported by CREDO. That is, we match exactly on grade-level, year, gender, race/ethnicity, free or reduced-price lunch eligibility, LEP status, special education status, and to students performing within 0.10 standard deviations on the given FCAT exam.[4] The reader should note that each charter student might have a different matched traditional public school student for the two subjects, math and reading.

The dependent variable is the change in test scores (A_i) from the year when the student first enters a charter school (*t-1*), to the following year (*t*):

$$(1)\ \Delta A_{i,t} = A_{i,t} - A_{i,t-1}$$

We then estimate the following equation:

$$(2)\ \Delta A_{i,t} = \beta_1 A_{i,t-1} + \beta_2 C_{i,t} + \beta_3 X_{i,t} + \varepsilon_{i,t}$$

Here, A_{it} is student *i*'s standardized score on the FCAT math or reading exam for period *t*. We control for baseline test score, $A_{i,t-1}$, because the rate at which scores change may be a function of their initial ranking in the baseline distribution. $C_{i,t}$ is an indicator for student *i* attending a charter school in year *t*, and the coefficient of interest, β_2, represents the estimated treatment effect of attending a charter school on test growth. $X_{i,t}$ is a set of control variables for student characteristics and time period. Finally, $\varepsilon_{i,t}$ is a stochastic error term. We match each student in

[4] We found it difficult to replicate CREDO's exact matching procedures because they were inconsistently described in the technical manuals accompanying their reports. We recommend that future technical reports explicitly state the cases in which a student's observational characteristics (eg. FRL-eligible) are assigned from the charter school records in *t* or from the TPS records in *t-1*.

the first year he or she is observed attending a charter school, *t-1*, and then estimate one-year charter effects.

We also employ an alternative matching model that modifies CREDO's approach by removing the three potentially troublesome program variables from both the matching process and from the regression controls. In this case, matches are based solely on baseline test score, race, gender and year.[5]

4.3 Internal Validity of the CREDO Model

We employ an instrumental variable approach to test the internal validity of CREDO's results. This has two advantages. First, it provides us with a set of arguably exogenous estimates against which to compare those generated by CREDO's observational approach for estimating charter school effectiveness. Second, by focusing on a high-performing subsample of schools, the IV analysis also allows us to test if the CREDO approach only produces accurate estimates when the true effects are close to zero or if the CREDO approach can be relied upon to produce accurate estimates when the true effects are non-zero. While we don't expect to observe identical effects using these two methods, as the IV estimates are local to compliers—that is, students who were induced to attend a charter because of the instruments—our confidence in the internal validity of the CREDO approach will increase if the IV estimates are close to the CREDO estimates.

Focusing on students who attended a charter school in eighth grade, we take advantage of variation in the geographic location of charter high schools to construct five potential

[5] In addition to the detailed specification checks described here, we also run a variety of related tests whose results are not presented here because of space limitations. These results are available from the authors by request. We find that the CREDO estimates of charter effectiveness aren't substantively impacted when we restrict the sample to transfer students only so that we can use program participation variables (e.g. FRL) that were assigned before a student left the traditional public school; or when we establish a control group that consists of public-to-public school transfer students to indirectly control for the negative effect of simply switching schools, an effect that all public-to-charter school transfer students are potentially exposed to.

instruments for charter school attendance. This approach follows that of Booker et al. (2011, 2014), who use a two-stage least squares framework to estimate charter attainment effects for those students who attended a charter school in eighth grade. The first stage of our model is the following:

(3) $C^*_{i,t} = \alpha_i X_{i,t-1} + \mu_{i,t}$

where $C^*_{i,t}$ is a latent variable indicating ninth grade charter school attendance for student *i* in year *t*, and X_i is a vector of exogenous variables based on eight-grade charter location. We use the same instruments as Booker et al. (2011, 2014) to predict high school charter attendance, as geographic proximity has been previously shown to predict students' choice of high school (Altonji, Elder, & Taber, 2005; Grogger & Neal, 2000; Neal, 1997). These variables include distance (in miles) to nearest traditional public school, distance to nearest other charter school, whether or not the eighth-grade charter school offers ninth grade, as well as the number of other charter schools and number of private schools within a five-mile radius, to predict whether an eighth-grade charter student will attend a charter school in ninth grade.

We calculate the first two instruments using longitude and latitude coordinates for each school provided by the National Center for Education Statistics' Elementary and Secondary Information System. The third instrument is an indicator variable for whether an eighth grader's charter school also offers ninth grade. The intuition here is that a student will be more likely to attend a charter school in ninth grade if they already attend this school. We identify the school structure separately for each year by evaluating whether any ninth-grade students participate in FCAT exams at the respective charter school the next year. The final two instruments are the number of charter and private schools offering ninth grade the subsequent year that are located within a five mile radius of a student's eighth-grade charter school. We expect the number of charters to be positively related to charter attendance and the number of private schools to be negatively related to charter attendance. Booker et al. (2011, p.391) reason that charter schools entering the market in areas with many private schools will display attributes that more closely resemble traditional public schools than if these private schools never existed. This is because when new charters are choosing which attributes will be most marketable for them, pre-existing private schools in that market will have already distinguished themselves by offering a significantly different set of characteristics than traditional public schools. Thus, students may be less likely to enroll in a charter school when many proximate private schools exist because the charter school will be more similar to the traditional public school than it otherwise would be. The negative relationship between charter school attendance and the number of private schools within five miles indicates that this intuition is supported by the data.[6]

Incorporating the predicted charter attendance estimate from (3), we estimate charter effects using the following second stage model:

$$(4)\ A_{i,t} = \gamma_1 A_{i,t-1} + \gamma_2 C_{i,t} + \gamma_3 X_{i,t} + \zeta_{i,t}$$

where A_{it} is student *i*'s score on the FCAT math or reading exam for period *t*. The coefficient of interest, γ_2, represents the estimated treatment effect of attending a charter school on test growth. $X_{i,t}$ is a set of control variables for student characteristics and time period, and ζ is the error term.

4.4 External Validity of Lottery Studies

We use a revised CREDO model to estimate effects for undersubscribed and oversubscribed charters by incorporating indicator variables for the different charter school types. This allows us to generate separate estimates for oversubscribed schools, which will offer some intuition about the generalizability of estimates from lottery studies. This question has

---

[6] We do not include the number of nearby traditional public schools, as Booker et al. (2011, p. 392) reason that students typically have to attend their neighborhood school, so the number of nearby traditional public schools is not relevant.

important policy implications because if undersubscribed schools are systematically worse than oversubscribed schools, this finding would be an argument to actually prefer the matching strategy to randomized trials that focus entirely on oversubscribed schools when evaluating the charter sector overall.

5. Findings

5.1 Unequal Measurement Error in VCRs

Consistent with the defense presented in CREDO (2009b), we find no evidence of bias introduced by the use of up to seven virtual twins in the generation of the amalgamated VCR. Using a one-to-one matching approach in place of CREDO's one-to-seven approach, the estimated charter effects are within 0.003 standard deviations and have nearly identical standard errors (within 0.001) to those estimates presented in CREDO (2009a) and Davis & Raymond (2012), as shown in Table 1. This suggests that this particular feature of the CREDO estimation approach does not seriously affect the estimation of charter school effects in this context. Thus, the rest of our estimations are conducted with one-to-one matches.[7]

<< TABLE ONE ABOUT HERE >>

5.2 Observational Studies' Use of Inconsistent Program Variables

Students are classified differentially for participation in government programs by charter and traditional public schools. The probability that a given student is classified as special education, free-lunch eligible or as LEP increases by 0.6 percent, 3.8 percent and 2.5 percent, respectively, while the student is enrolled at a traditional public school than when the same student is enrolled at a charter school (Table 2). Considering the average charter school rates, this means that a student is 4%, 8% and 26% more likely to be classified as special education,

[7] While we find little difference in the results produced by these two models, we wish to note that we are working with an extremely large dataset. The differences between these two model specifications might be more serious in an analysis limited to a smaller area.

free-lunch eligible and LEP, respectively, when enrolled in a traditional public school rather than at a charter school. Explanations for why this differential labeling across sectors might be occurring are provided in Appendix B.

<< TABLE TWO ABOUT HERE >>

Given these differences in program participation by sector, CREDO's matching and estimation procedures, which assume that these practices are consistent across sectors, is potentially problematic. To test how this affects CREDO's estimations of charter school effectiveness, we replicate CREDO's matching methodology and compare it to an otherwise-identical methodology that avoids these program variables. We show that the inclusion of baseline test scores is sufficient to capture most of the variation in student test scores that would have been explained by these program participation variables, thus removing concerns about the exclusion of these potentially troublesome variables.

Table 3 presents summary statistics for the treatment and control groups generated in our matched sample. Because demographic characteristics are incorporated into the matching process, the treatment and control groups have identical values for these variables. Students are matched exactly on all variables except test score, so this sample is not fully representative of all charter school students. Charter students with unusual characteristics or prior test scores are less likely to find a match. As a result, minority students, especially Asian, multiracial, and Native American students, are less likely than others to find a match.

<< TABLE THREE ABOUT HERE >>

Columns 1, 2, 5, and 6 in Table 4 report math and reading results that use CREDO's matching algorithm, while columns 3, 4, 7, and 8 report results using an alternative matching strategy that does not include the program variables. As shown in Column (1), a statistically significant negative one-year charter treatment effect of -0.016 standard deviations of students' normalized test scores is estimated for reading. Removing the program variables from the analysis (as shown in column 4) increases estimated charter effects to -0.005 standard deviations. However, this estimate is still negative and statistically significant. Likewise, removing the variables from the matching process for math increases estimated charter effects from -0.022 to -0.010 standard deviations. The increase of 0.011 and 0.012 standard deviations in reading and math, respectively, is noteworthy. To further investigate whether these changes in the estimated charter effects result from (1) removing the three program variables from the matching process, or (2) removing them as control variables in the regression models, we also present results from regressions in which we match on program variables but do not use the three program variables as controls in the regression and vice versa. As is evident in columns (2) and (6), results differ by the widest margin from the original CREDO estimates when we drop the program variables from the matching.

<< TABLE FOUR ABOUT HERE >>

5.3 Internal Validity of the CREDO Model

We start by presenting evidence for the reliability and validity of the instrumental variable approach. In Table 5, we present OLS estimates of charter school attendance in ninth grade, modeled as a function of the five instruments discussed above. Results are generally as expected. The further a students' eighth-grade charter school is from a traditional public school offering ninth grade, the more likely they are to attend a charter school the next year. Similarly, the further away a student's school is from a ninth-grade charter school, the less likely a student is to attend a charter school in ninth grade. Also as expected, students are more likely to enroll in a charter school for ninth grade if their eighth-grade charter school also offers ninth grade. The more charter schools are located within five miles, the more likely a student is to enroll in a charter school. Finally, as expected, there is a negative relationship between the number of private schools within five miles and charter attendance.

<< TABLE FIVE ABOUT HERE >>

In Table 6, we present F-tests of exclusion restrictions, which are commonly used in the empirical literature to test for instrument validity (e.g. Booker et al., 2011). Among the variables that were demonstrated in Table 4 to be significant determinants of attending a charter high school, the number of charter schools and private schools within five miles and whether the eighth-grade charter school offers ninth grade can be excluded from the test score model, implying that these variables meet the conditions required to serve as instruments for charter school attendance.

<< TABLE SIX ABOUT HERE >>

We present estimated charter effects using these three instrumental variables in Table 7. Attending a charter high school is associated with a test score increase of 0.019 and 0.049 standard deviations, respectively, for math and reading. Although estimates for both subjects are positive, they are insignificant when using standard errors clustered at the school level. One explanation for the reduction in power is that the sample of students is not as large as in the overall matching estimates previously presented. We also present CREDO matching estimates for the sample of eight grade students for comparison, with and without the potentially endogenous program variables. Estimated charter effects from these two matching models are similar in direction and magnitude to matching estimates from the IV models.

<< TABLE SEVEN ABOUT HERE >>

5.4 External Validity of Lottery Studies

To determine whether experimental and observational studies have different results because they rely on data from different types of charters, we estimate the charter effects for oversubscribed and undersubscribed schools separately.

In Table 8, we add a control for whether a student is attending an oversubscribed or undersubscribed charter school. Attending an undersubscribed charter has a significant negative effect of -0.058 standard deviations in reading and -0.040 standard deviations in math when program variables are excluded. In contrast, oversubscribed charter schools have positive estimated charter effects of 0.026 standard deviations in reading and 0.016 standard deviations in math. Notably, the difference between charter estimates for undersubscribed and oversubscribed schools is 0.075 and 0.050 standard deviations for reading and math, respectively, while the difference is only 0.011 and 0.012 standard deviations, respectively, for the difference in estimated charter effects with and without program variables.

<< TABLE EIGHT ABOUT HERE >>

6. Discussion

Four major findings emerge from this analysis. First, the use of up to seven students in CREDO's "virtual control record" is not a major issue, though we prefer to use only one "twin." Second, we show that the same student is significantly less likely to be classified as free or reduced-price lunch, special education or LEP when they are enrolled in a charter school, relative to when they are enrolled at a public school. When these program indicators are excluded from the matching process, the estimated one-year charter effect increases 0.011 and 0.012 standard deviations in reading and math, respectively. We recommend that researchers relying on these endogenous variables conduct sensitivity tests to confirm that their results aren't biased by their inclusion. This may be particularly important in states or regions where sector differences in program participation classification are especially large.

Third, using a quasi-experimental approach that exploits variation in the location of eighth-grade charter schools, we use an instrumental variables approach to estimate charter effects for both math and reading. In order to maximize comparability with the CREDO estimates, we limit the alternative analyses to identical samples of students. Whilst acknowledging the caveat that the IV analysis is limited to estimating the effect of attending a charter school in ninth grade on students who attended a charter school in eight grade, it is interesting to note that the IV results are similar to the CREDO estimates, coming within 0.003 standard deviations in reading and 0.029 standard deviations of CREDO estimates in math. These results imply that CREDO's methodology did not yield biased estimates for this sample.

Finally, we compare the performance of oversubscribed and undersubscribed charter schools to test the external validity of experimental studies. We find that oversubscribed schools significantly outperform undersubscribed schools. While this finding is certainly not surprising, as oversubscribed schools may elicit more parental interest precisely because they are better schools, it does prompt questions about the external validity of experimental studies. Such studies only examine effects for attending oversubscribed charter schools, which are not representative of all charter schools. As a result, they may generate misleading estimates of the average charter school's effectiveness. From a policy perspective, this finding points to the value of relying on matching models like the CREDO approach to evaluate the overall performance of the charter sector.

The present study has two important limitations—the first relates to external validity and the second concerns the sample size in the quasi-experimental section. Regarding external validity, we acknowledge that we only estimate charter effects in one state, using student-level data from 2001-2009. However, the intent of this paper is to investigate methodological biases, not to generate substantive estimates of charter school effectiveness. Thus, while it is true that the estimated charter effects from this study are not generalizable to other states and years, the results from this study can offer insight into methodological tradeoffs in charter school evaluations. The second limitation of this paper is that our instrumental variable approach relies on relatively few observations. Further, the parameter estimated by the IV estimator is a local average treatment effect, whereas the parameter estimated by CREDO's matching approach is a type of average treatment-on-treated effect. Thus, while the IV resolves many questions about the internal validity of the charter estimates, the discussion around the internal validity of the CREDO approach is by no means definitively finished. We encourage future research to return to the questions raised here as more and better data become available.

As the charter sector expands across the United States, a significant number of new charter schools do not have waiting lists, precluding the possibility of a randomized evaluation. This study contributes to a growing body of evidence suggesting that the CREDO approach is capable of overcoming the problem of student self-selection into charter schools to produce reliable estimates of charter effectiveness, and does so in a manner that ensures high rates of coverage for many different types of charter schools in diverse locations across the country.

References

Abdulkadiroğlu, A., Angrist, J. D., Cohodes, S., Dynarski, S. M., Fullerton, J., Kane, T. J. & Pathak, P. A. (2009). *Informing the debate: Comparing Boston's charter, pilot and traditional schools*. Boston MA: Boston Foundation Report.

Abdulkadiroğlu, A., Angrist, J. D., Hull, P. D., & Pathak, P. A. (2014). Charters without lotteries: Testing takeovers in New Orleans and Boston. *School Effectiveness and Inequality Initiative Discussion Paper #2014.03*. Retrieved from http://seii.mit.edu/wp-content/uploads/2014/12/SEII-Discussion-Paper-2014.03-Abdulkadiro%C4%9Flu-Angrist-Hull-Pathak1.pdf.

Altonji, J. G., Elder, T. E. & Taber, C. R. (2005). Selection on observed and unobserved variables. Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113 (1): 151-184.

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A. & Walters, C. R. (2010). Who benefits from KIPP? (Working Paper No. 15740). Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w15740.

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Fullerton, J.B., Kane, T. J., Pathak, P. A. & Walters, C. R. (2011). *Student achievement in Massachusetts' charter schools*. Cambridge MA: Center for Education Policy Research, Harvard University.

Angrist, J. D., Pathak, P. A. & Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics*, 5(4): 1-27.

Bass, D. N. (2010). Fraud in the lunchroom? *Education Next*, 10(1): 67-71.

Batdorff, M., Maloney, L. & May, J. (2010). *Charter school funding: Inequity persists*. Muncie, IN: Ball State University.

Bettinger, E. P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review*, 24(2): 133-147.

Betts, J. R. & Tang, Y. E. (2011). *The effect of charter schools on student achievement: A meta-analysis of the literature.* Seattle, WA: Center on Reinventing Education.

Bifulco, R. & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy*, 1(1): 50-90.

Booker, K., Gilpatric, S. M., Gronberg, T. & Janson, D. (2007). The impact of charter school attendance on charter performance. *Journal of Public Economics*, 91(5): 849-876.

Booker, K., Sass, T. R., Gill, B. & Zimmer, R. (2011). The effects of charter high schools on educational attainment. *Journal of Labor Economics*, 29(2): 377-415.

Booker, K., Sass, T. R., Gill, B. & Zimmer, R. (2014). Charter high schools' effects on long-term attainment and earnings. *Andrew Young School of Policy Studies Research Paper Series*, 14(5).

Center for Research on Education Outcomes (2009a). *Multiple choice: Charter school performance in 16 States*. Stanford, CA: CREDO. Retrieved from http://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf.

Center for Research on Education Outcomes (2009b). *Fact vs. fiction: An analysis of Dr. Hoxby's misrepresentation of CREDO's research*. Stanford, CA: CREDO. Retrieved from http://credo.stanford.edu/reports/CREDO_Hoxby_Rebuttal.pdf.

Center for Research on Education Outcomes (2013a). *National charter school study*. Stanford, CA: CREDO. Retrieved from http://credo.stanford.edu/documents/NCSS%202013%20Final%20Draft.pdf.

Center for Research on Education Outcomes (2013b). *Charter school performance in Florida*. Stanford, CA: CREDO. Retrieved from http://credo.stanford.edu/reports/FL_CHARTER%20SCHOOL%20REPORT_CREDO_2009.pdf.

Center for Research on Education Outcomes (2013c). *National charter school study technical appendix*. Stanford, CA: CREDO. Retrieved from http://credo.stanford.edu/documents/NCSS2013_Technical%20Appendix.pdf.

Center for Research on Education Outcomes (2015). *Urban charter school study 2015*. Stanford, CA: CREDO. Retrieved from http://urbancharters.stanford.edu/index.php.

Chingos, M. M., & West, M. R. (2015). The uneven performance of Arizona's charter schools. *Educational Evaluation and Policy Analysis*, *37*(1), 120–134.

Cook, T. D., Shadish, W. R. & Wong, V. C. (2008). The conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27(4): 724-750.

Davis, D. H. & Raymond, M. E. (2012). Choices for studying choice: Assessing charter school effectiveness using two quasi-experimental methods. *Economics of Education Review*, 31(2): 225–236.

Dobbie, W. & Fryer Jr., R. G. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 3(3): 158-187.

Federal Education Budget Project (2014a). *Individuals with Disabilities Education Act – funding distribution*. Washington D.C.: New America Foundation. Retrieved from http://febp.newamerica.net/background-analysis/individuals-disabilities-education-act-funding-distribution.

Federal Education Budget Project (2014b). *No child left behind funding*. Washington D.C.: New America Foundation. Retrieved from http://febp.newamerica.net/background-analysis/no-child-left-behind-funding.

Florida Department of Education (2015). *Fact Sheet: Florida's Charter Schools*. Tallahassee, FL: Florida Department of Education. Retrieved from http://www.fldoe.org/core/fileparse.php/7778/urlt/Charter_Oct_2015.pdf

Florida Statutes (2000). Funds for operation of schools. Title XVI, Chapter § 236.081

Florida Tax Watch Center for Educational Performance and Accountability (2010). *How charter school funding compares*. Tallahassee, FL: Florida Tax Watch. Retrieved from http://www.floridataxwatch.org/resources/pdf/charterschools20812.pdf.

Fortson, K., Verbitsky-Savitz, N., Kopa, E. & Gleason, P. (2012). *Using an experimental evaluation of charter schools to test whether nonexperimental comparison group methods can replicate experimental impact estimates*. NCEE Technical Methods Report No. 2012-

4019. Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gleason, P., Clark, M., Tuttle, C. C. & Dwoyer, E. (2010). *The evaluation of charter school impacts: Final report*. NCEE 2010-4029. Washington D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Greene, J. P. & Forster, G. (2002). *Effects of funding incentives on special education enrollment* (Civic Report Number 32). New York City, NY: The Manhattan Institute.

Grogger, J. & Neal, D. (2000). Further evidence on the effects of Catholic secondary schooling. *Brookings-Wharton Papers on Urban Affairs*, 151-201.

Hanushek, E. A., Kain, J. F., Rivkin, S. G. & Branch, G. F. (2007). Charter school quality and parental decision marking with school choice. *Journal of Public Economics*, 91(5): 823-848.

Harwell, M. & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher*, 39(2): 120-131.

Hastings, J. S., Neilson, C. A. & Zimmerman, S. D. (2012). The effect of school choice on intrinsic motivation and academic outcomes (Working Paper No. 18324). Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w18324.

Hosp, J. & Reschly, D. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Council for Exceptional Children*, 70(2): 185-199.

Hoxby, C. M. (2009). *A statistical mistake in the CREDO study of charter schools*. Unpublished paper. Stanford, CA: Stanford University. Retrieved from http://users.nber.org/~schools/charterschoolseval/memo_on_the_credo_study.pdf.

Hoxby, C. M., Murarka, S. & Kang, J. (2009*). How New York City's charter schools affect student achievement: August 2009 report*. Cambridge, MA: New York City Charter Schools Evaluation Project.

Hoxby, C., M. & Rockoff, J., E. (2004). *The impact of charter schools on student achievement*. Unpublished Manuscript. Retrieved from http://www.rand.org/content/dam/rand/www/external/labor/seminars/adp/pdfs/2005hoxby.pdf.

Imberman, S. (2011). Achievement and Behavior in Charter Schools: Drawing a More Complete Picture. *The Review of Economics and Statistics*, *93*(2), 416–435.

Levin, M. & Nueberger, Z. (2013). *Community eligibility: Making high-poverty schools hunger free*. Washington D.C.: Food Research and Action Center and Center on Budget and Policy Priorities. Retrieved from http://frac.org/pdf/community_eligibility_report_2013.pdf.

Mathematica Policy Research (2012). *Review of the CREDO charter school studies*. Washington D.C.: What Works Clearinghouse Quick Review, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/singlestudyreview.aspx?sid=220 - .

Mills, J. N. (2013). The achievement impacts of Arkansas open-enrollment charter schools. *Journal of Education Finance*, *38*(4), 320–342.

National Alliance for Public Charter Schools (2015). *Data Dashboard*. Washington DC: NAPCS. Retrieved from http://dashboard.publiccharters.org/dashboard/home.

Neal, D. (1997). The effects of Catholic secondary schooling on educational achievement, *Journal of Labor Economics*, 15(1): 98-123.

New America Foundation (2015). *Federal education budget project*. Washington D.C.: New America Foundation. Retrieved from http://febp.newamerica.net/k12/FL.

Parrish, T. (2002). Racial disparities in the identification, funding, and provision of special education. In D. J. Losen and G. Orfield (Eds.), Racial inequity in special education. Cambridge, MA: Harvard Education Press, 15-38.

Peterson, P. E. & Llaudet, E. (2006). On the public-private school achievement debate (Working Paper No. PEPG 06-02). Retrieved from the Program on Education Policy and Governance, Harvard University website: http://www.ksg.harvard.edu/pepg/PDF/Papers/PEPG06-02-PetersonLlaudet.pdf.

Sass, T. R. (2006). Charter schools and student achievement in Florida. *Education Finance and Policy*, 1(1): 91-122.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334–1356.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children,* 77(3): 317-334

Tuttle, C. C., Gill, B., Gleason, P., Knechtel, V., Nicols-Barrer, I. & Resch, A. (2013). *KIPP middle schools: Impacts on achievement and other outcomes*. Washington DC: Mathematica Policy Research, Inc.

United States Department of Agriculture (2010). *National school lunch program fact sheet*. Retrieved from http://www.fns.usda.gov.cnd/Lunch/AboutLunch/NSLPFactSheet.pdf.

Winters, M. A. (2013). *Why the gap? Special education and New York City charter schools*. New York City: Manhattan Institute for Public Research.

Winters, M. A. (2015). Understanding the Gap in Special Education Enrollments Between Charter and Traditional Public Schools Evidence From Denver, Colorado. Forthcoming in *Educational Researcher*.

Witte, J. F., Wolf, P. J., Carlson, D. & Dean, A. (2012). *Milwaukee independent charter schools study: Final report on four-year achievement gains*. Fayetteville, AR: Department of Education Reform, University of Arkansas.

Wolf, P. J., Witte, J., F. & Fleming, D. J. (2012). Special choices. *Education Next*, *12*(3), 16–22.

Wood, R.C., Chambers, M.L., Mendonca, S. & Birkett, K.F. (n.d.) *Florida.* Washington D.C.: U.S. Department of Education, National Center for Education Statistics, Education Finance Statistics Center. Retrieved from https://nces.ed.gov/edfin/pdf/StFinance/Florida.pdf.

Zimmer, R. & Buddin R. (2006). Charter school performance in two large urban districts. *Journal of Urban Economics*, 60(2): 307-326.

Zimmer, R., Gill, B., Booker, K., Lavertu, S. & Witte, J. (2012). Examining charter student achievement effects across seven states. *Economics of Education Review*, 31(2): 213-224.

Table 1. Comparing CREDO Estimates of Charter Effectiveness under One-to-One and One-to-Seven Virtual Twin Matches

	Davis and Raymond 2012				CREDO 2009			
	Reading	CREDO Reading	Math	CREDO Math	Reading	CREDO Reading	Math	CREDO Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Charter	-0.018**	-0.021**	-0.019**	-0.021**	-0.018**	-0.02**	-0.026**	-0.03**
	[0.003]	[0.002]	[0.002]	[0.002]	[0.002]		[0.002]	
Lagged read score	-0.241**	--			-0.243**	--		
	[0.002]				[0.001]			
Lagged math score			-0.201**	--			-0.208**	--
			[0.002]				[0.001]	
Special education	-0.021**	--	-0.031**	--	-0.049**	--	-0.054**	--
	[0.004]		[0.004]		[0.003]		[0.003]	
Free lunch	-0.098**	--	-0.079**	--	-0.101**	--	-0.083**	--
	[0.003]		[0.003]		[0.002]		[0.002]	
LEP	-0.068**	--	-0.015**	--	-0.037**	--	0.003	--
	[0.006]		[0.005]		[0.004]		[0.004]	
Constant	0.662**	--	0.556**	--	0.621**	--	0.546**	--
	[0.021]		[0.019]		[0.014]		[0.013]	
Matching variables:								
Special education	X	X	X	X	X	X	X	X
Free-lunch eligible	X	X	X	X	X	X	X	X
LEP	X	X	X	X	X	X	X	X
Maximum "virtual twins"	1	7	1	7	1	No limit	1	No limit
Years used:	2005-2008				2001-2008			
Charter students	58,907	--	58,687	--	96,300	--	96,125	--
R-squared	0.135	--	0.114	--	0.135	--	0.119	--

** p<0.01, * p<0.05, + p<0.1

Notes: The dependent variable is the change in test score from the previous year to the current year. Charter students and matched public students included. All test scores are standardized separately for each grade level and each year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test. All regressions control for prior test score, whether a student repeated a grade, gender, race, grade and year. Standard errors are in parentheses. Columns 2 and 4 come from Davis and Raymond (2012), and Columns 6 and 8 come from CREDO (2009).

Table 2. Individual-level Fixed Effects Estimates of Program Variables in TPS and Charter Schools

Dependent variable:	Special Education	Eligible for Free/Reduced Lunch	Limited English Proficiency
	(1)	(2)	(3)
Mean of dependent variable	0.143	0.483	0.095
TPS attendance	0.006**	0.038**	0.025**
	[0.001]	[0.001]	[0.001]
Constant	0.088**	0.572**	0.187**
	[0.002]	[0.003]	[0.002]
Student FE	X	X	X
Year controls	X	X	X
Grade controls	X	X	X
Observations	561,750	561,750	561,750
Students	103,570	103,570	103,570
R-squared	0.018	0.022	0.088

** p<0.01, * p<0.05, + p<0.1

Notes: Units of observation are the students' program variables in student-years. All students who transfer across the public-charter sector while they are in testable grades between 2001 and 2009 are included. Special education, Free-lunch eligibility and limited English proficiency are used as dependent variables. These three indicator variables equal 1 when a student participates in each respective program, and equal 0 otherwise. The interpretation is that the probability a given student is classified as special education is 0.6% higher while the student is attending a traditional public school than while the same student is at a charter school. Considering the average rate of special education participation, this means that a student is 4% more likely to be classified as special education while in a traditional public school.

Table 3. Summary Statistics for Matched Students

	Reading					Math				
	Charter students (N = 114,729)		TPS students (N = 114,729)		p-Value of Difference	Charter students (N = 114,407)		TPS students (N = 114,407)		p-Value of Difference
Prior Test Scores	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Baseline Math	0.024	[0.906]	0.086	[0.906]	0.00	0.038	[0.864]	0.038	[0.864]	1.00
Baseline Reading	0.052	[0.874]	0.052	[0.875]	1.00	0.055	[0.901]	0.037	[0.905]	0.00
Other Baseline Covariates	Percentage		Percentage			Percentage		Percentage		
Grade (from baseline year)										
3rd	32		32		1.00	31		31		1.00
4th	10		10		1.00	10		10		1.00
5th	23		23		1.00	23		23		1.00
6th	10		10		1.00	10		10		1.00
7th	8		8		1.00	8		8		1.00
8th	12		12		1.00	12		12		1.00
9th	6		6		1.00	6		6		1.00
10th	0		0		1.00	0		0		1.00
Sex										
Female	50.1		50.1		1.00	50.1		50.1		1.00
Male	49.9		49.9		1.00	49.9		49.9		1.00
Race/Ethnicity										
Asian or Pacific Islander	1.2		1.2		1.00	1.2		1.2		1.00
Black, not Hispanic	21.1		21.1		1.00	21.1		21.1		1.00
Hispanic	31.1		31.1		1.00	31.1		31.1		1.00
Multiracial	2.2		2.2		1.00	2.2		2.2		1.00
White, not Hispanic	44.3		44.3		1.00	44.3		44.3		1.00
Programs										
Special education	12		12		1.00	12.0		12.0		1.00
Free-lunch eligible	44.3		44.3		1.00	44.4		44.4		1.00
Limited English proficiency	10.7		10.7		1.00	10.7		10.7		1.00
Emotionally Disabled	0.29		0.51		0.00	0.53		0.27		0.00
Specific Learning-Disabled	3.98		4.18		0.00	4.01		4.31		0.00
Missing Value Indicators										
Baseline Math	0.4		0.3		0.00	0		0		1.00
Baseline Reading	0		0		1.00	0.3		0.3		0.29

Sex	0	0	1.00	0	0	1.00
Race/Ethnicity	0	0	1.00	0	0	1.00

Notes: Samples are charter students and matched public students. Charter students who transfer from a public to a charter school are matched to traditional public school students using program variables and lagged test scores from the traditional public school in the year before students switch to charter schools. Charter students who have not previously attended a public school are matched using program variables and lagged test scores in the first year observed at a charter school in the dataset. All test scores are standardized separately for each grade level and each year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test.

Table 4. Charter School Effectiveness With and Without Program Variables (2001-2009)

	Reading				Math			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Charter school	-0.016**	-0.016**	-0.012**	-0.005**	-0.022**	-0.020**	-0.018**	-0.010**
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Lagged reading score	-0.243**	-0.242**	-0.229**	-0.228**				
	[0.001]	[0.001]	[0.001]	[0.001]				
Lagged math score					-0.210**	-0.210**	-0.200**	-0.199**
					[0.001]	[0.001]	[0.001]	[0.001]
Special education	-0.033**	-0.044**			-0.045**	-0.060**		
	[0.003]	[0.002]			[0.002]	[0.002]		
Free-lunch eligible	-0.099**	-0.098**			-0.080**	-0.082**		
	[0.002]	[0.002]			[0.002]	[0.002]		
Limited English proficiency	-0.039**	-0.041**			-0.004	-0.013**		
	[0.004]	[0.003]			[0.003]	[0.003]		
Constant	0.606**	0.573**	0.564**	0.524**	0.547**	0.512**	0.511**	0.466**
	[0.013]	[0.012]	[0.013]	[0.012]	[0.012]	[0.011]	[0.012]	[0.011]
Matching Variables:								
Special Education	X		X		X		X	
Free Lunch	X		X		X		X	
LEP	X		X		X		X	
Charter students	114,729	118,888	114,729	118,888	114,407	118,751	114,407	118,751
R-squared	0.131	0.138	0.125	0.132	0.118	0.125	0.113	0.119
Match Rate:	94.1%	97.5%	94.1%	97.5%	93.8%	97.4%	93.8%	97.4%

** p<0.01, * p<0.05, + p<0.1

Notes: The dependent variable is the change in test score from the previous year to the current year. Charter students and matched public students are included. Charter students who transfer from a public to a charter school are matched to traditional public school students using program variables and lagged test scores from the traditional public school in the year before students switch to charter schools. Charter students who have not previously attended a traditional public school are matched using program variables and lagged test scores in the first year observed at a charter school in the dataset. All test scores are standardized separately for each grade level and each year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test. All regressions control for race, grade, year, and repeated grades. All students are matched for exactly two consecutive years, and 1-year effects are reported. The match rate reports the percentage of charter students with two consecutive years of test scores that match public students.

Table 5. OLS Estimates of Attending a Charter High School

	(1)	(2)
Distance to nearest traditional public school	0.073	0.074
	[0.059]	[0.059]
Distance to nearest other charter	-0.006	-0.006
	[0.005]	[0.005]
Eighth-grade charter offers ninth-grade	0.396**	0.397**
	[0.077]	[0.078]
Number of other charters	0.085**	0.085**
	[0.016]	[0.017]
Number of private schools	-0.038*	-0.038*
	[0.017]	[0.017]
Eighth-grade math score	0.044*	
	[0.018]	
Eighth-grade reading score		0.033*
		[0.015]
Observations	21,273	21,290
R-squared	0.311	0.308

** p<0.01, * p<0.05, + p<0.1

Notes: The dependent variable is an indicator variable for charter attendance in 9^{th} grade. Units of observation are eighth-grade charter school students. Distances are in miles and number of other charter and private schools are for a five-mile radius from the eight-grade charter school. All regressions control for race, gender and year. Standard errors adjusted for clustering at the school level are in brackets.

Table 6. F-Tests of Exclusion Restrictions for Instrument Validity

Instruments	Reading Score (F-Test)		Math Score (F-Test)	
	F	P-value	F	P-value
All five instruments	1.88	0.1063	1.98	0.0896
	(5, 88)		(5, 88)	
Distance to nearest traditional public school	0.22	0.6404	0.30	0.5851
	(1, 88)		(1, 88)	
Distance to nearest other charter	6.91	0.0101	6.32	0.0137
	(1, 88)		(1, 88)	
Eighth-grade charter offers ninth-grade	0.20	0.6555	0.18	0.6738
	(1, 88)		(1, 88)	
Number of other charters	1.42	0.2366	3.34	0.0710
	(1, 88)		(1, 88)	
Number of private schools	0.22	0.6389	0.38	0.5383
	(1, 88)		(1, 88)	
Minimum distance variables	3.65	0.0301	3.32	0.0407
	(2, 88)		(2, 88)	
No. charters, no. private schools	1.05	0.3546	2.41	0.959
	(2, 88)		(2, 88)	
No. charters, no. private schools, 8 to 9	0.70	0.5542	1.61	0.1938
	(3, 88)		(3, 88)	

Notes: The F-test values are reported for regressing the designated instruments on 9^{th} grade test scores. Degrees of freedom and number of clusters are in parentheses. Distances are in miles and number of other charter and private schools are within a five-mile radius from the eighth-grade charter school. Standard errors are adjusted for clustering at the school level.

Table 7. Instrumental Variable & Matching Estimates of Charter Effectiveness in Ninth Grade

Method:	Matching	Matching	IV	Matching	Matching	IV
Subject:	Reading	Reading	Reading	Math	Math	Math
	(1)	(2)	(3)	(4)	(5)	(6)
Charter attendance	0.022*	0.042**	0.019	0.020**	0.024**	0.049**
	[0.009]	[0.009]	[0.016]	[0.007]	[0.008]	[0.014]
	(0.022)	(0.020)	(0.047)	(0.030)	(0.035)	(0.039)
Special education	0.052**			0.020		
	[0.018]			[0.015]		
Free-lunch eligible	-0.051**			0.005		
	[0.010]			[0.009]		
Limited English proficiency	0.045+			0.055**		
	[0.023]			[0.019]		
Matching Variables:						
Special Education	X			X		
Free-lunch eligible	X			X		
Limited English proficiency	X			X		
Instruments:						
Eighth-grade charter offers ninth-grade			X			X
Number of other charters			X			X
Number of private schools			X			X
Charter students	7,080	7,169	21, 803	7,081	7,177	21,772
R-squared	0.090	0.089	0.610	0.048	0.044	0.687

** p<0.01, * p<0.05, + p<0.1

Notes: Samples for matching estimates are restricted to students from the IV regression for same subject. The dependent variable for regressions 1, 2, 4 and 5 is the change in test score from eighth-grade to ninth-grade. For these regressions, charter students and matched public students included. Charter students who transfer from a public to a charter school are matched to traditional public school students using program variables and lagged test scores from the traditional public school in the year before students switch to charter schools. Charter students who have not previously attended a public school are matched using program variables and lagged test scores in the first year observed at a charter school in the dataset. The dependent variable for regressions 3 and 6 is a student's ninth-grade test score. For these regressions, charter attendance is instrumented for by whether a student's eighth-grade charter school offers ninth grade, as well as the number of other charter and private schools within a five-mile radius. All regressions control for prior test score, race, gender and year, and all report 1-year effects. All test scores are standardized separately for each grade level and each year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test. Standard errors adjusted for clustering at the school level are in parentheses. Robust standard errors are in brackets.

Table 8. Effectiveness of Oversubscribed and Non-Oversubscribed Charter Schools (2005-2009)

	Reading	Reading	Reading	Reading	Math	Math	Math	Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Undersubscribed charter	-0.063**	-0.062**	-0.065**	-0.058**	-0.047**	-0.045**	-0.048**	-0.040**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Oversubscribed charter	0.012**	0.012**	0.020**	0.026**	0.003	0.003	0.011**	0.016**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Lagged reading score	-0.244**	-0.245**	-0.231**	-0.233**				
	[0.001]	[0.001]	[0.001]	[0.001]				
Lagged math score					-0.205**	-0.206**	-0.197**	-0.197**
					[0.001]	[0.001]	[0.001]	[0.001]
Repeated grade	0.128**	0.173**	0.127**	0.173**	0.110**	0.166**	0.109**	0.166**
	[0.012]	[0.012]	[0.012]	[0.012]	[0.011]	[0.012]	[0.011]	[0.012]
Special education	-0.009**	-0.021**			-0.023**	-0.038**		
	[0.003]	[0.003]			[0.003]	[0.003]		
Free-lunch eligible	-0.090**	-0.091**			-0.074**	-0.077**		
	[0.002]	[0.002]			[0.002]	[0.002]		
Limited English Proficiency	-0.054**	-0.052**			-0.018**	-0.026**		
	[0.005]	[0.005]			[0.004]	[0.004]		
Constant	0.568**	0.508**	0.533**	0.466**	0.564**	0.506**	0.530**	0.462**
	[0.018]	[0.017]	[0.018]	[0.017]	[0.017]	[0.016]	[0.017]	[0.016]
Matching Variables:								
Special Education	X		X		X		X	
Free Lunch	X		X		X		X	
LEP	X		X		X		X	
Observations	77,608	77,251	77,608	77,251	80,667	80,460	80,667	80,460
R-squared	0.128	0.139	0.123	0.133	0.113	0.122	0.109	0.117
Match rate:	89.1%	92.6%	89.1%	92.6%	88.7%	92.3%	88.7%	92.3%

** p<0.01, * p<0.05, + p<0.1

Notes: The dependent variable is the change in test score from the previous year to the current year. Charter students and matched public students are included. Charter students who transfer from a public to a charter school are matched to traditional public school students using program variables and lagged test scores from the public sector in the year before students switch to charter schools. Charter students who have not previously attended a traditional public school are matched using program variables and lagged test scores in the first year observed at a charter school in the dataset. All test scores are standardized separately for each grade level and each year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test. Oversubscribed is a binary variable that equals 1 if the charter school has a waiting list, indicating oversubscribed charter schools, and equals 0 otherwise. All regressions control for year, grade and race. All students are matched for exactly two consecutive years, and 1-year effects are reported. Years 2005-2009 are used because these are the years oversubscribed charter school data are available. The match rate reports the percentage of charter students with two consecutive years of test scores that match public students.

Appendix A. Other Potential Concerns with CREDO's Methodology

In the literature review, we present the three most relevant criticisms of CREDO's methodology. Here, we discuss two other minor concerns that may bias its charter estimates.

First, CREDO specifies that matched control group students perform within 0.10 standard deviations above or below the baseline performance of a given charter student. Because students are less densely concentrated the further you get from the center of the test score distribution, this can create unintended biases for matches that occur in the far right and left tails. For example, if a charter student performs at the 95th percentile, his or her normalized test score (z-score) is 1.645. Approximately 0.95 percent of students will perform within the z-score range 0.10 standard deviations higher, while approximately 1.12 percent of students will perform within the z-score range 0.10 standard deviations lower than this student. This means there is a higher probability that this high-performing charter student will be matched with a lower-performing public school student than a higher-performing one. This could lead to an over-estimation of the charter school effect for a high-performing student, as the baseline test scores of the matched control group students are actually slightly lower than those of the high-performing charter school student.

This problem is exacerbated if more extreme test scores are more likely to regress to the mean. Because a very high-performing charter student has a higher probability of being matched to a control group peer whose score is within 0.10 standard deviations below the charter student's score, rather than above it, the charter student has a higher probability of regressing towards the mean. If present, this type of bias would result in over-estimating charter effectiveness for lower-ability students and under-estimating charter effectiveness for higher-ability students.

To this point, it is interesting to note that CREDO estimates positive charter effects for students with lower baseline test scores and negative charter effects for students with higher baseline test scores (CREDO, 2009a). Although it is possible that this finding is influenced by the attenuation bias described above, it is more likely to reflect true heterogeneity in the treatment effect. This is because only a small number of matches could be affected by this potential source of bias, given that less than four percent of all matches are so extreme that test scores are even 0.05 SD apart. CREDO has responded to this criticism by citing an independent analysis from Mathematica Policy Research (Fortson et al., 2012), which finds that restricting the variation in baseline test scores during the matching process does not significantly affect estimated charter effects (CREDO, 2013a). CREDO also demonstrates that the baseline test scores for charter students and their controls are not significantly different for any subgroup for which they present results.

Second, the group of matched charter students CREDO assembles for its analysis may not be fully representative of the universe of charter school students. If certain student subgroups are under-represented in the matched sample, the CREDO results may not be appropriately extrapolated to these groups. Further, if sufficiently large proportions of students are excluded from the analysis, the overall charter effect may be biased by their omission. Indeed, CREDO (2013c) reports that minority charter students and students with lower baseline test scores are underrepresented in the matched charter student sample. However, CREDO consistently matches at least eighty percent of all charter students in their studies, so this particular criticism does not hold much weight. Ultimately, the CREDO studies are far more representative of all charter students than any experimental study on this subject.

Appendix B. Differential Program Participation

There are strong reasons to believe that there are differences in program participation by charter and traditional public sector. In this appendix we outline those reasons. We begin by discussing the role of programs in school funding formulas and for identifying subgroups for state accountability systems. Next, we review reliability problems associated with using student eligibility for these programs as indicator variables. Finally, we discuss how systematic differences in the incentives to classify students as program participants can lead to discrepancies in student classification between charter and public schools.

Appendix B.1 Role of Program Variables in Funding Formulas and Accountability Systems

Student participation in free or reduced-price lunch, special education, and limited English proficiency programs is often used in calculations to determine the level of federal and state aid to local schools. At the federal level, spending on the National School Lunch Program reached $10.8 billion dollars nationally (United States Department of Agriculture, 2010) and $725 million for Florida alone (New America Foundation, 2015) by the 2010 fiscal year. Federal spending through the Individual with Disabilities Education Act, meanwhile, reached $12.5 billion dollars for the 2014-15 school year (Federal Education Budget Project, 2014a). Student participation in the free or reduced-price lunch program also influences each school's eligibility for Title 1 status and its corresponding federal funding, which reached $14.49 billion nationally (FEBP, 2014b) and $675 million for Florida alone (NAF, 2015) in the 2009 fiscal year. Both traditional public and non-profit charter schools are eligible for these funding streams.

To determine the level of state and local funding for its traditional public schools, Florida's sixty-seven public school districts participate in the Florida Education Finance Program (FEFP), a modified foundation aid plan. The FEFP determines public school funding by multiplying the number of full-time equivalent students in a school that have been identified for each educational program by a cost factor to generate weighted full-time equivalent counts (Wood, Chambers, Mendonca & Birkett, n.d.). These weighted student counts are then multiplied by a base per-pupil funding amount and a district cost differential (Florida Statutes, § 236.081). Supplemental allocations may also be added to account for declining enrollment or dropout prevention. Thus, as enrollment in each of the educational programs increases, so does the total value of state and local dollars received by each public school through the FEFP.

Program variables are also often used as an important background factor in school evaluations, such as No Child Left Behind, which stipulates that schools must demonstrate that student subgroups defined by these programs make adequate yearly progress. The fact that these programs operate nationwide makes them attractive control variables to use in state-by-state analyses of charter effectiveness, but there are various reasons why these variables may be mismeasured generally and why participation in these three programs may differ systematically, depending on which school sector a student attends.

Appendix B.2 Reliability Problems with Program Variables

Although there is an official income cutoff that determines student eligibility for subsidized lunch, research has shown that participation in this program is an unreliable indicator of family income due to non-compliance, error, and fraud in the application process (Bass, 2010).[8] To apply for the lunch program, parents simply self-report their incomes, but do not have to provide any type of official income documentation. Additionally, districts are only required to verify household incomes for up to three percent of the program participants. Because schools get additional funding for each student that participates in the lunch program, administrators

[8] Free or reduced-price lunch eligibility is based on household income cutoffs (Bass, 2010). Students with household incomes that are less than 130 percent of the federally designated poverty level qualify for free meals. Students in between 131 and 180 percent of the federal poverty level qualify for reduced-price meals.

have an incentive to potentially over-classify students as eligible for this program. Over the last two decades, national participation in the program has increased by over twenty percent, even though the poverty rate among children has declined by ten percent, suggestive evidence of growing non-compliance with the program rules (Bass, 2010).

There are other reasons to believe that lunch subsidies are not credible identifiers of socioeconomic status. For instance, in many Floridian schools and districts, the Community Eligibility Option now qualifies all school or district students for free lunch if at least 40 percent of school or district students are low income (Levin and Neuberger, 2013). Harwell and LeBeau (2010) perform a review of indicators for socioeconomic status and find free or reduced-price lunch eligibility to be an unreliable indicator. They conclude that it should not be used in studies because “it suffers from important deficiencies that can bias inferences” (p. 120).

There are similar measurement problems with special education and LEP labels, which identify students who are eligible for additional support services based on these unique needs. As with the subsidized lunch program variable, the process for identifying a student for LEP or special education services has an element of subjectivity that can introduce measurement error, particularly if the identification process is incentivized with the offer of additional "bounty" funding for students who receive these labels (Greene & Forster, 2002). One particularly compelling piece of evidence demonstrating the subjective nature of special education identification is the disproportionate representation of certain races/ethnicities among this population (Hosp & Reschly, 2004; Parrish, 2002; Sullivan, 2011).

Appendix B.3 Differential Incentives for Program Classification Across Sectors

The subjective identification of these program variables may bias the CREDO results if there are differential incentives for student identification across sectors (i.e., between traditional public schools and charter schools). This is a risk in Florida, where school funding policies present the two school sectors with different incentives for identifying students for these programs. Throughout the time period covered by this study, Florida did not classify charter schools as "Local Education Agencies," so local districts had the authority to determine how to distribute state and federal funding to charter schools. According to Batdorff, Maloney and May (2010), Florida's local districts provided over 25 percent lower per-student funding to charter schools than to traditional public schools in the 2006-07 school year. Florida Tax Watch (2010) finds that the majority of this differential funding stems from federal sources, which includes the funding associated with student participation in special education, LEP, and free lunch programs.[9] This incentive structure is not unique to Florida. Discrepancies in new special education classifications have been observed in Denver, Colorado, for instance, where Winters (2015) shows that new special education classifications are disproportionately more likely to occur in traditional public schools relative to charter schools.[10]

There are other concerns that arise with the LEP and special education variables that relate to variation within these categories. CREDO groups all 24 distinct learning disabilities and

[9] Effective in 2013, after the period from which the data for this paper draws, Florida statutes permit charter schools to constitute their own "Local Education Agencies" for the purpose of receiving equal federal funding to public schools.

[10] Charter schools often have a limited administrative staff that more closely resembles that of a private school than a public school. They must deal directly with the state department of education, which can add an administrative burden to classifying individual students as LEP, special education, or free or reduced price lunch. Consequently, charter schools may be less likely to classify qualified students as eligible for these programs than public schools simply because of sector differences in the administrative costs of classification. This is potentially problematic for CREDO's matching strategy. If schools are less likely than public schools to classify students for the free or reduced-price lunch, LEP and special education programs, then CREDO risks inaccurately matching charter students to dissimilar public school counterparts. Specifically, the CREDO approach risk matching students that may be disadvantaged, but attend a charter school and thus are not labeled as such, with public school students that are almost certainly not disadvantaged. If eligibility for these programs is correlated with lower test score growth, then a naïve match on this characteristic that doesn't address the systematic measurement error can introduce a negative bias when estimating charter effects.

six LEP levels together,[11] creating one special education indicator variable and one LEP variable, determined by program participation, thus ignoring the severity of the cognitive disability or language deficiency. This means that the CREDO approach may match students with modest learning disabilities, such as Attention Deficit Hyperactivity Disorder (ADHD), with students who have severe learning disabilities, such as deafness, blindness or autism. Similarly, matched LEP students may speak different languages and be at different stages in their English language acquisition, but will be matched as if their language abilities are equal. This issue has the potential to introduce systematic bias if the charter and public sectors classify students differently, in a way that correlates with the severity of the disability or proficiency with the English language.[12]

Using New York City charter school lotteries, Winters (2013) finds that charter schools are less likely than public schools to classify students as in need of special education. In particular, Winters finds that charter schools are less likely to classify students as “specific learning disabled” and as having an “emotional disability,” which are commonly regarded as the most subjective and least severe of the special education disabilities. They also happen to be among the most common special education classifications, together compromising over half of all special education observations.

These program variables are often used for matching and control purposes because they predict student performance. However, it may be preferable to avoid using these variables if they do not add much predictive power to the estimation model and if they are potentially endogenous

[11] The Florida Department of Education identifies students as one of six different levels of English proficiency, but does not indicate what other language the student uses, if not English.

[12] Peterson and Llaudet (2006) assert that there is no definitive basis for what qualifies as “limited English proficiency” across schools, particularly across sectors, and suggest replacing this variable with a more objective measure, such as frequency a language other than English is spoken at home, as self-reported by students. They also recommend distinguishing whether special education status is due to a profound or moderate disability, if available.

because assignment for program participation depends on sector selection. Baseline test score and racial/ethnic information alone may be sufficient to explain the variation in student outcomes that would be captured by these three program variables, and allow for an increased number of observations. In our replication of CREDO's model (Table 3), removing the three program variables only lowers the proportion of variance that is explained in the dependent variable by 0.006 and 0.005 standard deviations for reading and math, respectively. That is, matching on baseline test score, race/ethnicity among students in similar geographic locations may be sufficient to account for individual student characteristics and avoids the use of potentially endogenous program variables.

Finally, it is important to note that differential participation biases can affect any study that uses these variables across school sectors. Although we focus on CREDO specifically, the implications of differential participation across sectors have much broader implications that should be accounted for in all research using these variables.

Appendix C: Analysis of Transfer Students Only

We further analyze the mismeasurement of the three program variables that CREDO incorporates as matching and control variables—special education, free or reduced-price lunch, and limited English proficiency—by reducing the sample to transfer students who transfer in particular directions. In Table C1, we compare changes in individual students' program participation classification among students who transfer from a public school to a charter school between 3rd and 10th grade (columns 2, 5, and 8) and students who moved from a charter school into a public school during those grades (columns 3, 6, 9).

All three designations seem to vary depending on the direction of the student transfer, but the increase in student classification for these three programs is particularly large among students who transfer from a charter to a public school. As reported in Table C1, the probability that a student who transfers from a charter to a public school at some point between 3rd and 10th grade is classified as special education, free-lunch eligible or as LEP increases 1.0 percent, 4.7 percent, and 1.5 percent, respectively, while they are in the public sector. All three increases are statistically significant.

We also repeat the primary analysis of charter school effectiveness using a subsample of both charter and public students who transfer from a public school in the prior year to investigate whether the change in estimated charter effect from removing the program variables is due to a bias in the variables or the regressions needing the variables as controls. In Table C2, we present regressions that match the same charter and public students using program variables from before they transfer schools, but using regression controls from both before and after the students transfer to see if the bias in classification across sectors influences results.

In both reading and math, the estimated charter effects appear more negative when the program variables from the charter sector are used. The negative reading effect increases from -0.023 to -0.019 standard deviations and the negative math effect increases from -0.034 to -0.030 standard deviations as the program variables are used in the public sector, before students transfer schools, which does not carry the bias. The difference in estimated charter effect using variables from the two sectors indicates that the bias in variable classification does indeed cause the change in estimated charter effect. Table C1 indicates that the difference in program classification across sectors is smaller among students who transfer from a public to a charter school than among all transfer students, so there is reason to believe that this 0.004 standard deviation bias for math and reading is an underestimate of the true bias.

Table C1. Individual-level Fixed Effects Estimates of Program Variables in TPS and Charter Schools for Subsamples

Dependent variable:	Special education			Free-lunch eligible			Limited English proficiency		
Sample:	All transfers	Public to charter transfers	Charter to public transfers	All transfers	Public to charter transfers	Charter to public transfers	All transfers	Public to charter transfers	Charter to public transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean of dependent variable	0.143	0.142	0.142	0.483	0.507	0.476	0.095	0.104	0.073
TPS Attendance	0.006**	0.004**	0.010**	0.038**	0.030**	0.047**	0.025**	0.018**	0.015**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Constant	0.088**	0.091**	0.086**	0.572**	0.602**	0.564**	0.187**	0.210**	0.145**
	[0.002]	[0.002]	[0.002]	[0.003]	[0.003]	[0.003]	[0.002]	[0.002]	[0.002]
Student FE	X	X	X	X	X	X	X	X	X
Year controls	X	X	X	X	X	X	X	X	X
Grade controls	X	X	X	X	X	X	X	X	X
Observations	561,750	423,398	317,959	561,750	423,398	317,959	561,750	423,398	317,959
Students	103,570	75,774	55,556	103,570	75,774	55,556	103,570	75,774	55,556
R-squared	0.018	0.019	0.015	0.022	0.025	0.022	0.088	0.099	0.062

** p<0.01, * p<0.05, + p<0.1

Notes: Units of observation are the students' program variables in student-years. Samples are restricted to all students who transfer across the public-charter sector, all students who at one point transfer from a public to a charter school, and all students who at one point transfer from a charter to a public school. These transfers must take place while the students are in tested grades (3rd to 10th) between 2001 and 2009. Special education, free-lunch eligibility and limited English proficiency are used as dependent variables. They equal 1 when a student participates in each respective program, and equal 0 otherwise. The interpretation would be that the probability a given student is classified as special education is 0.6% higher while the student is attending a traditional public school than while the same student is at a charter school. Considering the average rate of special education participation, this means that a student is 4% more likely to be classified as special education while in a traditional public school.

Table C2. Charter School Effectiveness for Public to Charter Transfer Students with Program Variables Measured in Either the Public or Charter Sector

	Reading		Math	
	(1)	(2)	(3)	(4)
Charter school	-0.023**	-0.019**	-0.034**	-0.030**
	[0.003]	[0.003]	[0.003]	[0.003]
Lagged reading score	0.760**	0.758**		
	[0.002]	[0.002]		
Lagged math score			0.791**	0.789**
			[0.002]	[0.002]
Special education	-0.053**		-0.066**	
	[0.005]		[0.005]	
Free lunch	-0.097**		-0.082**	
	[0.004]		[0.003]	
Limited English proficiency	-0.030**		0.000	
	[0.006]		[0.006]	
Lagged special education		-0.055**		-0.066**
		[0.005]		[0.005]
Lagged free lunch		-0.110**		-0.094**
		[0.004]		[0.003]
Lagged limited English proficiency		-0.019**		0.007
		[0.006]		[0.006]
Constant	0.107**	0.113**	0.023**	0.026**
	[0.009]	[0.009]	[0.009]	[0.009]
Matching variables:				
Special Education	X	X	X	X
Free Lunch	X	X	X	X
LEP	X	X	X	X
Charter students	65,906	65,906	66,036	66,036
R-squared	0.111	0.113	0.103	0.105

** $p<0.01$, * $p<0.05$, + $p<0.1$

Notes: The dependent variable is the change in test score from the previous year to the current year. Charter students and matched public students included. Both the treatment and control samples are limited to students who transferred from a traditional public school the prior year. Students are matched using program variables from before they transfer schools, while they are all in the public sector. Lagged variables refer to those variables while both treatment and control subjects are in the traditional public school sector the year before they transfer schools. Otherwise, the variables refer to after the student transfers schools. All test scores are standardized separately for each grade level and each year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test. All regressions control for prior test score, whether a student repeated a grade, gender, race, grade and year. All students are matched for exactly two consecutive years, and 1-year effects are reported. Standard errors are in parentheses.

Appendix D: Summary Statistics for Over and Undersubscribed Charter Schools

Table D1 presents summary statistics for oversubscribed and undersubscribed charter schools. Notably, students' prior test scores are higher for students attending oversubscribed charters. The magnitude of the difference between over and undersubscribed charters is 0.154 standard deviations compared to -0.142 in math and 0.190 standard deviations compared to -0.089 in reading. Regarding student demographic characteristics, students in oversubscribed charters are less likely to be female, less likely to be Black or White, and more likely to be Hispanic. Students in oversubscribed charters are also less likely to be eligible for free or reduced-price lunch and less likely to be classified as requiring special education services or as limited English proficient.

Table D1. Summary Statistics for Over and Undersubscribed Charter Schools (2005-2009)

	TPS		Charter Schools		P-Value	Undersubscribed Charters		Oversubscribed Charters		P-Value
Students	2,419,009		115,207		of	49,716		76,477		of
Student-year observations	5,919,941		207,780		Difference	72,461		135,319		Difference
Prior Test Scores	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Math score	-0.002	[1.003]	0.051	[0.906]	0.00	-0.142	[0.938]	0.154	[0.871]	0.00
Reading score	-0.003	[1.002]	0.093	[0.914]	0.00	-0.089	[0.943]	0.190	[0.883]	0.00
Other Baseline Covariates		Percentage		Percentage			Percentage		Percentage	
Grade										
3rd		12.9		14.3	0.0		16.0		13.3	0.0
4th		12.3		13.3	0.0		14.3		12.7	0.0
5th		12.4		12.9	0.0		13.6		12.6	0.0
6th		12.2		15.2	0.0		13.1		16.3	0.0
7th		12.4		14.0	0.0		11.5		15.3	0.0
8th		12.5		12.4	0.3		9.9		13.8	0.0
9th		12.8		8.8	0.0		11.1		7.6	0.0
10th		12.4		9.2	0.0		10.5		8.4	0.0
Sex										
Female		50.8		49.0	0.0		49.8		48.5	0.0
Male		48.9		50.8	0.0		49.9		51.2	0.0
Race/Ethnicity										
Asian or Pacific Islander		2.4		1.9	0.0		1.5		2.1	0.0
Black, not of Hispanic origin		22.9		18.1	0.0		25.3		14.3	0.0
Hispanic		24.1		34.6	0.0		25.3		39.6	0.0
Multiracial		3.3		3.4	0.1		3.8		3.2	0.0
White, not of Hispanic origin		47.0		41.7	0.0		43.6		40.7	0.0
Programs										
Special education		18.9		13.5	0.0		13.8		13.3	0.0
Free-lunch eligible		48.0		38.2	0.0		45.0		34.6	0.0
Limited English proficiency		9.5		8.0	0.0		8.9		7.5	0.0

Missing Value Indicators						
Math score	1.6	1.5	0.0	1.8	1.4	0.0
Reading score	1.7	1.2	0.0	1.9	0.9	0.0
Sex	0.3	0.3	0.4	0.3	0.3	0.3
Race/Ethnicity	0.0	0.0	0.1	0.0	0.0	0.0

Notes: All test scores are standardized separately for each grade level and year, such that the reading/math scores have a mean of 0 and a standard deviation of 1 among all students in Florida who have taken the test. Oversubscribed is a binary variable that equals 1 if the charter school has a waiting list, indicating oversubscribed charter schools, and equals 0 otherwise. Years 2005-2009 are used because these are the years oversubscribed charter school data are available. Some students attend both undersubscribed and oversubscribed charter schools and some charter schools are oversubscribed one year and undersubscribed another year.