NATIONAL
CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

AIR

One Day Too Late? Mobile Students in an Era of Accountability

UMUT ÖZEK

WORKING PAPER 82 • OCTOBER 2012

One Day Too Late? Mobile Students in an Era of Accountability

Umut Özek
American Institutes for Research

Contents

Acknowledgements

This is a considerably revised version of a paper with the same name that first appeared as CALDER working paper 82. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110968 to the American Institutes for Research. I would also like to thank the Florida Department of Education for providing the data, David Figlio, Rajashri Chakrabarti and the seminar participants at the Urban Institute, APPAM, AEFP, AERA, and AIR for useful comments. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education. All errors are mine.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

One Day Too Late? Mobile Students in an Era of Accountability

Umut Özek

CALDER Working Paper No. 82
October 2012

Abstract

How to incorporate mobile students, who enter schools/classrooms after the start of the school year, into educational performance evaluations remains to be a challenge. As mandated by the No Child Left Behind Act of 2001 (NCLB), all states currently require that a school is accountable only if the student has been enrolled in the school for a full academic year. This paper investigates the school response to this eligibility requirement in a regression-discontinuity framework. Comparing students who enter schools right before and after the eligibility cutoff, I find no evidence that schools behave strategically in response to this requirement.

1. Introduction

Accountability has become a mantra in public education more than a decade after the No Child Left Behind Act of 2001 (NCLB) was signed into law. The enactment of this federal law accelerated the national trend towards an educational regime where schools are held accountable for the performance of their students, primarily by imposing sanctions such as the threat of losing federal funds unless a state implemented a school accountability system meeting several requirements. Furthermore, demands for greater accountability have been intensifying beyond simple school-level accountability as the focus of educational accountability shifts from institutions to individual educators. Over the last decade, several federal laws and policies have incentivized states to develop individual-level systems where teachers and principals are personally held responsible for their students' performances.[1] A recent example is the Race to the Top (RTTT) competition, which provided significant impetus for states to require evidence of student learning in teacher evaluations.[2]

The centerpiece in a sustainable accountability system is a fair assessment mechanism that yields the correct allocation of the blame/reward for the failure/success of individual students among educational production function inputs (e.g. schools, teachers, parents). An important challenge in efforts to isolate the contribution of individual schools/educators on student outcomes is mobile students who enter schools and/or classrooms after the beginning of the school year. Unless taken into account, student mobility across schools/classrooms might lead to the incorrect attribution of student performance to the effectiveness of schools/teachers in the spring semester. This misattribution is particularly consequential for schools and educators serving disadvantaged populations where within-semester student turnover rates are typically higher. For instance, in Florida, one of the few states that

[1] The introduction of individual-level accountability is particularly important because of the role of teachers and principals as the most consequential school-level factors in the production of education (Goldhaber et al. (1999); Hanushek (1986); Rivkin et al. (2005) and Clark et al. (2009)).

[2] Specifically, the section on Great Teachers and Leaders of RTTT, which requires states to develop data-based teacher evaluation systems and use these evaluations to inform key decisions such as hiring, compensation and retention, carries the largest weight in RTTT application reviews.

keeps track of student mobility during the school year, roughly 9 percent of all public school students each year enter the schools at which they are ultimately tested in the spring at least a month after the beginning of the school-year. On the other hand, at schools where at least 80 percent of students are free or reduced priced lunch (FRPL) eligible, 'late-entrants' account for approximately 15 percent of spring enrollment.[3]

This paper investigates the unintended consequences of the full academic year (FAY) eligibility requirement - the current strategy to incorporate mobile students into school evaluations in all states. Under NCLB, a school is accountable for a student's performance only if the student has been enrolled in the school for a full academic year. In other words, even though all students are required to take standardized tests in certain grades and subjects, the test score of a given student can only be used to evaluate his/her school if she has attended that school for a 'full academic year', a critical element that must be defined by each state and approved by the Department of Education in order to comply with the federal law. As of 2014, all 50 states and the District of Columbia had the FAY requirement incorporated into their school accountability systems. Almost all of these systems identify two critical dates, typically one at the beginning of the school year and the other close to the testing window, and define eligible students for a given school as those who were enrolled in that school during both dates.[4]

This interesting aspect of the policy, which is intended to ensure that schools with high within-semester student turnover are not unfairly punished, creates a clear incentive for schools to behave strategically in two ways. First, schools might attempt to strategically reclassify low-performing students to alter the composition of 'eligible' test-takers whose scores are used in the assessment of schools.[5]

[3] Author's calculations from administrative student-level data for years between 2002 and 2011.
[4] More specifically, 39 states and the District of Columbia currently define FAY in this way.
[5] Several studies in the school accountability literature have found evidence of such strategic reclassification in other contexts. For instance, Cullen and Reback (2006), Figlio and Getzler (2006) and Jacob (2005) have presented evidence that schools classify low-performing students into special education categories that are either exempt from test-taking or are not used to evaluate school performance. Similarly, Figlio (2006) has found that schools tend to assign harsher punishments to low-performing students during the testing period compared to their higher achieving peers, manipulating the test-taker pool.

Second, schools might find it beneficial to strategically assign students to classrooms based on their eligibility in an attempt to allocate their resources to boost the 'assessed' student performance. In order to investigate these hypothesis, I utilize detailed student-level administrative data from Florida, which uses surveys conducted in October and in February to identify the FAY eligible students under its accountability system, 'Florida's A+ Plan'.

Comparing the students who enter the school at which they take the test right before the October eligibility cutoff to those who enter right after the cutoff and thus become ineligible, regression-discontinuity results provide no evidence for the existence of such strategic behavior. In particular, just-ineligible students are statistically indistinguishable from their just-eligible peers in the same school along a multitude of observable dimensions including current and past year test performance, misbehavior, socioeconomic status and demographics. Further, I find no significant discontinuity in the density of entering students around the eligibility cutoff, providing evidence against the strategic sorting hypotheses. I also compare the teacher and classroom characteristics of just-eligible and ineligible students and find no evidence of strategic placement of ineligible students into different classrooms. These findings remain unchanged when the analysis is constrained to 'near-failing' and 'failing' schools that face the highest accountability pressure.

2. Policy Background

2.1. NCLB and the FAY Requirement

The No Child Left Behind Act of 2001, signed into law on January, 8, 2002, authorized the Department of Education to withhold federal funds unless a state implemented an accountability system incorporating various 'critical elements' of the federal legislation such as the mandate to cover all public schools and students in the state, several factors that determine adequate yearly progress of schools and local education agencies, and subgroup accountability requirements. As part of NCLB, all states

were required to submit detailed implementation information on these elements to the Department of Education by January 31, 2003, and apply them during the 2002-2003 school year.

One of these critical elements is that the state accountability system has a consistent definition of FAY. This requirement arose from Section 1111(b)(3)(C)(xi) of the federal law, which prohibits states from using the test scores of FAY ineligible students, who have attended more than one school in any academic year, for school accountability purposes.[6] Consequently, all states and the District of Columbia adopted accountability systems that define FAY in various ways. Appendix A lists these definitions as of 2014. Several states use 'the number of days enrolled at the school before statewide testing' to define FAY-eligible students, whereas the majority of the states, including Florida, and the District of Columbia identify two dates, one typically at the beginning of the school year and the other before the statewide testing window, and define FAY-eligible students as those who were enrolled at the school in which they were tested during both dates. In what follows, I describe the school accountability system in Florida, which took effect prior to the adoption of NCLB, and how it incorporates the FAY requirement.

2.2. School Accountability in Florida: Florida's A+ Plan

Enacted in 1999, Florida's A+ Plan employs school-level, performance-based rewards, sanctions and assistance in order to achieve the set of proficiency benchmarks described in the Sunshine State Standards and approved by the State Board of Education in 1996. Beginning in the summer of 1999, each public school is assigned a grade from A to F based on the performance of its students in curriculum standards-based Florida Curriculum Assessment Test (FCAT-SSS).

On the rewards side, monetary awards are given to schools that improve a letter grade or maintain an 'A'. Sanctions include increased scrutiny and oversight for schools that receive a 'near-failing' grade ('D') or a 'failing' grade ('F') as well as a voucher program called Opportunity Scholarship

[6] Section 1111(b)(3)(C)(xi) of the legislation, in its entirety, reads as follows: "Such assessments shall…include students who have attended schools in a local educational agency for a FAY but have not attended a single school for a FAY, except that the performance of students who have attended more than 1 school in the local educational agency in any academic year shall be used only in determining the progress of the local educational agency."

for students attending chronically low-performing (CLP) schools that receive a grade of 'F' in two out of the past four years including the current year. Opportunity Scholarship allows students in CLP schools to attend a higher-performing public school of their choice. Additionally, up through the 2005-06 school-year, Opportunity Scholarship allowed students to attend an eligible private school. Florida's accountability system also provides schools with recommendations on how to improve as well as and technical and instructional support, prioritizing 'D' and 'F' schools. Furthermore, as shown in Goldhaber and Hannaway (2004), the receipt of 'D' or 'F' carries significant social stigma for teachers and principals, providing schools additional motivation to improve.

Between the 1998-99 and 2001-02 school years, FCAT-SSS achievement levels were the primary determinants of school grades. During this time period, students in fourth grade were tested in FCAT-SSS reading and writing, fifth graders in math, and eighth and tenth graders in all three subjects. During the 2001-02 school year, the grading formula under the A+ Plan went through a major revision. Under the new formula, school grades incorporate FCAT-SSS reading and math achievement levels in all grades between three and ten along with the year-to-year progress of students in these subjects with special attention to the reading gains of students in the lowest quartile in reading at each school.[7]

While students in grades three through ten have been required to take FCAT-SSS in reading and math since 2002, the calculation of school grade does not incorporate the scores of all students. Under the A+ Plan, there are three criteria that determine student eligibility in school assessments:

i. Limited English Proficiency (LEP) eligibility: LEP students are included in the school grading formula if they have been in the English for Speakers of Other Languages (ESOL) program for more than two years prior to testing.

[7] Since the 2006-2007 school year, math gains of students in the lowest quartile at their corresponding schools along with the achievement levels of students in grades 5, 8 and 10 in science began to be incorporated in the grading formula.

ii. Exceptional Student Education (ESE) eligibility: ESE students are included in the school grade calculations if their only exceptionality is gifted, hospital/homebound, speech impaired, or a combination of these three.

iii. FAY eligibility: Students are included in the school grading formula if they were present in the same school during the October and February full-time equivalency (FTE) counts (surveys). The October survey typically takes place in mid-October whereas the February survey is conducted in the first week of February, roughly a month before the standardized testing window in Florida.[8]

The first two eligibility requirements have been in effect since the adoption of the A+ Plan in 1999 whereas the FAY-eligibility requirement was introduced in 2000. Beginning with the 2004-2005 school year, school grade calculations incorporated gains in reading and math achievement of LEP and ESE students; however, calculations have excluded the test score levels of students in all three categories along with the test score gains of FAY-ineligible students.

Aside from the October and February surveys, the Florida Department of Education (FLDOE) conducts three other surveys throughout the school year. The primary purpose of these surveys, each of which is conducted over a week, is to determine full-time equivalency counts of students, which are then used for school and school district funding decisions. In order for a student to be included in the full-time equivalency count of a school, he/she must have at least one day of membership in that school during the survey week. This requirement creates an eligibility cutoff for students at the schools in which they are tested where those who enter on or before the Friday of the October survey week are considered FAY-eligible and the students who enter on or after the Monday of the following week are excluded from school grade calculations. As further discussed in the following section, this discontinuity is the key element of my identification strategy.

[8] For K-12 public schools in Florida, instruction typically begins in August and ends at the end of May.

3. Data Description and Empirical Strategy

3.1. Data Description

In the analyses that follow, I utilize student-level administrative data on all elementary and middle school students between grades four and eight from 2003-2004 to 2011-2012 with non-missing prior year test scores in Florida. The dataset includes demographic information on students such as race, gender, FRPL eligibility, LEP status, LEP program entry and exit dates, ESE status at the time of each survey as well as FCAT-SSS reading and math scores. The most critical piece of information contained in the dataset for the purposes of this study is the entry date of each student to the school(s) she attended in a given school year, which enables me to identify FAY-eligible students using the eligibility cutoff dates given in Table 1. In order to examine the school response to accountability by school grade, I also utilize the accountability grades for all public schools in Florida between 2003-2004 and 2011-2012 school years. Table 2 presents the grade distribution for the subset of public schools that are used in the analysis.

Table 3 presents the descriptive statistics for all FAY eligible students in the sample (first column) along with all FAY ineligible students (second column), eligible students who enter their test schools in the month before the October cutoff (third column), and ineligible students who enter in the month after the cutoff (fourth column). During this time frame, approximately 6% of the students in the sample enter their test schools after the October eligibility cutoff. Compared to their eligible peers, FAY ineligible have significantly lower prior reading and math scores, are significantly more likely to have had disciplinary issues and to have been retained during the prior school year, more likely to come from economically disadvantaged families, more likely to belong to a racial/ethnic minority group (other than Asian), and more likely to be LEP ineligible and special education students. Restricting the sample to the mobile students around the cutoff reduces these differences considerably, yet these 'just-ineligible'

students still have worse academic outcomes in the prior school year compared to their 'just-eligible' peers.

3.2. Empirical Framework

Let S_i denote the number of school weeks between the entry week of student *i* to the school she was tested and the October cutoff date, with negative values indicating entry before the cutoff. Defining treatment, T_i, as being FAY-ineligible, a common regression model representation of this evaluation problem would become:

$$Y_i = \alpha + \beta T_i + \varepsilon_i \tag{1}$$

where Y_i is the test score of student *i*, standardized to mean zero and unit variance, and T_i is a deterministic function of S_i where $T_i = 1(S_i \geq 0)$. Provided that the conditional mean function $E[\varepsilon \mid S]$ is continuous at the eligibility cutoff, the causal impact of eligibility on student achievement is given by:

$$\beta = \lim_{S \downarrow 0} E[Y \mid S] - \lim_{S \uparrow 0} E[Y \mid S] \tag{2}$$

There are several ways to estimate β in this context. First is to estimate equation (2) non-parametrically using kernel-weighted local polynomial smoothing (Hahn et al. (2001), Porter (2003)). However, when the selection variable is discrete, as in this case, non-parametric estimator might lead to biased estimates as it is not feasible to compare averages within arbitrarily small neighborhoods around the cutoff (Card and Lee (2008)). Therefore, following Card and Lee (2008), I estimate equation (2) parametrically using the following framework:

$$Y_i = \alpha + \beta T_i + k(S_i) + k(S_i) * T_i + \varepsilon_i \tag{3}$$

where $k(S_i)$ is a polynomial function of the relative entry week. In the preferred specification, I limit the analysis to students within a bandwidth of 4 weeks, since increasing bandwidth is expected to

produce biased estimates in situations such as the case examined here where the selection variable is correlated with the outcome conditional on treatment status. I use linear and quadratic polynomials to check the robustness of the estimates, and two-way cluster the standard errors at the school-year and relative entry week level (Cameron et al. (2011), Card and Lee (2008)).

4. Results

4.1. Strategic Classification of Students around the Eligibility Cutoff

I first check to see whether schools strategically classify low-performing students into the FAY-ineligible category. There are several symptoms of such behavior. First, if schools sort students based on their observed characteristics (e.g. prior achievement and socioeconomic status), one would expect to observe significant discontinuities at the eligibility cutoff along these dimensions. Even in the absence of such differences between just-eligible and just-ineligible students, unusual changes in the density of entering students at the eligibility cutoff might hint at strategic classification along unobservable student traits. In either case, significant discontinuities in current year student test scores are expected.

Figure 1 explores this hypotheses and compares the current year reading and math scores of students around the eligibility cutoff for all schools in the first panel, 'A' schools in the second panel, and 'D' and 'F' schools in the third panel, with solid circles representing the average test score for each entry week that are centered at the October survey week (i.e., positive values of relative entry week represent ineligible students). The figures in the first two panels suggest a negative correlation between student achievement and entry week, while student test scores in both reading and math seem to be smooth around the eligibility cutoff. However, for 'D' and 'F' schools, average student achievement trends up in the weeks prior to the October eligibility cutoff, yet declines during the weeks after the cutoff, with considerably larger discontinuities at the cutoff.

Table 4 presents the estimated discontinuities in current year test scores at the FAY eligibility cutoff. In the first two columns, I estimate equation (3) using a bandwidth of 4 weeks and a linear $k(S_i)$ whereas the last two columns use a quadratic polynomial. All specifications include cohort fixed-effects to take differences between cohorts into account. The estimated effects reported in columns (I) reinforce the earlier graphical analysis. There are no statistically significant discontinuities in student achievement at the eligibility cutoff for all school types, while the estimated discontinuities for near-failing and failing schools are larger in magnitude. Columns (II) introduce school-by-year fixed effect to take into account the differences in schools attended between just-eligible and just-ineligible students, which might take place, for instance, if the vacant seats at higher quality schools fill up faster than those at other schools. Comparing the just-ineligible students with students who enter the same school in the week right before the eligibility cutoff, results remain unchanged.

Figures 2 and 3 compare the students on the two sides of the eligibility cutoff along their baseline characteristics including prior year test scores (results given in Figure 2), prior year disciplinary incidents and retentions, free or reduced priced lunch eligibility, race/ethnicity, native language, immigrant status, and LEP and ESE eligibility (Figure 3). This graphical analysis suggests no significant differences along these observable dimensions between just-eligible and just-ineligible students. Table 5 replicates the analysis in Table 4 using these baseline characteristics, yielding the same conclusion. While a few of the estimated discontinuities are statistically different from zero, there are no consistent differences between the two student groups that would suggest strategic sorting.

Finally, Figure 4 examines the entry date density around the eligibility cutoff for all schools (first panel), for 'A' schools (second panel), and for 'D' and 'F' schools (third panel). The findings present no evidence of strategic sorting along unobservable traits, with the number of entering students in the week after the cutoff almost identical to the number of entering students during the October survey

week (21,408 versus 21,148). This holds true for 'A' schools (8,954 versus 9,035), and 'D' and 'F' schools (1,209 versus 1,232).

4.2. Strategic Classroom Assignments

Another way schools might behave strategically in response to the eligibility requirement is by allocating their resources strategically in order to boost the test performance of students whose scores are used for accountability purposes. In particular, schools might find it beneficial to strategically assign eligible students to 'better' classrooms with more effective teachers and higher-performing peers. If this is the case, one would expect to observe significant differences between classroom characteristics (e.g. teacher attributes, classroom peer performance) of students around the eligibility cutoff.

In order to test this hypothesis, I first identify the reading and math classrooms of students, and drop classrooms with fewer than 10 students (5 percent of students in the sample are in such classrooms). I also drop schools with only one classroom per subject-grade-year (3 percent of students in the sample are in such schools) for which strategic assignments would not be possible. I then compare just-eligible students with their just-ineligible peers along several classroom attributes including average current and prior year test scores of their classroom peers in reading and math, teacher experience, whether the teacher is early career (i.e., in the first three years of his/her teaching career), has an advanced degree, holds a regular teaching license, and certified in the corresponding subject.

Table 6 presents the estimates for all schools (columns 2 and 3), 'A' schools (columns 4 and 5), and 'D' and 'F' schools (columns 6 and 7), examining the discrepancies in reading classrooms in the top panel and the math classrooms in the bottom panel. All specifications include cohort and school-year fixed effects, so I compare the classroom characteristics of eligible students with students at the same grade level who enter the same school in the weeks after the eligibility cutoff. I also include student baseline characteristics to ensure that the possible differences in classroom assignments are not driven by differences in student attributes between eligible and ineligible students. The results indicate that

just-ineligible students are assigned to comparable classrooms as their just-eligible peers in both reading and math, providing evidence against the strategic classroom assignments hypothesis.

5. Concluding Remarks

The No Child Left Behind Act of 2001 mandates states to implement a consistent definition of FAY incorporated into their accountability systems and prohibits the use of performance of students who have attended more than one school in any academic year in school assessments. As of 2014, all states had adopted the FAY requirement whereby some students are typically labeled ineligible based on their entry dates to the school and a predetermined eligibility cutoff.

In this study, I investigate two possible school responses to NCLB's FAY eligibility requirement using Florida's school accountability policy. First, I check to see whether schools strategically reclassify low-performing mobile students in an attempt to change the test-taking student pool whose test scores are used for accountability purposes. This is similar to the evidence presented in the previous literature in special education context where several studies have shown that schools reclassify low-performing students into special education categories that are excluded from school performance evaluations (Cullen and Reback (2006), Figlio and Getzler (2006) and Jacob (2005)). Second, schools might find it beneficial to allocate their resources strategically to boost the assessed student performance by clustering ineligible students into classroom with less effective teachers and low-performing peers. While both of these responses are undesired, they carry starkly different implications for students, with the latter scenario expected to be more detrimental for FAY ineligible students.

Comparing students who enter their test schools before and after the October eligibility cutoff, regression discontinuity estimates present no evidence of strategic behavior. In particular, I find that just-eligible students are comparable to their just-ineligible peers along a number of outcomes including current and prior year test scores, disciplinary incidents, retention, and other baseline characteristics,

ruling out the strategic sorting hypotheses. Further, the results indicate that just-ineligible students attend classrooms with similar teachers and peers as their just-eligible peers at the same school, providing evidence against the strategic resource allocation hypotheses. These findings remain unchanged when the analysis is constrained to 'near-failing' and 'failing' schools that face the highest accountability pressure.

References

Cameron, C.A., Gelbach J., Miller D., (2011). Robust Inference with Multi-way Clustering. Journal of Business and Economic Statistics, 29 (2): 238-49.

Card, David and David Lee, 2008. Regression Discontinuity Inference with Specification Error. Journal of Econometrics, 142 (2): 655-674.

Clark, Damon, Paco Martorell and Jonah E. Rockoff, 2009. School Principals and School Performance. Working Paper 38, National Center for Analysis of Longitudinal Data in Education Research.

Cullen, Julie Berry and Randall Reback, 2006. "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." in Improving School Accountability: Check-Ups or Choice, Advances in Applied Microeconomics, eds. Timothy J. Gronberg and Dennis W. Jansen, 14, Amsterdam: Elsevier Science.

Figlio, David, 2006. Testing, Crime and Punishment. Journal of Public Economics 90(4): 837-851.

Figlio, David and Lawrence Getzler, 2006. "Accountability, Ability and Disability: Gaming the System?" in Improving School Accountability: Check-Ups or Choice, Advances in Applied Microeconomics, T. Gronberg and D. Jansen, eds., 14 (Amsterdam: Elsevier Science).

Figlio, David and Joshua Winicki, 2005. Food for thought: the effects of school accountability on school nutrition. Journal of Public Economics 89 (2–3), 381–394.

Goldhaber, Dan and Jane Hannaway, 2004. Accountability with a Kicker: Preliminary Observations on the Florida A+ Accountability Plan. Phi Delta Kappan 85(8), 598-605.

Goldhaber, Dan, Brewer, Dominic, and Anderson, Deborah, 1999. A Three-Way Error Components Analysis of Educational Productivity. Education Economics 7(3), 199-208.

Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw, 2001. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. Econometrica 59 ,201-209.

Hanushek, Eric A., 1986. The Economics of Schooling: Production and Efficiency in Public Schools. Journal of Economic Literature 49(3), 1141-1177.

Jacob, Brian, 2005. Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools. Journal of Public Economics 89(5), 761–96.

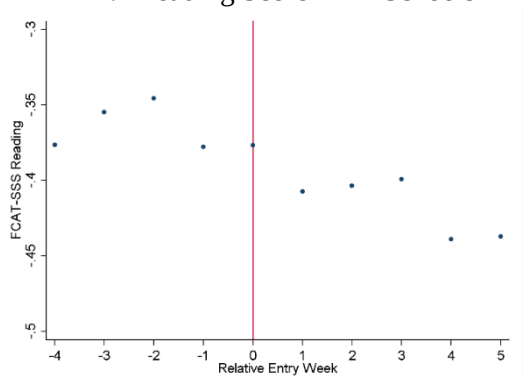
Porter, Jack, 2003. Estimation in the Regression Discontinuity Model.", Unpublished manuscript, Harvard University.

Rivkin, Steven G., Eric Hanushek and John F. Kain, 2005. Teachers, Schools and Academic Achievement. Econometrica 73(2), 417-458.

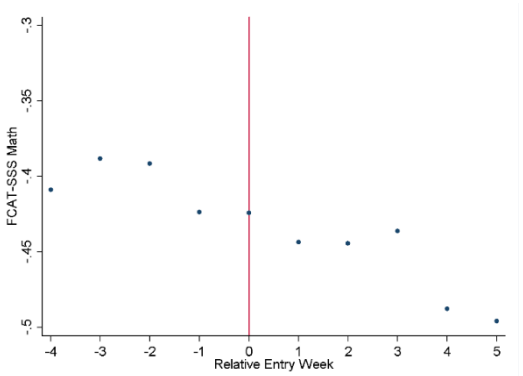
Figures

Figure 1 – FAY Eligibility and Student Achievement

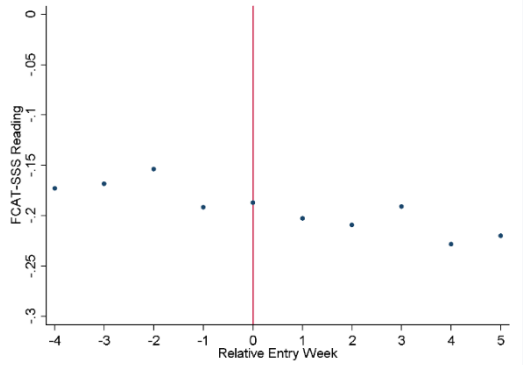
A. Reading Score – All Schools



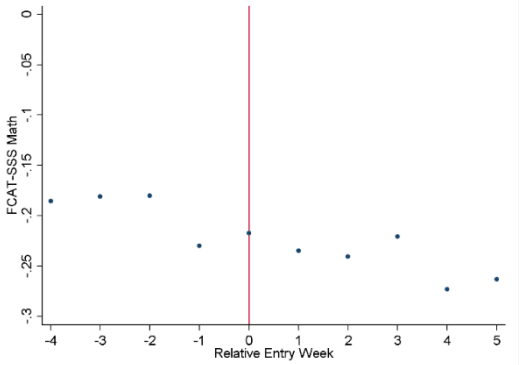
B. Math Score – All Schools



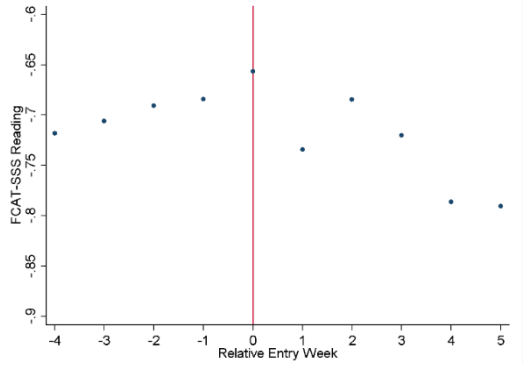
C. Reading Score – 'A' Schools



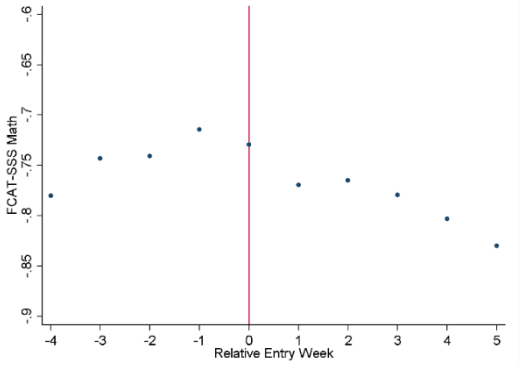
D. Math Score – 'A' Schools



E. Reading Score – 'D' and 'F' Schools



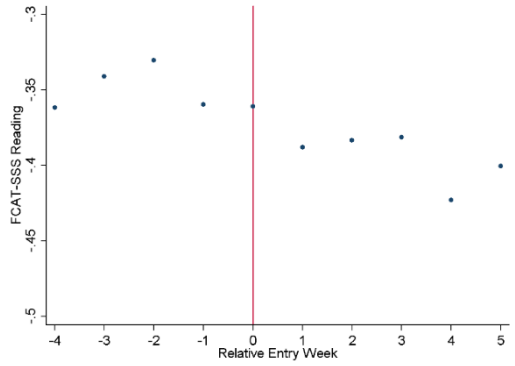
F. Math Score – 'D' and 'F' Schools



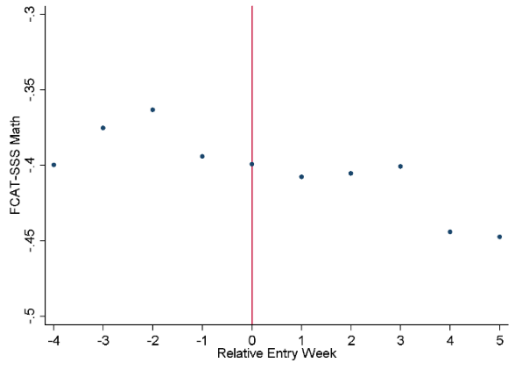
Notes: The solid circles represent raw averages of the corresponding outcomes for each entry week. The relative entry week of zero corresponds to the week before the eligibility cutoff (i.e., the October survey week).

Figure 2 – FAY Eligibility and Prior Year Student Achievement

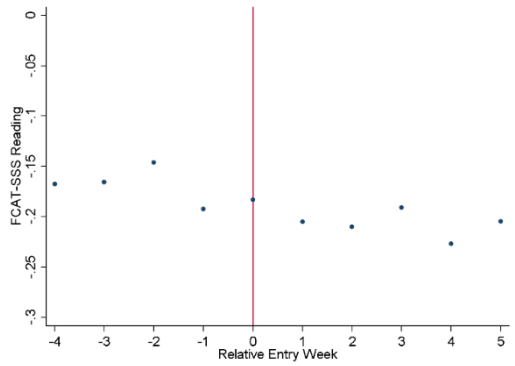
A. Reading Score – All Schools



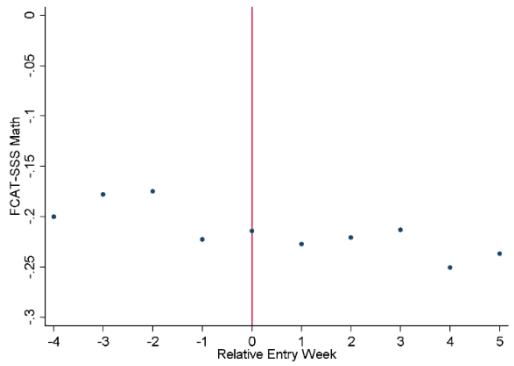
B. Math Score – All Schools



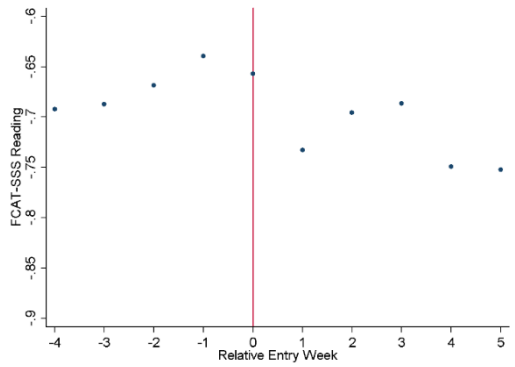
C. Reading Score – 'A' Schools



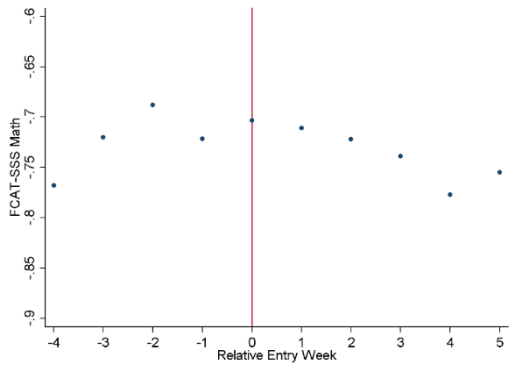
D. Math Score – 'A' Schools



E. Reading Score – 'D' and 'F' Schools

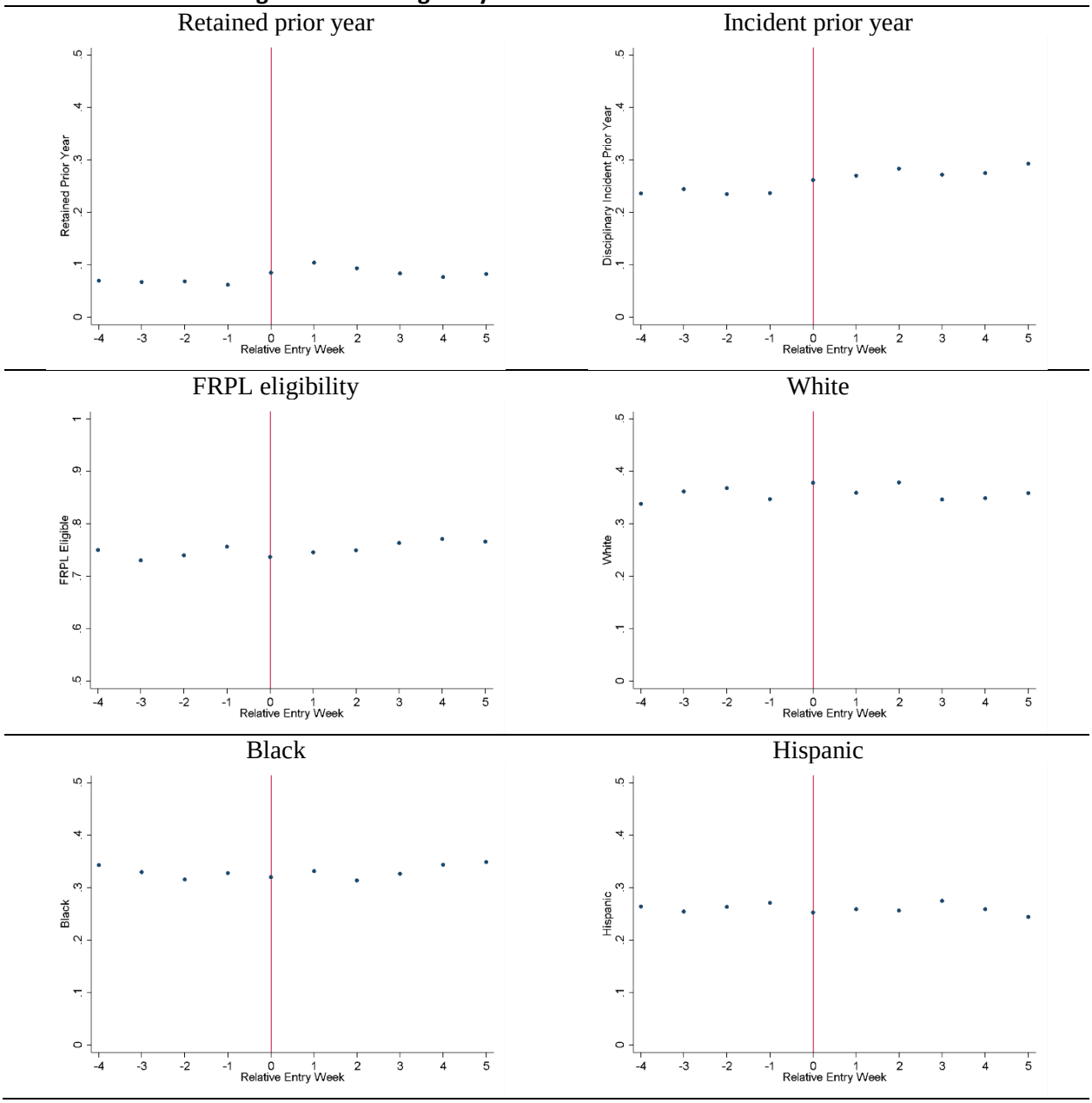


F. Math Score – 'D' and 'F' Schools



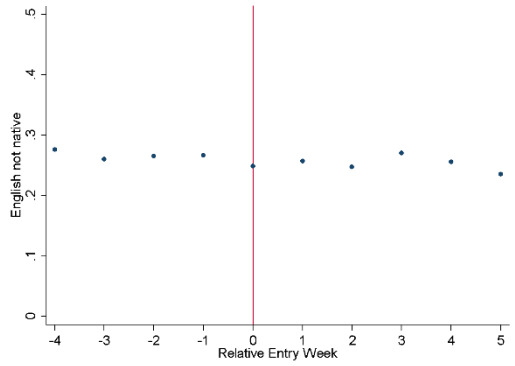
Notes: The solid circles represent raw averages of the corresponding outcomes for each entry week. The relative entry week of zero corresponds to the week before the eligibility cutoff (i.e., the October survey week).

Figure 3 – FAY Eligibility and Student Baseline Characteristics

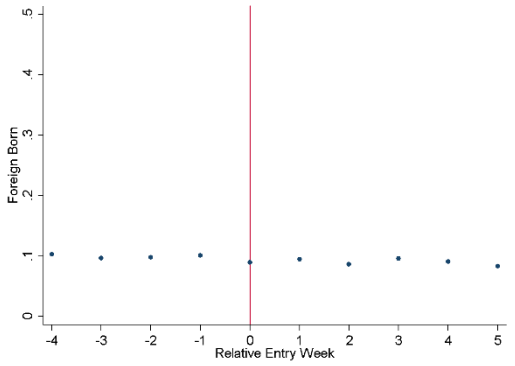


(Figure 3 continued)

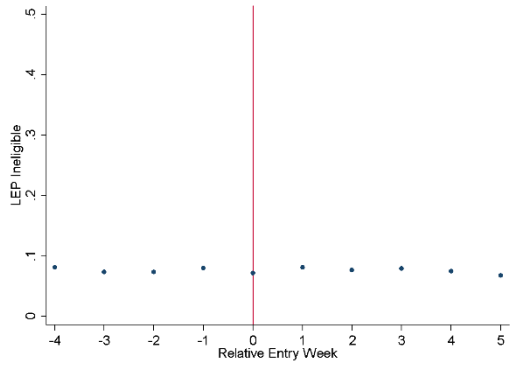
English not native



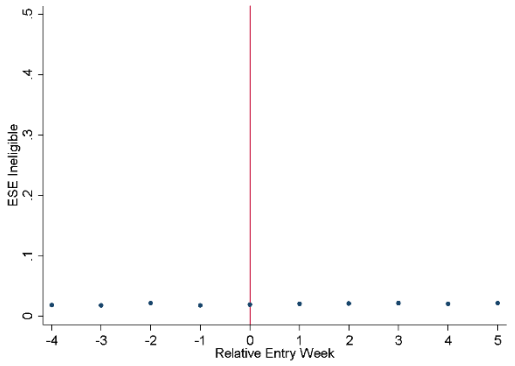
Foreign born



LEP ineligible



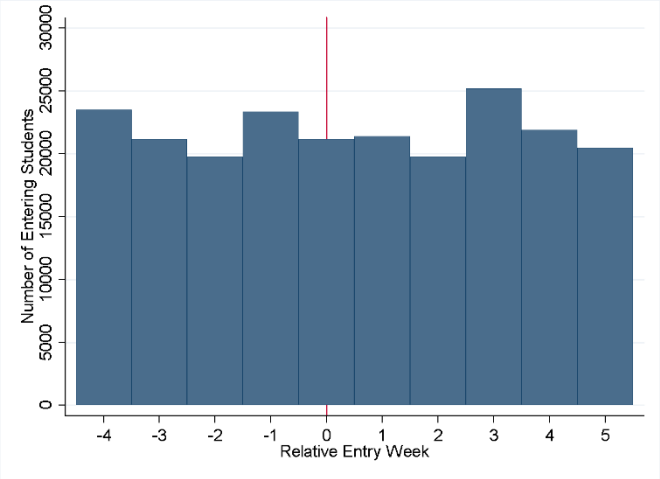
ESE ineligible



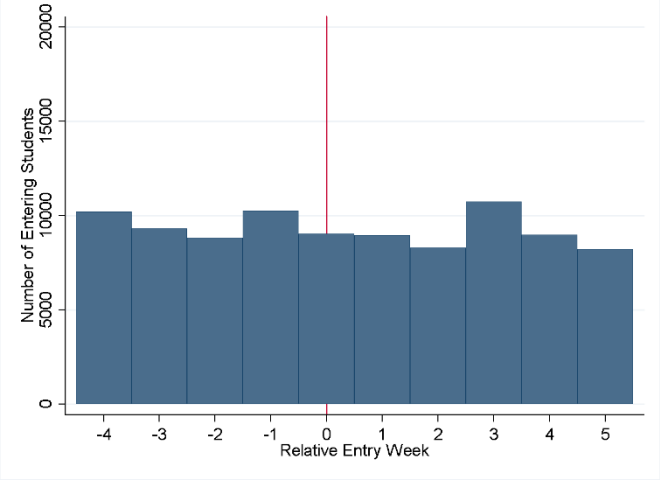
Notes: The solid circles represent raw averages of the corresponding outcomes for each entry week. The relative entry week of zero corresponds to the week before the eligibility cutoff (i.e., the October survey week).

Figure 4– Entering Student Density around the Eligibility Cutoff

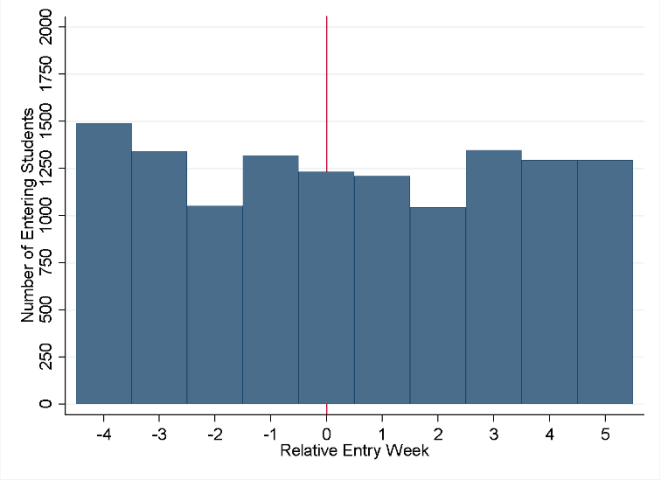
A. All Schools



B. 'A' Schools



C. 'D' and 'F' Schools



Notes: The three panels present the number of entering students in the sample between 5 weeks before and 5 weeks after the October eligibility cutoff. The relative entry week of zero corresponds to the week before the eligibility cutoff (i.e., the October survey week).

Tables

Table 1 - October Survey Dates

School Year	Survey Week	Eligibility Cutoff Date
2003-2004	October, 13-17	October, 20
2004-2005	October, 11-15	October, 18
2005-2006	October, 10-14	October, 17
2006-2007	October, 9-13	October, 16
2007-2008	October, 8-12	October, 15
2008-2009	October, 13-17	October, 20
2009-2010	October, 12-16	October, 19
2010-2011	October, 11-15	October, 18
2011-2012	October, 10-14	October, 17

Notes: Eligibility cutoff date is defined as the first day of FAY ineligibility. The dates were compiled from Appendix B in the User Manuals for each year, posted on Florida Department of Education (FLDOE) website: http://www.fldoe.org/eias/dataweb/archive.asp , accessed 03/27/2015.

Table 2 - School Grade Distribution: 2002-2003 to 2011-2012

	School Grade					
School Year	A	B	C	D	F	Total
2003-04	1,243	566	538	136	35	2,518
2004-05	1,263	541	616	184	49	2,653
2005-06	1,255	589	619	230	78	2,771
2006-07	1,467	610	570	122	21	2,790
2007-08	1,483	469	587	216	83	2,838
2008-09	1,585	542	565	155	45	2,892
2009-10	1,823	493	422	171	43	2,952
2010-11	1,533	673	557	147	41	2,951
2011-12	1,629	676	530	138	29	3,002

Notes: Author's calculations from state data. Elementary and middle schools used in the analysis include all schools serving any of the tested elementary and middle grades in Florida, namely grades 3 through 8.

Table 3 – Descriptive Statistics

	All		One month around cutoff	
	FAY-Eligible	FAY-Ineligible	FAY-Eligible	FAY-Ineligible
Prior year reading score	0.0644	-0.403	-0.349	-0.394
	(0.971)	(0.989)	(0.994)	(0.992)
Prior year math score	0.0656	-0.443	-0.384	-0.414
	(0.970)	(1.010)	(1.010)	(0.998)
Retained prior year	0.0199	0.0861	0.0708	0.0891
	(0.140)	(0.280)	(0.257)	(0.285)
Prior year incident	0.123	0.311	0.244	0.275
	(0.328)	(0.463)	(0.430)	(0.446)
FRPL eligible	0.525	0.756	0.741	0.758
	(0.499)	(0.430)	(0.438)	(0.428)
White	0.475	0.368	0.363	0.357
	(0.499)	(0.482)	(0.481)	(0.479)
Black	0.220	0.340	0.324	0.329
	(0.414)	(0.474)	(0.468)	(0.470)
Hispanic	0.243	0.241	0.261	0.263
	(0.429)	(0.428)	(0.439)	(0.440)
English not native	0.257	0.234	0.261	0.258
	(0.437)	(0.424)	(0.439)	(0.438)
Foreign born	0.0933	0.0821	0.0963	0.0921
	(0.291)	(0.274)	(0.295)	(0.289)
Special education	0.135	0.211	0.205	0.206
	(0.341)	(0.408)	(0.404)	(0.404)
LEP ineligible	0.0451	0.0661	0.0749	0.0782
	(0.208)	(0.248)	(0.263)	(0.269)
ESE ineligible	0.0134	0.0186	0.0193	0.0210
	(0.115)	(0.135)	(0.137)	(0.144)
N	7,565,799	451,447	84,709	87,148

Notes: Standard deviations are given in parentheses and brackets respectively.

Table 4 – FAY Eligibility and Current Year Student Achievement

	Linear		Quadratic	
Entry week range	4	4	4	4
All schools				
Reading score	-0.008	0.007	-0.015	0.018
	(0.012)	(0.012)	(0.024)	(0.024)
Math score	0.009	0.013	-0.014	0.002
	(0.012)	(0.012)	(0.024)	(0.025)
N	171,857	171,857	171,857	171,857
'A' schools				
Reading score	0.001	0.023	0.007	0.053
	(0.017)	(0.019)	(0.035)	(0.038)
Math score	0.015	0.040*	-0.010	0.013
	(0.018)	(0.019)	(0.035)	(0.038)
N	74,009	74,009	74,009	74,009
'D' and 'F' schools				
Reading score	-0.055	-0.005	-0.098	0.015
	(0.043)	(0.048)	(0.092)	(0.103)
Math score	-0.043	-0.037	-0.024	0.064
	(0.043)	(0.050)	(0.087)	(0.099)
N	9,699	9,699	9,699	9,699
Cohort FE	Yes	Yes	Yes	Yes
School-year FE	No	Yes	No	Yes

Notes: Robust standard errors, two-way clustered at the relative entry week and school-by-year level, are given in parentheses. Discontinuity estimates are obtained parametrically using the specified polynomial order and the score range. Columns labeled as (I) present the estimates from the base specification in equations (3) with the addition of cohort fixed-effects, whereas the columns labeled as (II) add school-by-year FE. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 5 – FAY Eligibility and Student Baseline Characteristics

	Linear		Quartic	
	(I)	(II)	(I)	(II)
Entry week range	4	4	4	4
All schools -				
Prior year reading score	-0.007	0.018	-0.015	0.032
	(0.012)	(0.012)	(0.024)	(0.025)
Prior year math score	0.011	0.027*	0.010	0.052*
	(0.012)	(0.012)	(0.024)	(0.025)
Retained prior year	0.020***	0.017***	-0.004	-0.010
	(0.005)	(0.004)	(0.009)	(0.008)
Incident prior year	0.012**	0.015**	-0.018*	-0.026*
	(0.005)	(0.005)	(0.009)	(0.010)
FRPL eligible	-0.006	-0.009	0.027*	0.014
	(0.006)	(0.005)	(0.012)	(0.010)
White	-0.003	0.005	-0.039**	-0.014
	(0.006)	(0.005)	(0.012)	(0.011)
English not native	0.003	-0.004	0.029**	0.011
	(0.005)	(0.005)	(0.011)	(0.010)
N	171,857	171,857	171,857	171,857
'A' schools-				
Prior year reading score	-0.005	0.023	-0.004	0.048
	(0.018)	(0.019)	(0.036)	(0.038)
Prior year math score	0.018	0.035	-0.002	0.042
	(0.018)	(0.019)	(0.036)	(0.039)
Retained prior year	0.018*	0.017**	-0.009	-0.009
	(0.007)	(0.006)	(0.014)	(0.011)
Incident prior year	0.010	0.011	-0.020	-0.028
	(0.006)	(0.007)	(0.012)	(0.015)
FRPL eligible	-0.011	-0.019*	0.017	0.002
	(0.009)	(0.009)	(0.019)	(0.018)
White	0.010	0.020*	-0.032	-0.007
	(0.010)	(0.009)	(0.019)	(0.018)
English not native	-0.001	-0.004	0.014	0.007
	(0.008)	(0.008)	(0.016)	(0.016)
N	74,009	74,009	74,009	74,009
'D' and 'F' schools-				
Prior year reading score	-0.071	-0.041	-0.054	0.052
	(0.043)	(0.048)	(0.090)	(0.102)
Prior year math score	0.005	0.041	0.007	0.125
	(0.042)	(0.048)	(0.086)	(0.098)
Retained prior year	0.008	0.012	-0.009	-0.046
	(0.019)	(0.017)	(0.032)	(0.033)
Incident prior year	0.046*	0.032	0.056	0.015
	(0.021)	(0.024)	(0.039)	(0.046)

(Table 5 continued)				
FRPL eligible	-0.007	-0.008	0.021	0.026
	(0.014)	(0.014)	(0.029)	(0.030)
White	-0.008	-0.0004	-0.027	0.006
	(0.017)	(0.016)	(0.032)	(0.032)
English not native	-0.016	-0.013	-0.001	0.011
	(0.021)	(0.020)	(0.041)	(0.037)
N	9,699	9,699	9,699	9,699
Cohort FE	Yes	Yes	Yes	Yes
School-year FE	No	Yes	No	Yes

Notes: Robust standard errors, two-way clustered at the relative entry week and school-by-year level, are given in parentheses. Discontinuity estimates are obtained parametrically using the specified polynomial order and the score range. Columns labeled as (I) present the estimates from the base specification in equations (3) with the addition of cohort fixed-effects, whereas the columns labeled as (II) add school-by-year FE. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 6 – FAY Eligibility and Classroom Assignments

	All schools		'A' Schools		'D' and 'F' Schools	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	4	4	4	4	4	4
Reading classrooms -						
Peer reading score average	0.004	0.007	0.001	0.001	0.003	0.001
	(0.005)	(0.011)	(0.008)	(0.017)	(0.019)	(0.039)
Peer prior reading score average	0.008	0.011	0.007	0.007	-0.004	0.004
	(0.005)	(0.011)	(0.009)	(0.017)	(0.020)	(0.044)
Teacher experience	0.014	0.200	-0.023	-0.188	-0.385	0.248
	(0.103)	(0.205)	(0.177)	(0.353)	(0.329)	(0.710)
Teacher early career	-0.001	-0.002	0.001	0.006	0.001	-0.016
	(0.004)	(0.009)	(0.007)	(0.013)	(0.019)	(0.037)
Teacher advanced degree	-0.005	-0.003	-0.005	-0.002	0.004	-0.028
	(0.005)	(0.011)	(0.009)	(0.017)	(0.022)	(0.043)
Teacher regular license	0.001	0.005	0.0001	0.005	0.010	0.039
	(0.002)	(0.005)	(0.003)	(0.007)	(0.013)	(0.026)
Teacher certified in reading	0.003	0.003	0.006	0.017	0.029	-0.007
	(0.004)	(0.008)	(0.006)	(0.012)	(0.019)	(0.033)
N	167,057	167,057	72,375	72,375	9,312	9,312
Math classrooms -						
Peer math score average	-0.004	0.004	-0.002	-0.004	-0.005	0.019
	(0.006)	(0.012)	(0.009)	(0.016)	(0.021)	(0.045)
Peer prior math score average	0.001	0.004	-0.002	0.004	-0.002	-0.001
	(0.006)	(0.012)	(0.009)	(0.017)	(0.021)	(0.045)
Teacher experience	0.113	-0.011	0.053	0.001	-0.124	0.487
	(0.113)	(0.231)	(0.193)	(0.341)	(0.340)	(0.755)
Teacher early career	-0.003	0.0037	-0.006	-0.001	-0.018	-0.025
	(0.005)	(0.010)	(0.007)	(0.017)	(0.019)	(0.042)
Teacher advanced degree	-0.002	0.011	-0.005	0.010	-0.006	0.020
	(0.006)	(0.012)	(0.009)	(0.017)	(0.023)	(0.046)
Teacher regular license	0.002	0.007	0.002	0.009	-0.007	-0.001
	(0.003)	(0.005)	(0.004)	(0.007)	(0.013)	(0.024)
Teacher certified in math	0.001	0.011	0.001	0.009	0.009	-0.001
	(0.004)	(0.008)	(0.006)	(0.013)	(0.016)	(0.035)

(Table 6 continued)

N	162,103	162,103	70,469	70,469	9,103	9,103
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
School-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Student baseline characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors, two-way clustered at the relative entry week and classroom level, are given in parentheses. Discontinuity estimates are obtained parametrically using the specified polynomial order and the score range. All models include cohort fixed-effects, school-by-year FE, and student baseline characteristics including prior year test scores, incidents, retention, FRPL eligibility, race/ethnicity, LEP and special education status, foreign born and English not native indicators. *, ** and *** represent statistical significance at 10, 5, and 1 percent respectively.

Appendix

FAY Definitions in 50 States and the District of Columbia as of 2014

State	A student is considered to be enrolled in a school for a full academic year if he/she is…
Alabama	Enrolled as of September 1 and remains enrolled as of the first day of testing.
Alaska	Enrolled continuously from October 1 through the first day of the annual test administration.
Arizona	Enrolled at the start of the school year (within the first two weeks of instruction) and presently enrolled during the first day of administration of AIMS.
Arkansas	Enrolled continuously from October 1 through and including the initial day of testing.
California	Enrolled continuously from a date in October (generally the first Wednesday) to the date of testing in the spring.
Colorado	Enrolled from one CSAP, Lectura, or CSAPA administration (annual test administration) to the next, unless the student is enrolled in the lowest grade in the school. In that case, students who have been continuously enrolled in the district and have been enrolled in the school on or before October 1st are included.
Connecticut	Enrolled as of October 1st of any school year and remains enrolled at that school up to and including the dates of the CAPT test administration in the spring of that school year.
Delaware	Enrolled continuously in the school from September 30 through May 31 of a school year.
District of Columbia	Enrolled on the official state (fall) enrollment date in October of each year and the first day of testing (typically in late April).
Florida	Enrolled and in attendance by the fall term as documented in Survey 2 conducted the second week of October and Survey 3 conducted the second week of February.
Georgia	Enrolled continuously from the Fall FTE count (which occurs on the first Tuesday in October each year) through the end of the State's Spring testing window (which occurs in March for the GHSGT and April/May for the CRCT).
Hawaii	Enrolled continuously from May 1st of one school year to May 1st of the next school year.
Idaho	Enrolled continuously from the end of the first eight (8) weeks or fifty-six (56) calendar days of the school year through the spring testing administration period.
Illinois	Enrolled on May 1 of the previous school year until state testing in the spring of that school year.
Indiana	Enrolled continuously from October 1 through and including the initial day of testing.
Iowa	Enrolled on the first day of the testing period for ITBS and ITED in the previous school year and enrolled through the academic year to the first day of the testing period for ITBS and ITED for the current school year.
Kansas	Enrolled in that school on the September 20 enrollment date of the fall preceding the spring test administration.
Kentucky	Enrolled in the school any 100 instructional days from the first instructional day of the school year through the first day of the testing window.
Louisiana	Enrolled in a school on October 1 and the test date.
Maine	Enrolled on or before October 1 in the academic year of testing through the date of testing.
Maryland	Enrolled by September 30 and attending that school through the dates of testing.
Massachusetts	Enrolled as of October 1 of any school year and remains enrolled at that school up to and including the dates of MCAS test administration in the spring of that school year.
Michigan	Enrolled in the school for the two most recent semi-annual official count days, held on the 4th Wednesday of September and 2nd Wednesday of February.

(Table continued)

Minnesota	Enrolled on October 1 of the current school year and also enrolled at the time of testing.
Mississippi	Enrolled in the same school at the end of month 6, 7 and 8 and has spent 75% of the instructional time at that school for Spring test data. Various criteria are used for fall test data and students with irregular schedules.[1]
Missouri	Enrolled the last Wednesday in September (the state's official attendance count date) and enrolled as of the MAP administration, without transferring out of the school for one more than half of the eligible days between the two dates.
Montana	Enrolled continuously from the October enrollment reporting date (first Monday in October) through the March test administration.
Nebraska	Enrolled from the last Friday in September (the official enrollment date for the State) until the end of the assessments or the end of the school year.
Nevada	Enrolled in a school on the state's official enrollment count day (the fourth Friday after the beginning of the school's academic calendar) and remain continuously enrolled in the same school up to and during each of the spring testing windows.
New Hampshire	Enrolled continuously in the school since the first business day in October of the previous school year.
New Jersey	Enrolled during the term that begins on July 1 and ends on or about June 30.
New Mexico	Enrolled from 120th day prior year to 120th day current year, for a period not to exceed 365 days.
New York	Enrolled from the first Wednesday in October until the dates of test administration.
North Carolina	Enrolled for 140 days of the first day of EOG testing (which occurs during the final three weeks of school.)
North Dakota	Enrolled at a school for a period equal to or exceeding 173 instructional days.
Ohio	Enrolled continuously from the October enrollment accounting period through the March or May test administration.
Oklahoma	Enrolled continuously beginning within the first ten days of the school year and has not experienced an enrollment lapse of ten or more consecutive days.
Oregon	Enrolled for more than half the number of instructional days in the school's calendar prior to May 1.
Pennsylvania	Enrolled from October 1 of the academic year to the close of the testing period.
Rhode Island	Enrolled in the same school from October 1 to the end of that prior school year.
South Carolina	Enrolled continuously from the time of the 45-day enrollment count until the time of testing.
South Dakota	Enrolled continuously from October 1 to the last day of the testing window.
Tennessee	Enrolled from at least one day of the first reporting period (consisting of the first 20 days of the school year and reported October 31) until test administration.
Texas	Enrolled during the Fall snapshot (typically the last Friday in October) and the spring test date.
Utah	In membership, in the same school, for not less than160 days.
Vermont	Continuously enrolled from the first day until the last.
Virginia	In membership in the school, LEA or the State by September 30 of the school year and continues in membership through test administration.
Washington	Enrolled continuously from October 1st in the current school year through the testing administration period.
West Virginia	Enrolled continuously in that school from the fifth instructional day of school to the spring testing window.
Wisconsin	Continuously enrolled since the third Friday of the September enrollment report of the previous academic year at the time of test administration.

(Table continued)

Wyoming	Enrolled on October 1 and on the first day of the official PAWS testing window.

[1] Compiled from the list of approved state accountability plans on Department of Education website accessed 03/30/2015. For more information, visit http://www2.ed.gov/admins/lead/account/stateplans03/index.html