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

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

UMUT ÖZEK

WORKING PAPER 82 • NOVEMBER 2012

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

Umut Özek
American Institutes for Research

Contents

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110242 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 using regression-discontinuity framework. I find that schools that face accountability pressure behave strategically in an attempt to boost the 'assessed' student performance, creating significant achievement gaps between eligible and ineligible mobile students. The findings also suggest that these achievement gaps are primarily driven by the strategic classification of students by failing schools to alter the eligible test-taker pool. I propose an alternative approach to mobile students in educational performance evaluations that eliminates this undesired incentive, which ironically affects students whom accountability systems specifically aspire not to leave behind.

1. Introduction

Accountability has become a mantra in public education nearly 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, intrinsic ability etc.). 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

[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.

where within-semester student turnover rates are typically higher. For instance, in Florida, one of the few states that 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 (FRL) eligible, 'late-entrants' account for approximately 15 percent of spring enrollment.[3]

This paper investigates the unintended consequences of the full academic year 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 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 2009, all 50 states and the District of Columbia had the 'full academic year' 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 allocate their resources strategically: schools that face accountability pressure might find it beneficial to focus on eligible students to boost the 'assessed' school performance so as to evade the stigma and sanctions

[3] Author's calculations from administrative student-level data for years between 2002 and 2006.
[4] More specifically, 39 states and the District of Columbia currently define full academic year in this way.

associated with being labeled as 'failing'.[5] In order to investigate this possible behavior, I utilize detailed student-level administrative data from Florida, which uses surveys conducted in October and in February to identify the full academic year eligible students under its accountability system, 'Florida's A+ Plan'.

By comparing the test performance (high-stakes and low-stakes) of 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 evidence for the existence of such strategic behavior. While I find no significant differences between eligible and ineligible students at 'safe' schools, students who enter 'near-failing' or 'failing' schools right after the eligibility cutoff perform significantly worse than the students on the other side of the cut-off especially in reading tests. Specifically, 'just-ineligible' students at 'near-failing' or 'failing' schools, on average, score 0.47σ lower in the high-stakes reading test (0.23σ in low-stakes reading) and 0.37σ lower in the high-stakes math test (0.34σ lower in the low-stakes math) than 'just-eligible' students, even though the previous year test performances and other observed characteristics of these two groups are statistically indistinguishable. These achievement gaps persist in the following year, yet I find no differences in non-cognitive outcomes such as disciplinary behavior and attendance. The findings also suggest that these gaps are primarily driven by students with entry dates right around the cutoff, providing evidence that schools might be manipulating the recorded entry dates of students in an attempt to alter the eligible test-taker pool.

This study raises an important concern with the current federal policy for incorporating mobile students into performance evaluations in accountability systems. Even though it is well-intended, the

[5] Along similar lines, recent literature has revealed other undesired school responses to accountability pressure including changing the test-taker composition or reclassifying students in an attempt to alter the composition of 'eligible' test-takers whose scores are used in the assessment of schools. 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.

full academic year requirement, as implemented in almost all states, creates undesirable incentives for failing schools. Furthermore, similar incentives will likely persevere at the teacher-level as the focus of educational accountability shifts from institutions to individual educators with policies such as Race to the Top. A straightforward solution is to replace the current assessment regime, which holds schools/educators fully responsible for the performances of some mobile students and not accountable at all for others based on their entry dates, with one that holds schools partially responsible for all mobile student depending on their 'exposure rates'. I further discuss this alternative approach in the fifth section.

2. Policy Background

2.1. NCLB and the Full Academic Year (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 full academic year. This requirement arose from Section 1111(b)(3)(C)(xi) of the federal law which prohibits states from using the test scores of full academic year (FAY) ineligible students, who have attended more than one school in any academic year, for school accountability purposes.[6]

[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 full academic year but have not attended a single school for a full academic year, except that the performance of students who have attended more than 1

Consequently, all states and the District of Columbia adopted accountability systems that define full academic year in various ways. Appendix A lists these definitions as of 2009. 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 full academic year 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). Every year between 1999 and 2008, Florida public school students in grades three through ten also took the norm-referenced Stanford-9 or Stanford-10 Achievement Tests as the FCAT-NRT, the results of which were not used for accountability purposes

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 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 2005-2006 school

---

school in the local educational agency in any academic year shall be used only in determining the progress of the local educational agency."

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 1998-1999 and 2001-2002 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 2001-2002 school year, the grading formula under the A+ Plan went through a major revision.[7] 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.[8]

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 account for 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.
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.

[7] Detailed information about the new grading formula, as it was used in the 2008-2009 school year, is provided in Appendix B.
[8] 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.

iii. Full Academic Year (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.[9]

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, 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 will be the key element of my identification strategy.

3. Data Description and Empirical Strategy

[9] For K-12 public schools in Florida, instruction typically begins in August and ends in May.

3.1. Data Description

In the analyses that follow, I utilize student-level administrative data on all elementary students between grades three and five from 2002-2003 to 2005-2006 in Florida. The dataset includes demographic information on students such as race, gender, FRL-eligibility, LEP status, LEP program entry and exit dates, ESE status at the time of each survey as well as reading and math scores in both FCAT-SSS and FCAT-NRT. The most critical piece of information contained in the dataset for the purposes of this study is the exact 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 obtained the accountability grades for all public schools in Florida between 2002-2003 and 2005-2006 school years. Table 2 presents the grade distribution of the subset of public schools that are utilized in the analysis.[10]

Over the four year period, approximately 97 percent of all elementary students in grades three through five took the FCAT-SSS and FCAT-NRT in both reading and math, leaving me with 2,147,639 student-year observations. Ninety percent of the elementary school test-takers entered the school in which they were tested in the first week of the school year. Of the remaining 223,105 'late-entrants', 63,756 students (29%) entered the elementary school at which they were tested during the two-month window around the October eligibility cutoff. The majority of the latter group (75%) was received from another public school in or out of the school district whereas the remaining students were either attending private schools (13%) or another attendance taking unit within the same school (e.g. a magnet program) before entering the school at which they were tested.

While only a small fraction of students in Florida during that time period were FAY-ineligible (6.5%), they constitute the primary target of NCLB and all other state accountability systems. As shown in Table 3, ineligible elementary students had significantly lower prior year proficiency rates compared

[10] This subset contains all public schools that served at least one of the tested elementary grades (3rd grade, 4th grade or 5th grade) during the time period examined in the study.

to eligible elementary students in reading (51% versus 67%) and math (47% versus 64%); were more likely to be FRL-eligible (71% versus 52%) and were more likely to belong to a racial/ethnic minority group (61% versus 51%).[11] These gaps widen even further when comparing all students with ineligible students in near-failing and failing elementary schools, who are expected to be most adversely affected by the eligibility requirement. Only one-third of the ineligible students in 'D' and 'F' schools had performed at or above the proficiency levels in reading and math in the previous school year and roughly 90 percent of these students were FRL-eligible and/or non-white.

3.2. Empirical Framework

In order to estimate the causal impact of FAY-eligibility on student achievement, I rely on regression-discontinuity (RD) design. Let S_{it} denote the number of school days between the entry date of student *i* to the school she was tested and the October cutoff date in year *t*, with negative values indicating entry before the cutoff. Defining treatment, T_{it}, as being FAY-ineligible and combining observations over time for a given student, 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:

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

I estimate (2) non-parametrically using kernel-weighted local polynomial smoothing initially proposed by Hahn et al. (2001) and later developed by Porter (2003) to include higher-order polynomial

[11] More information on the proficiency thresholds used under the A+ Plan is provided in Appendix B.

estimators.[12] This method has been shown to reduce the misspecification bias possibility of parametric models and achieve the optimal rate of convergence. The estimation technique is essentially equivalent to estimating a polynomial regression using the kernel-weights ($w_i = K(S_i/h)$) with the observations to the right and left of the October eligibility cutoff date and then calculating the treatment effect using the left limits and right limits of these regressions at the cutoff. I prefer the triangle kernel in the estimation, since it has been shown to be boundary optimal (Cheng et al., 1997).

The critical decision left to the researcher in this context is the choice of bandwidth parameter, *h*, since increasing bandwidth is expected to produce biased estimates especially in situations where the selection variable (S_i) is correlated with the outcome (Y_i) conditional on treatment status (T_i). This is likely to be the case in this context, since the entry date of a student is directly related to her instructional exposure at the school she was tested, which is expected to impact the ultimate test score. In order to minimize this concern, I choose a rather narrow bandwidth of 5 school days and also report the results for alternative bandwidths of 2 and 10 school days. Throughout the rest of the non-parametric analysis, I implicitly assume that the exposure rates of students on either side of the cutoff do not differ significantly within these bandwidths.[13]

An important concern in the non-parametric approach in this context is the discrete nature of the selection variable. As described in Card and Lee (2008), students who enter right before the eligibility cutoff might not provide a good counterfactual for those just above as it is not feasible to compare averages within arbitrarily small neighborhoods around the cutoff. To alleviate this concern

[12] For this purpose, I used the Stata command 'rd' by Nichols (2007). In the remainder of the paper, I report estimated treatment effects that were obtained using local linear smoothing, yet the results are robust to specifications with higher-order polynomials.

[13] For the sake of computational feasibility, I restrict the sample to students with $-20 \leq S_i < 20$ in the non-parametric analysis. Given the bandwidth values chosen in the estimation, this restriction does not have any impact on the estimated treatment effect, since the excluded observations would receive zero weights in calculating the discontinuity at the cutoff even with the highest preferred bandwidth. This restriction practically converts the dataset into a repeated cross-section of individual students and thus justifies the notation used in (1), since only 3 percent of students are observed multiple times in the restricted sample.

and check the robustness of the non-parametric estimates, following Card and Lee (2008), I also estimate (2) parametrically using the following regression 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 day. I use linear, quadratic, cubic and quartic functions to check the robustness of the findings to different specifications, and two-way cluster the standard errors at the school and relative entry day level as described in Cameron et al. (2006).

4. Results

4.1. Full Academic Year Eligibility and Student Outcomes

Figures 1A and 1B present a graphical inspection of the effects of eligibility on student achievement. The four panels in these two figures present local linear smoothing of the standardized FCAT-SSS and FCAT-NRT scores (Y_i) in reading and math on relative entry dates of students with respect to the October cutoff (S_i) for 'safe' elementary schools that received 'A', 'B' or 'C' the previous summer as well as 'near-failing' ('D') and 'failing' ('F') elementary schools. The triangle kernel and a bandwidth of 5 school days is used in the estimation, and the vertical lines cutting smoothed lines on each graph represent 95% confidence intervals. The four panels in Figure 1 provide striking evidence of strategic behavior among schools facing accountability pressure. While no apparent significant difference is observed between the high-stakes test performances of 'just-eligible' and 'just-ineligible' students for safe elementary schools (panels (A) and (C)), students who enter 'D' and 'F' schools, which face the highest accountability pressure under the A+ Plan, just after the cutoff perform significantly worse in both reading and math than those who enter just-before the October cutoff (panels (B) and (D)).

In order to assess the magnitude of this achievement gap, Table 4 gives the estimates of β in (2) where the outcome is the standardized current year test scores. The findings, which are obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten school days, reinforce the evidence presented in Figures 1A and 1B. Using the preferred bandwidth of five school days, students who enter a 'D' or 'F' school just after the cutoff perform 0.47σ worse in high-stakes reading (0.23σ in low-stakes reading) and 0.37σ worse in high-stakes math (0.34σ in low-stakes reading) tests compared to their just-eligible counterparts. The impact estimates are considerably large by educational standards, ranging from 0.47σ to 0.32σ for high-stakes reading (0.24σ to 0.19σ for low-stakes reading) and 0.39σ to 0.18σ for math (0.36σ to 0.21σ for low-stakes math) depending on the chosen bandwidth and almost all of them are statistically significant at conventional levels. Using the mid-point bandwidth of 5 school days, these differences correspond to 80 percent of the control mean, i.e. the left-hand-side prediction of the non-parametric regression at the cutoff, of -0.58 in reading and 70 percent of the control mean of -0.52 in math. Furthermore, the results present no consistent evidence of strategic behavior for 'safe' schools that received 'A', 'B' or 'C'. The latter finding provides moderate evidence that accountability pressure the school faces is associated with the observed achievement gap between eligible and ineligible students.

The first panel in Table 5 presents the parametric estimates. For this exercise, I restrict the sample to students with non-missing prior test scores (i.e. 4th and 5th graders in the last three years). For better causal inference, I also limit the sample to students with $-20 \leq S_i < 20$ in specifications where $k(S_i)$ is linear or quadratic. Reading results are comparable to the non-parametric analysis with discontinuity estimates ranging from -0.28σ to -0.44σ for the high-stakes test and from -0.21σ to -0.36σ for the low-stakes test. However, math performances of eligible and ineligible students are only statistically distinguishable in the quartic specification. This is similar to the non-parametric estimates

obtained using only the students with non-missing prior year test scores.[14] Once again, no consistent achievement differences emerge between these two types of students at 'safe' schools.

The first panel of Table 6 compares eligible and ineligible students along other outcomes in the current year. Non-parametric estimates reveal that ineligibility does not lead to significantly higher rates of disciplinary incidents, suspensions, absences or grade retention. The sole exception is the absence rate for the ineligible students at safe schools, who have slightly higher attendance rates. The second panel gives the discontinuity estimates for student outcomes in the following year. The eligible students who had attended 'D' and 'F' schools in the first year still outperform their just-ineligible peers in the following year as further illustrated in Figure 2. The gaps are slightly narrower for the high-stakes reading test, comparable for the high-stakes math test, yet significantly higher for the low-stakes reading and math tests. The results also suggest that the just-ineligible 3rd and 4th graders in 'D' and 'F' schools are slightly less likely to stay in the same school the following year, but the estimates are statistically insignificant at conventional levels. Similar to the first year, ineligibility is not associated with higher probabilities of disciplinary problems.

4.2. Disentangling the Mechanisms behind the Achievement Gap

4.2.1 Differences in Student Attributes

The first explanation behind the achievement gaps is the differences in student attributes (e.g. prior achievement, demographics, family characteristics, and other observed and unobserved traits) between eligible and ineligible students. Possible differences in these attributes might hint at several scenarios. For instance, 'educationally motivated' parents might be strategically manipulating the entry dates of their children anticipating the adverse effects of being FAY-ineligible, leading to the differences in student attributes. Similarly, schools might be manipulating the entry dates of low-performing students to change the composition of 'eligible' test-takers whose scores are used in the assessment of

[14] These findings, not reported here for brevity, are available upon request.

schools, similar to the evidence presented in Cullen and Reback (2006), Figlio and Getzler (2006) and Jacob (2005) in the special education context.

Obviously, differences in prior achievement levels might explain the differences between reading and math performances of just-eligible and ineligible students. In other words, even if schools do not engage in strategic behavior, one might observe differences in test scores purely due to the fact that those who are just-ineligible have different 'starting points' than the students to the left of the cutoff. Likewise, different starting points might also lead to false conclusions of no strategic behavior in this context. Figure 3A and 3B present evidence rejecting this possibility by replicating the analysis reported in Figure 1A and 1B using prior year test scores, which seem to be continuous around the cutoff for both samples. The first four rows of Table 7 report the non-parametric discontinuity estimates of this graphical analysis, reaching the same conclusion.

I also check the continuity of other student characteristics around the cutoff including prior year disciplinary incidents, suspensions, ineligibility, and grade promotion (rows 5-8 in the first panel); disciplinary incidents and suspensions at the school the student withdrew from (second panel); and student attributes in the current year including FRL-eligibility, limited English proficiency status, special education status, ESE and LEP eligibility, whether the student is on track (has never been retained), whether she was born in the United States, whether English is the language spoken at home, gender, and race/ethnicity (third panel). The findings reveal no statistically significant difference between just-eligible and just-ineligible students who entered 'near-failing' or 'failing' schools. At the safe schools, just-ineligible students are more likely to have been retained in the previous year and born in the U.S., more likely to be black, and less likely to be ESE ineligible.

To further examine whether differences in student attributes explain the achievement gaps, the second panel in Table 5 presents the parametric estimates controlling for observed student attributes listed in Table 7 as well as indicators for where the student came from (e.g. another public school within

the district, private school, another attendance taking unit at the same school), entry day of the week indicators, and grade and year indicators. The inclusion of these covariates does not seem to alter the conclusions, while it does reduce the magnitude of the estimates considerably in less conservative specifications that do not impose entry day restrictions. For instance, the high-stakes reading discontinuity estimated using the cubic specification decreases from 0.37σ to 0.21σ (from 0.31σ to 0.18σ for low-stakes reading) when controlling for student covariates.

While strongly suggestive, the evidence presented above is not sufficient to rule out the possible discontinuity of unobserved student characteristics at the cutoff. For instance, one might claim that those who enter failing schools at the end of a week are different than those who enter at the beginning along unobserved dimensions, which, in turn, is responsible for the achievement gaps at the cutoff. To test this possibility, I construct four 'pseudo' cutoffs (two weeks before/after, one week before/after) away from the actual October cutoff dates and check for discontinuities in test scores. Table 8 suggests that, away from the actual cutoff, students who with entry dates at the beginning of a week either outperform or perform similarly to those with entry dates at the end of a school week. This is in contrast to the discontinuity at the actual cutoff.

Another symptom of unobserved differences between eligible and ineligible students is an unusual increase in the number of entrants in the day and/or week prior to the cutoff, as noted in McCrary (2008). The two panels in Figure 4 give the number of entrants to 'D' and 'F' schools in raw form and in deseasonalized form (i.e. removing the weekly cyclicality apparent in the first panel) between a week after the beginning of the school year and before the winter break in Florida. The findings provide evidence against this selection possibility as the discontinuity in the number of entrants at the cutoff is statistically indistinguishable from the average Monday-Friday discontinuity in the time frame examined (p-value of 0.963). Further, the number of entering students during the survey week (599) is comparable to the number of entrants the week after (580).

An interesting detail about the non-parametric discontinuity estimates reported in Table 4 is the significant decline in the estimated achievement gaps at 'D' and 'F' schools as we move from the bandwidth of 5 school days to 10. In high-stakes reading, for instance, the discontinuity drops from 0.47σ to 0.25σ, suggesting that the achievement differences are driven primarily by differences between students who enter during the week right before and right after the cutoff. The upper panel in Figure 5 presents the average raw reading and math scores by relative entry week around the cutoff, providing striking evidence supporting the latter statement. Average student achievement, except for high-stakes math, gradually increases during the weeks leading to the October enrollment survey, drops dramatically during the week after the cutoff, and then reverts back to the levels before the survey week. In high-stakes reading, for which this pattern is most apparent (and for which the largest achievement gap is observed), average achievement increases from -0.7σ to -0.6σ during the three weeks leading to the eligibility cutoff, plummets to -0.8σ during the week right after the survey, and then returns back to -0.7σ in a week. If the survey week and the week right after are excluded, the average high-stakes reading score for eligible students around the cutoff is almost identical to that of the ineligible students (-0.723σ versus -0.716σ).

The lower panel repeats the same exercise replacing the raw scores with test scores that are regression-adjusted by the student covariates listed in Table 5 along with indicators for where the student came from, and grade and year fixed-effects. A similar, yet even more noticeable, pattern is evident in reading scores, especially for high-stakes reading. Students with entry dates during the survey week perform 0.08σ better in high-stakes reading than predicted by their background characteristics whereas students right above the cutoff, on average, perform 0.11σ worse than expected. There is a somewhat downward trend in residualized math scores, yet no unusual drop exists between the two weeks around the eligibility cutoff. Table 6 repeats the same exercise using schools that received 'A', 'B' or 'C' in the previous year and finds no sizeable changes in raw or adjusted student performance by

week of entry. All these findings present evidence that schools when facing accountability pressure might be strategically manipulating the recorded entry dates of students around the cutoff to alter the eligible test-taker pool. Further, the observed student attributes such as prior achievement fail to capture the dimensions along which such manipulation takes place.

4.2.2. Differences in School Effects

Second, even if the eligible and ineligible students are comparable around the cutoff along individual attributes, ineligible students might perform worse if they attend lower value-added schools. This might take place, for instance, if the vacant seats at higher value-added schools fill up faster than those at other schools. While the analysis presented above is conducted separately by school grade (and thus controls for the heterogeneity in school effects to some extent), it is still plausible that ineligible students might be experiencing different school effects than eligible students. I investigate this possibility in three ways. First, comparisons along observed school characteristics reported in Table 9 reveal that just-eligible and just-ineligible students are attending similar schools. Second, parametric estimates controlling for school covariates (in addition to the student attributes) provide evidence that it is not the differences in school effects that is driving the achievement gaps: discontinuity estimates presented in the second panel of Table 5 are almost identical to the third panel.

Finally, to better control for the differences between schools, I incorporate school fixed-effects in a 'pseudo-RD' framework. The biggest challenge in this exercise is that the sample size around the cutoff is not sufficiently large to make 'within-school' comparisons using the parametric framework outlined above. Therefore, I combine students who enter during the same week and use relative entry week as the selection variable. Thus, each school in the estimation sample has at least one student in each cell. However, it is important to note that, in this approach, the identification assumption (i.e. that the students who enter during the weeks before the cutoff are comparable to those who enter after) is significantly stronger. To account for the possible differences between students, I include the set of

student covariates listed above in the regression. As suggested by Card and Lee (2008), the standard errors are clustered at the entry week level. The first panel in Table 10 presents discontinuity estimates without school fixed-effects. The findings are similar to those presented in the second panel of Table 5 with the exception of high-stakes math for which we observe significant gaps at 'D' and 'F' schools. Further, including school fixed-effects does not seem to change the conclusions: students who enter 'D' and 'F' schools during the week(s) after the cutoff perform significantly worse than their peers who enter the same school during the weeks before. The findings reveal no consistently significant achievement gaps for safe schools with the exception of low-stakes reading.

4.2.3. Differences in Classroom Effects

Finally, schools might strategically assign students to classrooms based on their eligibility status whereby eligible students are assigned to different classrooms with possibly more 'effective' teachers and/or more 'accomplished' peers, leading to differences in 'classroom-effects' between eligible and ineligible students. If this is indeed the case, one would not only expect to find achievement gaps between eligible and ineligible students, but also expect to see differences in the performances of other students in their classrooms. For this exercise, I first select the reading and math classrooms of students. I then identify their primary reading and math teachers, with whom students spend at least fifty percent of their instruction time in that subject per week, and drop classrooms with fewer than 10 and more than 40 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 non-parametrically estimate the differences between the 'out-of-sample' average classroom performances of eligible and ineligible students. That is, for each bandwidth (2, 5 and 10 school days), I exclude the students that receive a non-zero weight in the estimation to calculate the average classroom performance so that the possible performance gaps between classrooms are not driven by the differences in performance between just-eligible and just-

ineligible students.[15] In the estimation, each observation is weighted by the number of out-of-sample students in the classroom and the standard errors are clustered at the classroom level.

The estimates presented in Table 11 reveal significant differences between the peer performances of eligible and ineligible students, especially between students who enter during the week before and after the eligibility cutoff. Specifically, using the bandwidth of two school days, the classroom peers of just-eligible students in 'D' and 'F' schools perform 0.15σ better in high-stakes reading and math, 0.18σ better in low-stakes reading and 0.14σ better in low-stakes math. These findings provide evidence that just-ineligible students are in different classrooms than their eligible counterparts, which might explain some of the achievement differences right around the cutoff. Also important to note, however, is that these discontinuities seem to dissipate using the bandwidth of 10 school days. This contradicts the scenario under which schools systematically assign ineligible students to different classrooms (regardless of their entry date), because, in that case, one would also expect to find significant differences between classroom performances of eligible and ineligible students away from the cutoff as well. The estimates also reveal no significant discontinuities at safe schools.

Another classroom-related explanation to the observed achievement gaps, while not directly testable in this context, is differential teacher value-added within the classroom based on student eligibility. That is, all else constant, achievement differences between eligible and ineligible students might arise if teachers allocate their efforts strategically and focus more on the eligible students in the classroom. This is analogous to the commonly documented case of 'educational triage' where teachers have been shown to focus on students just below the proficiency threshold when facing accountability pressure (Chakrabarti (2006), Krieg (2008), Neal and Schanzenbach (2010)).

[15] This statement is only true if there are no spillover effects of the students in the estimation sample on their peers in the classroom.

4.3. Accountability Pressure versus Unobserved School Traits

Can accountability pressure explain the achievement gaps observed in near-failing and failing schools or is there an underlying factor that is simultaneously causing some schools to fail and creating these differences? In order to address this question, I restrict the sample to schools with accountability scores between 300 and 340, twenty points below and above the C-D cutoff in the years examined. The assumption here is that the near-failing and safe schools within this band are comparable along observed and unobserved attributes, and any differences between ineligible and eligible students are caused by differences in accountability pressure faced by the schools. Since the sample size is significantly reduced in this exercise, I estimate the discontinuities parametrically. Table 12 presents the findings where the first two columns report the estimates from the base specification in (3), the third and fourth incorporates student covariates, and grade and year indicators, and the last two columns add the school covariates. The findings are similar in this restricted sample. There are significant discontinuities in reading at just-failing schools, yet no consistent gaps at just-safe schools with the exception of a few significant differences in high-stakes math for several specifications.

I also check to see whether the entry date manipulation evidenced above is driven by the accountability pressure. Figure 6 repeats the exercise in Figure 5 using regression-adjusted test scores for just-failing and just-safe schools. The findings, albeit noisier due to smaller sample size, indicate that accountability pressure is likely responsible for the strategic behavior evidenced in failing schools. The unusual drop in reading scores at the cutoff still exists for just-failing schools, yet no such discontinuity is apparent for just-safe schools. There is no such pattern in math for either type of schools except for the moderate drop in high-stakes math scores at just-safe schools.

5. An Alternative Approach to Mobile Students in Accountability Systems

The main reason behind the strategic behavior evidenced in the previous section is the use of dichotomous full academic year eligibility requirements where the test scores of students who arrive at a school 'a day too late' are excluded from school assessments. These requirements are intended to avoid situations where low-performing schools with high within-semester student turnover are implicitly punished by holding them fully responsible for the performances of new students who have not spent 'enough' time at the school before standardized testing. However, such distinctions between students can easily be detected by schools, providing failing schools incentives to reallocate their resources and focus on students whose performances count for school accountability purposes.

Alternatively, one can arguably attain the same objective by holding schools partially responsible for the performances of mobile students rather than using a binary eligibility indicator. How much each student's performance contributes to the school's evaluation can be determined by the student's 'exposure rate', which is defined as the ratio of the number of school days the student was in membership at the school prior to testing to the number of available school days prior to testing. In this way, the incentive for strategic behavior no longer exists, since all tested students will be assigned a non-zero weight in school assessment calculations provided that they satisfy other eligibility requirements. This alternative approach also carries significant implications for the design of future policy practices as educational accountability begins to target individual educators, since the aforementioned undesired incentives will likely present themselves at the classroom level if teacher evaluation mechanisms treat mobile students similarly as under the existing school accountability systems.

6. Concluding Remarks

The No Child Left Behind Act of 2001 mandates states to implement a consistent definition of 'full academic year' 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 2009, all states had adopted the full academic year requirement whereby some students are typically labeled ineligible based on their entry dates to the school and a predetermined eligibility cutoff. This requirement provides schools the undesirable incentive to behave strategically to boost the assessed student performance and thus evade the sanctions imposed by accountability systems.

This study investigates the existence of such strategic behavior and the impact on student achievement levels using detailed administrative data from Florida. The analyses reveal striking evidence of such behavior, suggesting that schools that face accountability pressure manipulate the recorded entry dates of students, creating significant achievement gaps between just-eligible and just-ineligible students, who are shown to be otherwise similar. For the students who remain in the public school system in the second year, these gap narrows considerably, yet still exists.

I propose an alternative approach to mobile students in accountability systems in which schools are partially held responsible for the performances of all mobile students depending on their 'exposure rates', rather than the current method in which schools are held fully responsible for some of these students and not accountable at all for the performances of others. This alternative, which can also be applied to teacher performance evaluation mechanisms, removes the unintended incentive created by the current system, while accomplishing the underlying objective of the full academic year requirements.

References

Cameron, C.A., Gelbach J., Miller D., forthcoming. Robust Inference with Multi-way Clustering. Journal of Business and Economic Statistics.

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

Chakrabarti, Rajashri, 2006. Vouchers, Public School Response and the Role of Incentives: Evidence from Florida. Working Paper, Federal Reserve Bank of New York.

Cheng, Ming-Yen, Jianqing Fan, and JS Marron, 1997. On Automatic Boundary Corrections. Annals of Statistics 25, 1691-1708
.

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.

Jacob, Brian and Steven Levitt, 2003. Rotten Apples: An Investigation of the Prevalance and Predictors of Teacher Cheating. Quarterly Journal of Economics 118(3), 843-877.

Krieg, John, 2008. "Are Students Left Behind? The Distributional Effects of the No Child Left Behind Act." Educational Finance and Policy 3(2), 250-281.

McCrary, Justin, 2008. Testing for Manipulation of the Running Variable in the Regression Discontinuity Design. Journal of Econometrics 142(2), 698-714.

Neal, Derek and Diane Whitmore Schanzenbach, 2010. Left Behind by Design: Proficiency Counts and Test-Based Accountability. Review of Economics and Statistics 92(2), 263-283.

Nichols, Austin. 2007. rd: Stata module for regression discontinuity estimation. http://ideas.repec.org/c/boc/bocode/s456888.html

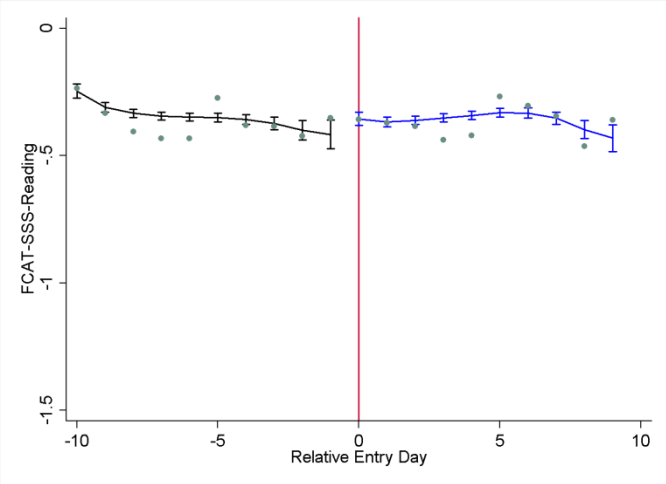
Porter, Jack, 2003. Estimation in the Regression Discontinuity Model.", Unpublished manuscript, Harvard University.

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

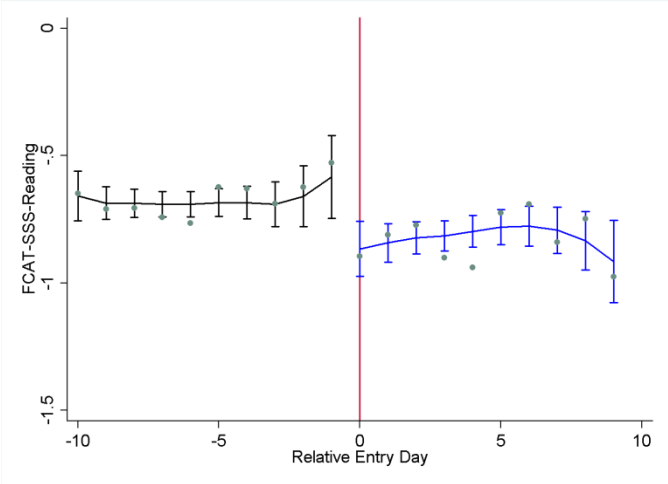
Figures

Figure 1A - FAY-Eligibility and FCAT-SSS Scores

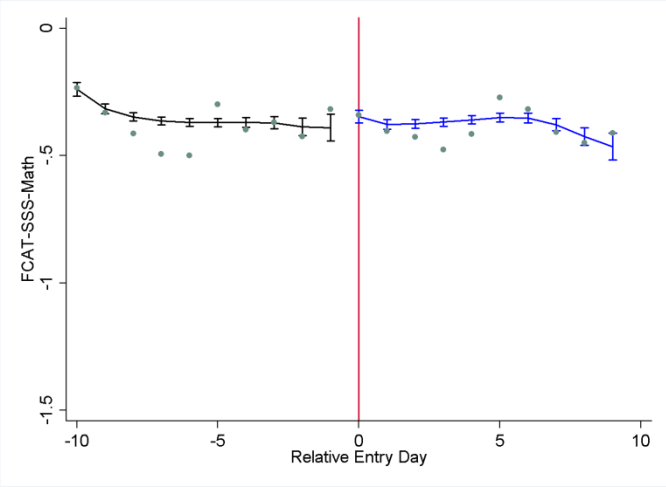
A. ‘A’, ‘B’ and ‘C’ Schools: Reading



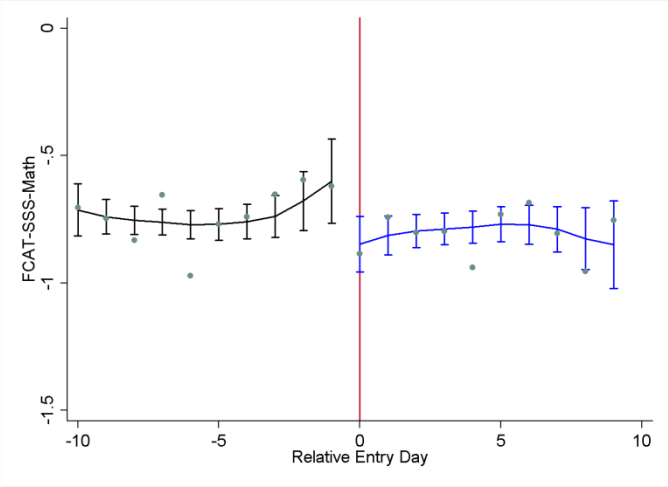
B. ‘D’ and ‘F’ Elementary Schools: Reading



C. ‘A’, ‘B’ and ‘C’ Schools: Math



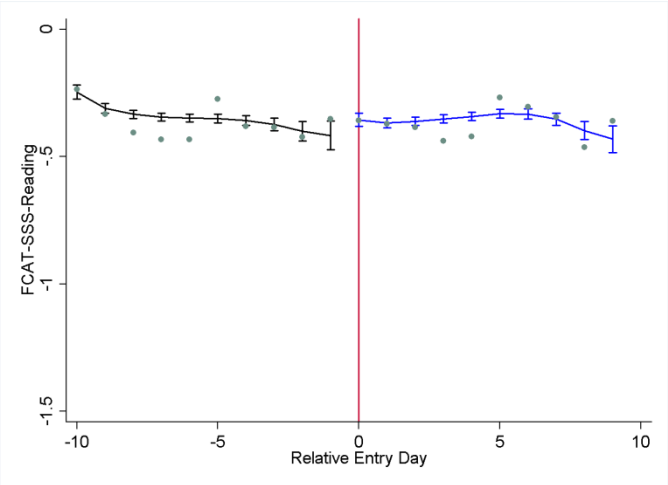
D. ‘D’ and ‘F’ Elementary Schools: Math



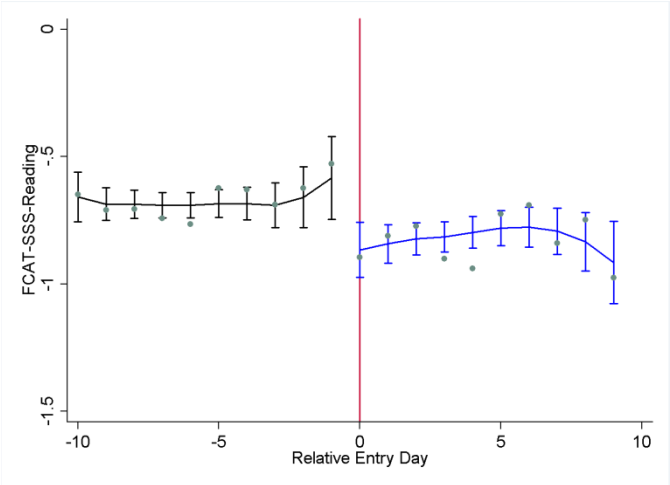
Notes: The four panels present the local linear smoothing of the current year standardized FCAT-SSS scores in reading and math on relative entry date of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 5 school days are used in the estimation. The vertical lines cutting smoothed lines on each graph represent 95% confidence intervals and the solid circles represent raw cell means.

Figure 1B - FAY-Eligibility and FCAT-NRT Scores

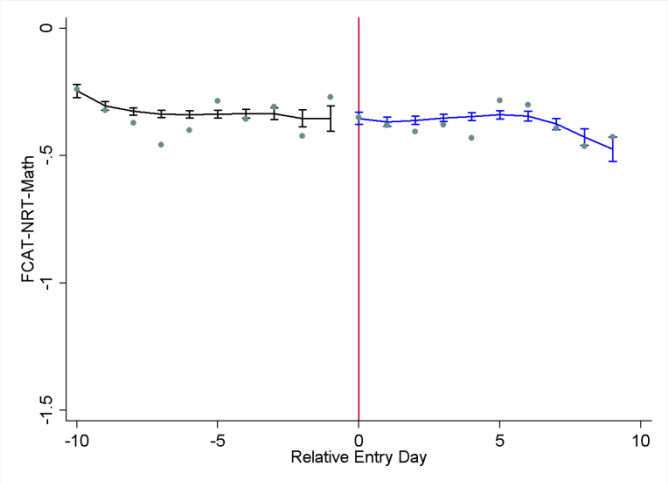
A. ‘A’, ‘B’ and ‘C’ Schools: Reading



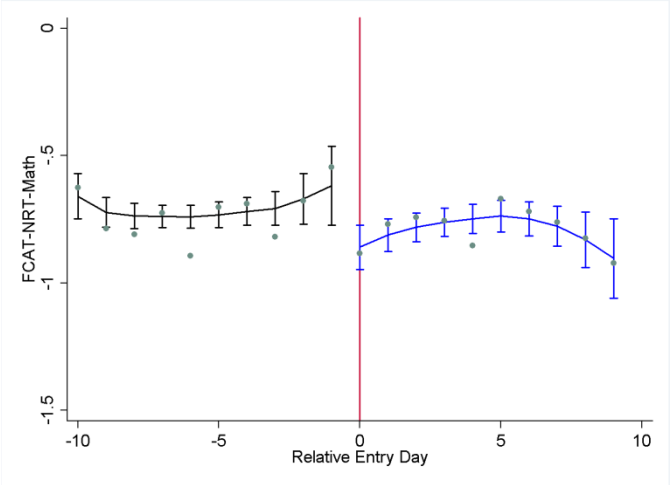
B. ‘D’ and ‘F’ Elementary Schools: Reading



C. ‘A’, ‘B’ and ‘C’ Schools: Math



D. ‘D’ and ‘F’ Elementary Schools: Math

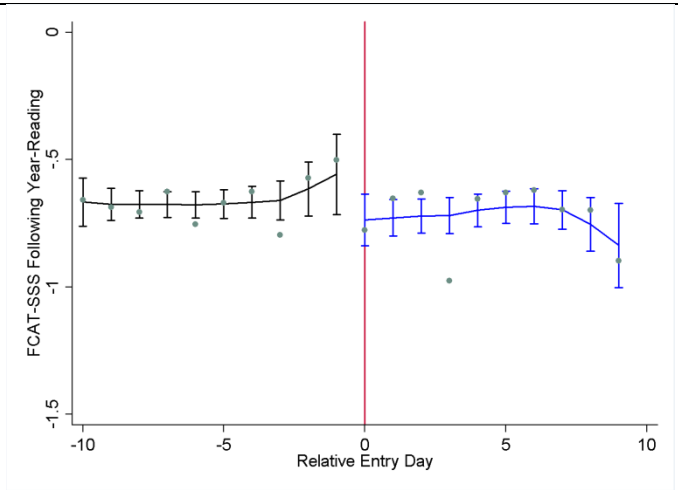


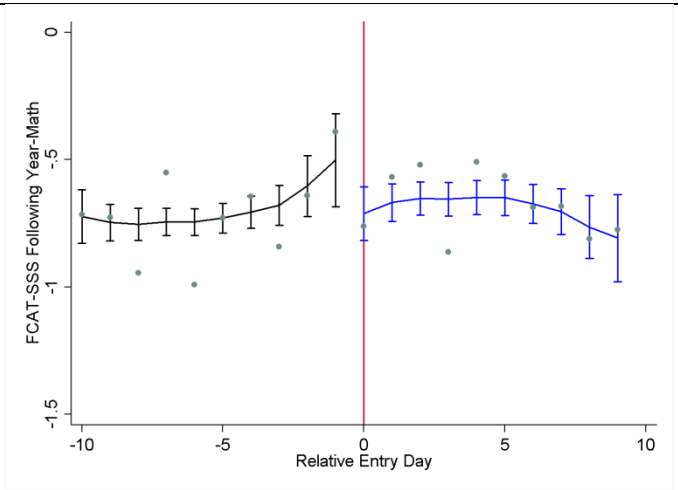
Notes: The four panels present the local linear smoothing of the current year standardized FCAT-NRT scores in reading and math on relative entry date of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 5 school days are used in the estimation. The vertical lines cutting smoothed lines on each graph represent 95% confidence intervals and the solid circles represent raw cell means.

Figure 2 - FAY-Eligibility and Following Year Student Achievement: ‘D’ and ‘F’ Schools

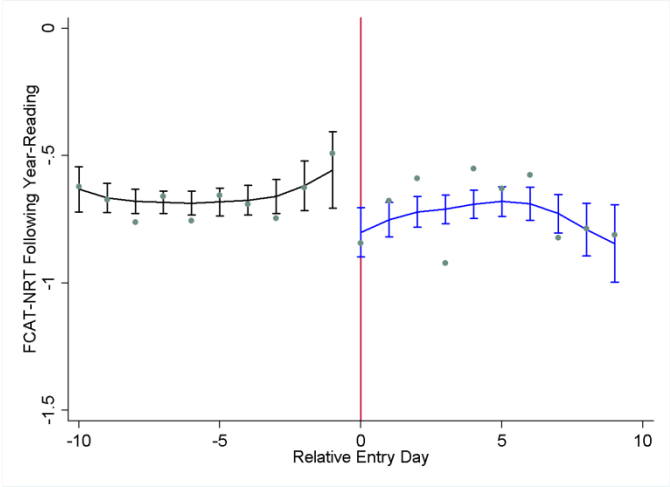
A. FCAT-SSS Reading

B. FCAT-SSS Math

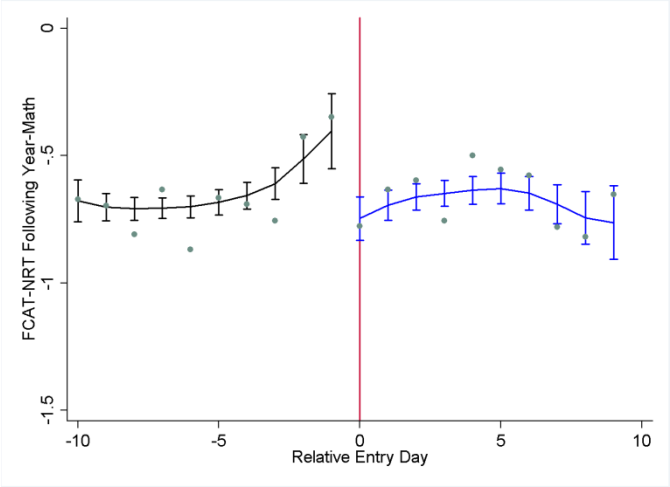




C. FCAT-NRT Reading



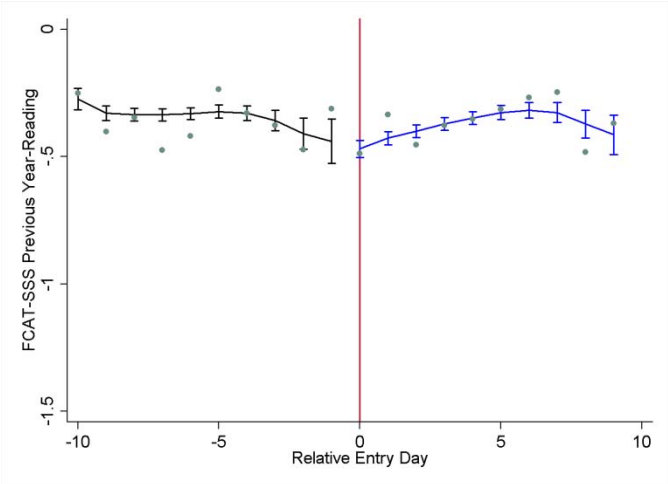
D. FCAT-NRT Math



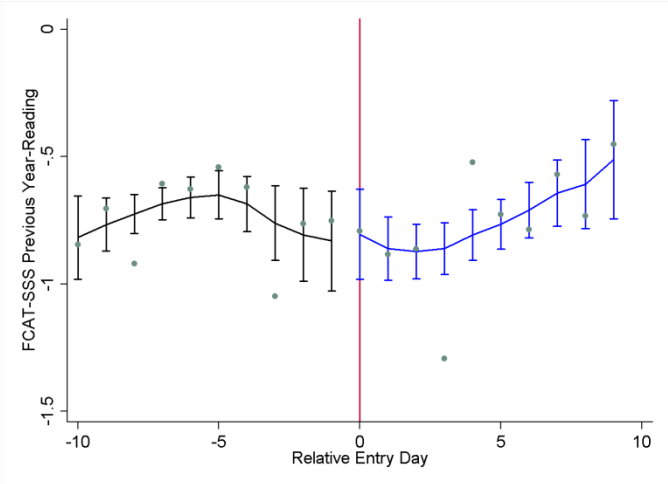
Notes: The four panels present the local linear smoothing of the following year standardized FCAT-SSS and FCAT-NRT scores in reading and math on relative entry date of the students at near-failing or failing schools separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 5 school days are used in the estimation. The vertical lines cutting smoothed lines on each graph represent 95% confidence intervals and the solid circles represent raw cell means.

Figure 3A - FAY-Eligibility and Prior Year FCAT-SSS Scores

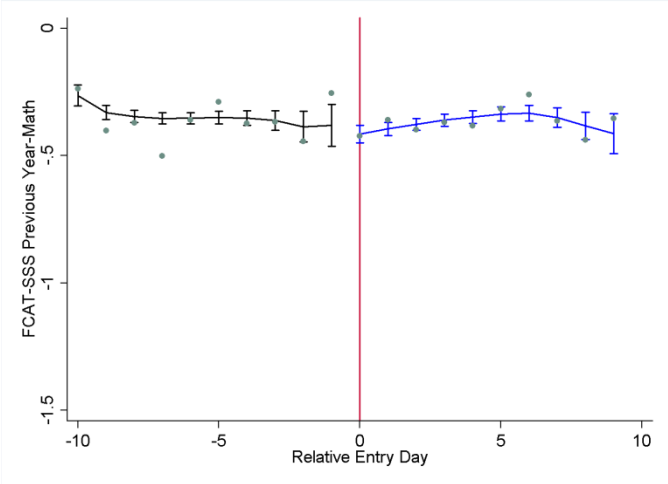
A. ‘A’, ‘B’ and ‘C’ Schools: Reading



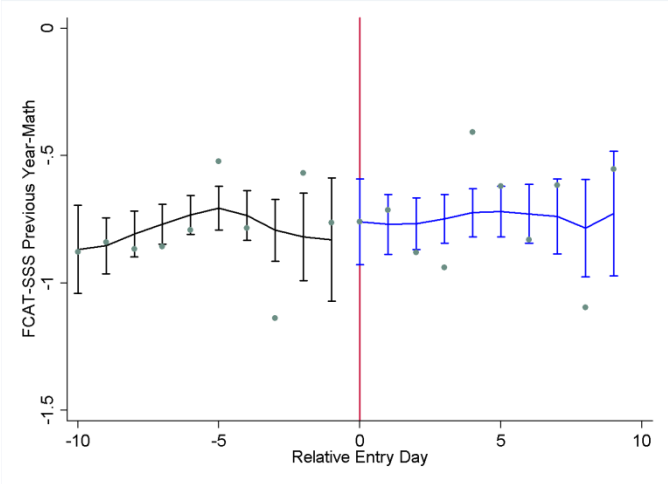
B. ‘D’ and ‘F’ Elementary Schools: Reading



C. ‘A’, ‘B’ and ‘C’ Schools: Math



D. ‘D’ and ‘F’ Elementary Schools: Math

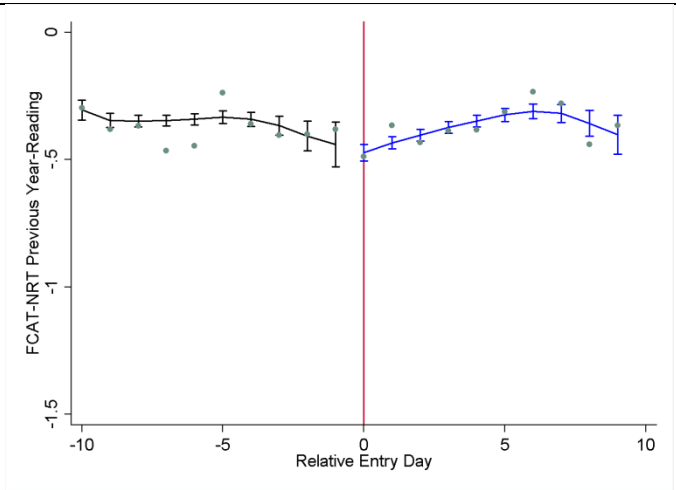


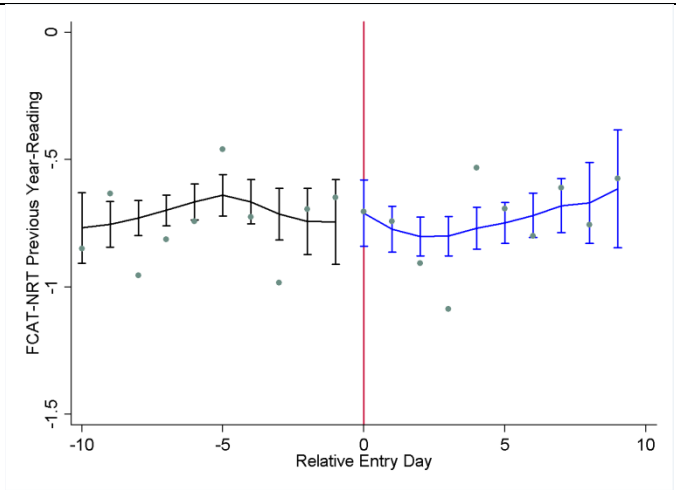
Notes: The four panels present the local linear smoothing of the previous year standardized FCAT-SSS scores in reading and math on relative entry date of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 5 school days are used in the estimation. The vertical lines cutting smoothed lines on each graph represent 95% confidence intervals and the solid circles represent raw cell means.

Figure 3B - FAY-Eligibility and Prior Year FCAT-NRT Scores

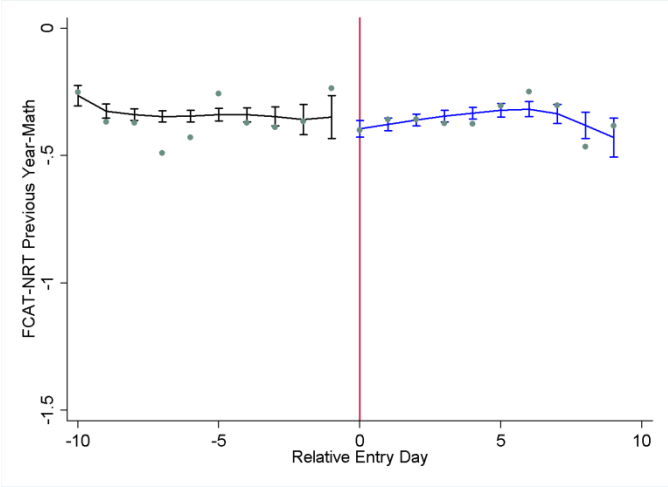
A. ‘A’, ‘B’ and ‘C’ Schools: Reading

B. ‘D’ and ‘F’ Elementary Schools: Reading

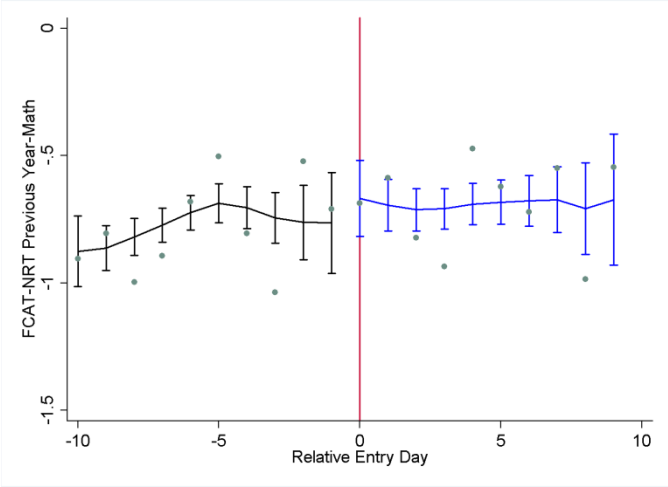




C. ‘A’, ‘B’ and ‘C’ Schools: Math



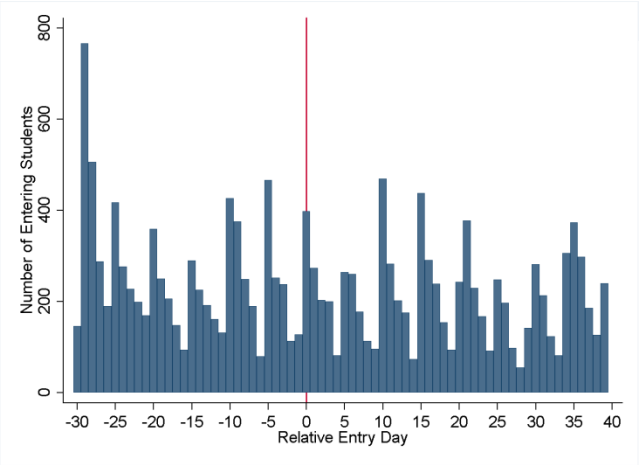
D. ‘D’ and ‘F’ Elementary Schools: Math



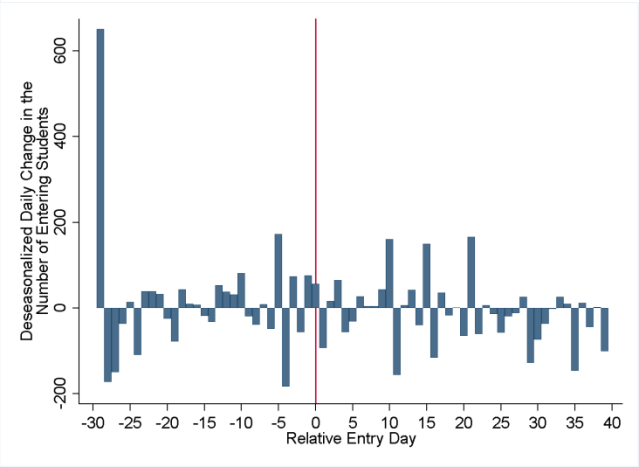
Notes: The four panels present the local linear smoothing of the previous year standardized FCAT-NRT scores in reading and math on relative entry date of the student separately for the left of the cutoff date and the right. The triangle kernel and a bandwidth of 5 school days are used in the estimation. The vertical lines cutting smoothed lines on each graph represent 95% confidence intervals and the solid circles represent raw cell means.

Figure 4 - Selection into/out of Treatment: 'D' and 'F' Schools

A. Number of Entering Students

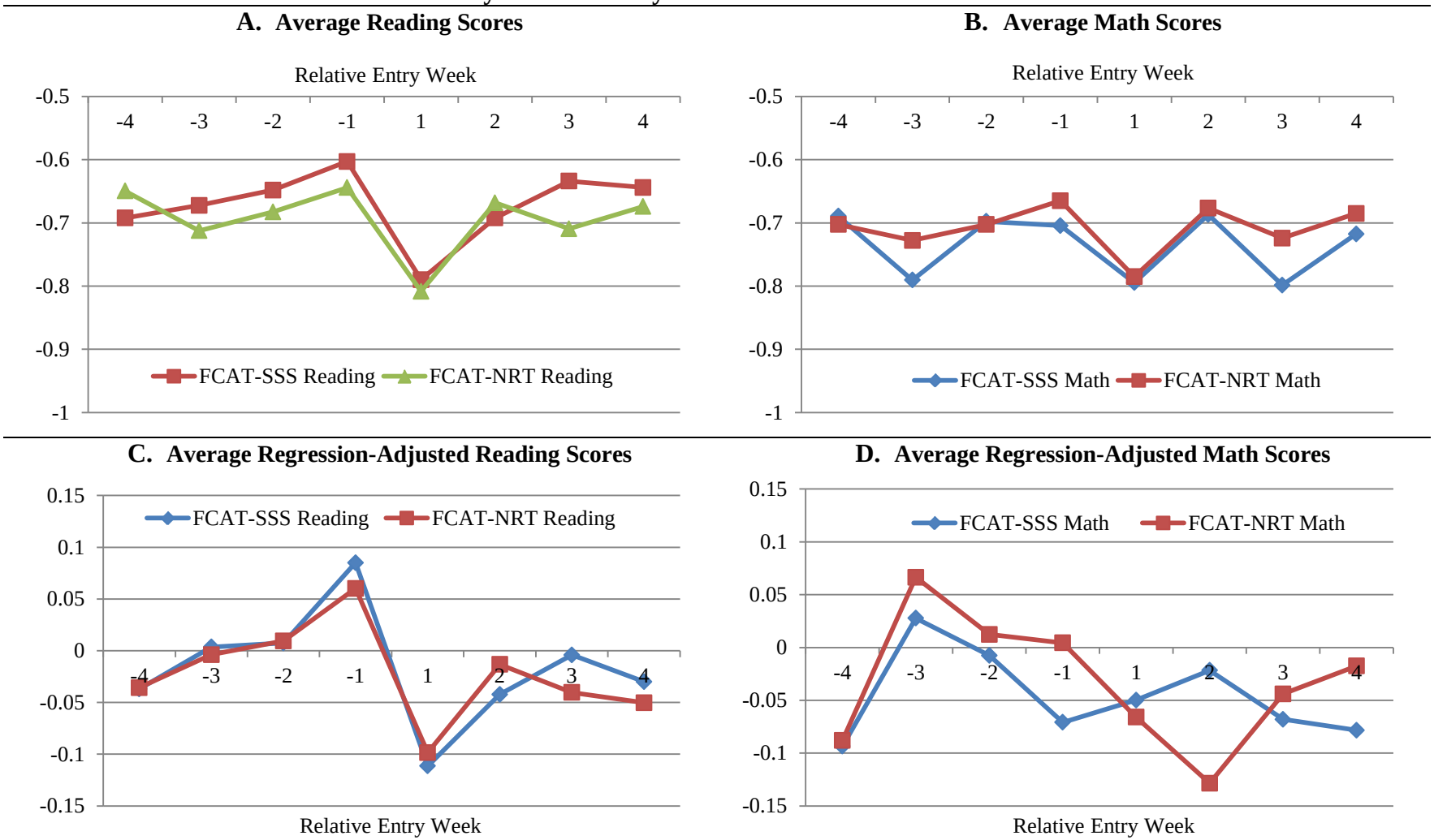


B. Deseasonalized Daily Change in the Number of Entering Students



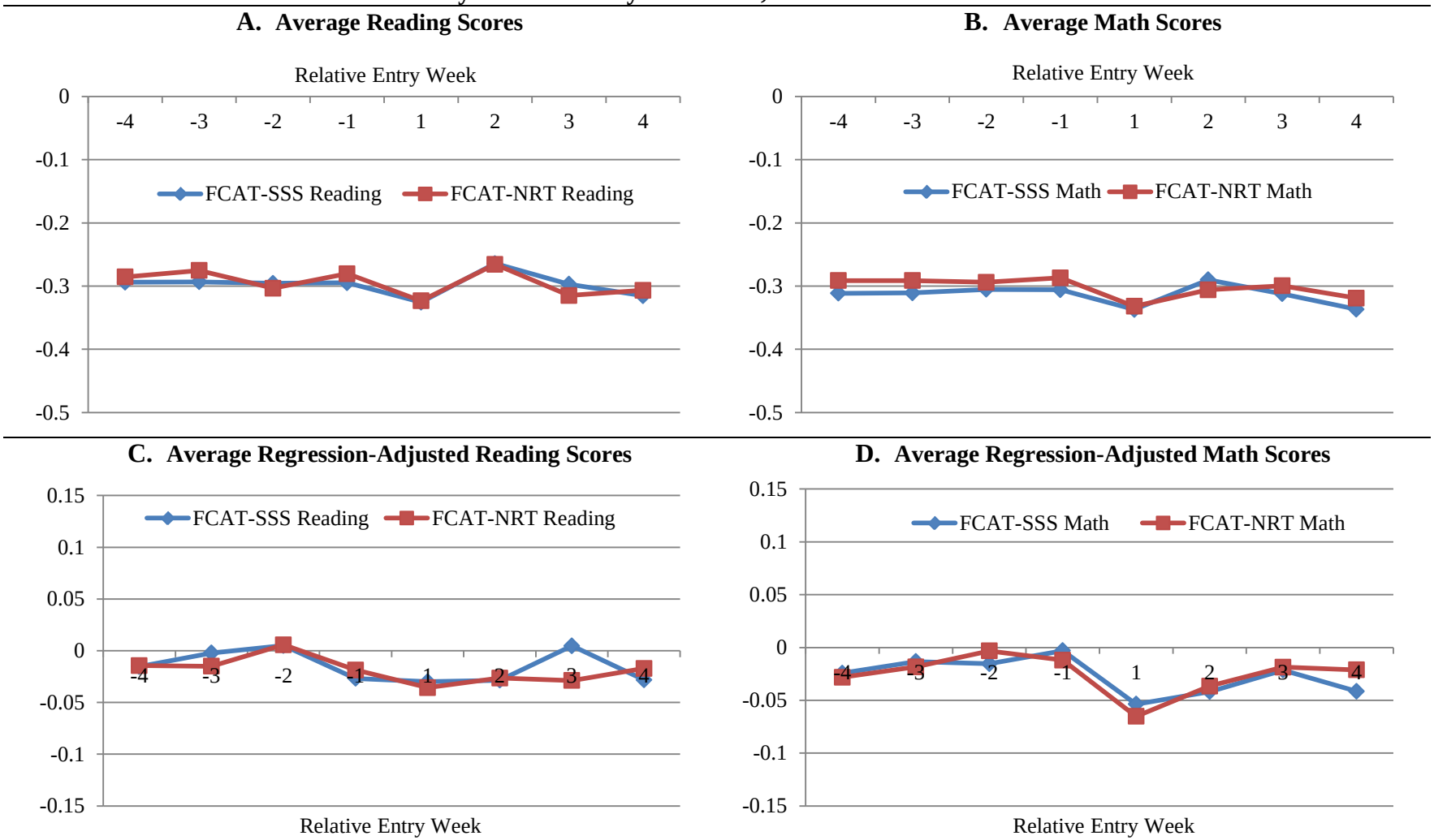
Notes: The two panels present the number of entering students and the 'deseasonalized' daily changes in the number of entering students to the 'D' and 'F' schools in the sample between six weeks before (roughly a week after the beginning of the school year) and two months after (roughly a week before the winter break) the October eligibility cutoff, which is shown by the vertical line.

Figure 5 – Average Raw and Regression-Adjusted Student Achievement by Relative Entry Week: 'D' and 'F' Schools



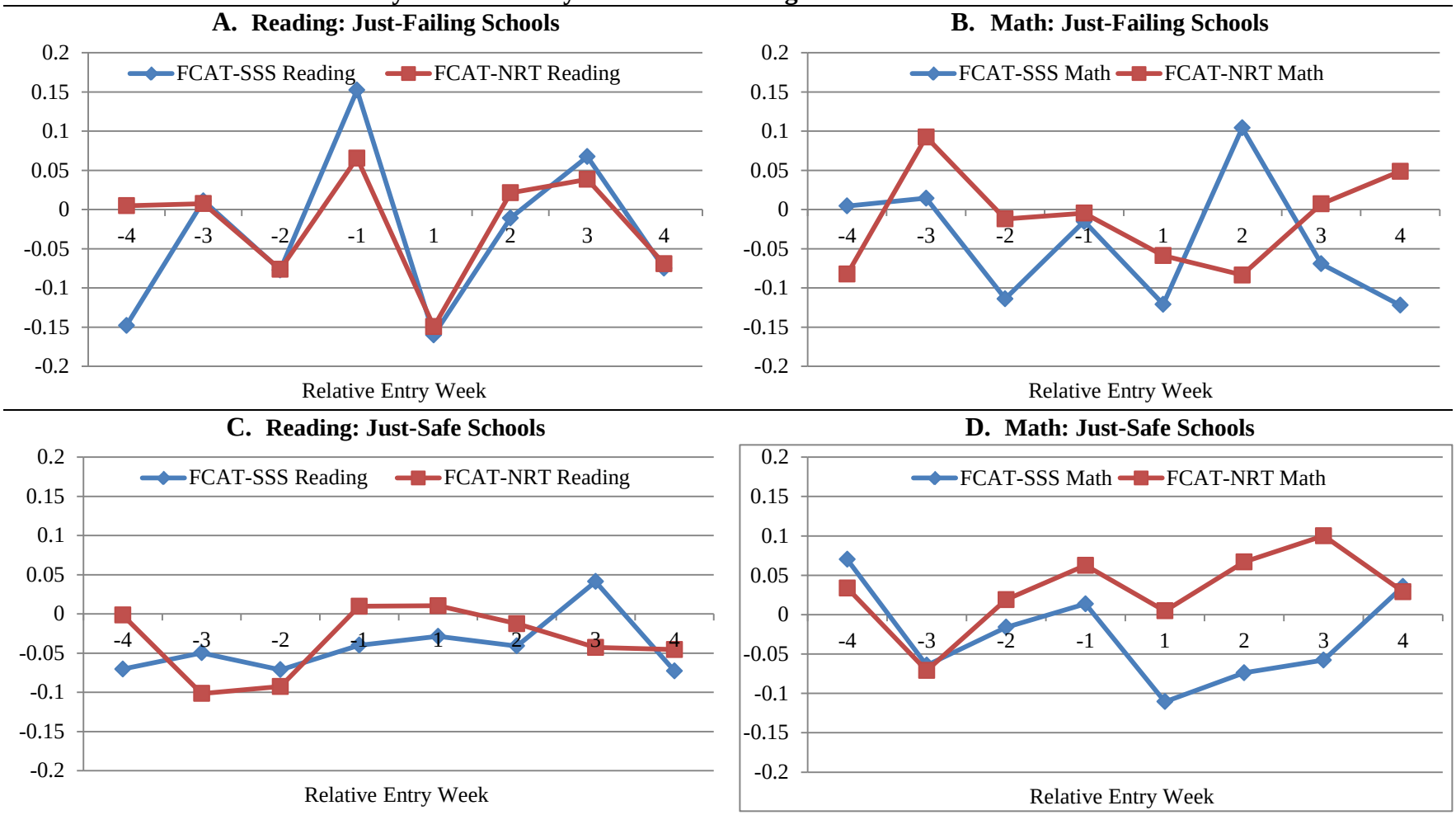
Notes: Panels A and B present the average raw test scores of students at near-failing or failing schools by their entry week whereas Panels C and D present the average test scores that are regression-adjusted by student covariates in listed in Table 7 along with indicators for where the student came from, and grade and year fixed-effects.

Figure 6 – Average Raw and Regression-Adjusted Student Achievement by Relative Entry Week: 'A', 'B' and 'C' Schools



Notes: Panels A and B present the average raw test scores of students at safe schools by their entry week whereas Panels C and D present the average test scores that are regression-adjusted by student covariates in listed in Table 7 along with indicators for where the student came from, and grade and year fixed-effects.

Figure 7 – Average Raw Regression-Adjusted Student Achievement by Relative Entry Week: Just-Failing versus Just-Safe Schools



Notes: The four panels present the average test scores that are regression-adjusted by the student covariates listed in Table 7 along with the indicators for where the student came from, and grade and year fixed-effects. Panels A and B report the findings for 'just-failing schools' that received an accountability score between 300 and 320, and the lower panels present the findings for 'just-safe schools' whose scores fall in the 320-340 band

Tables

Table 1 - October Survey Dates

School Year	Survey Week	Eligibility Cutoff Date
2002-2003	October, 7-11	October, 14
2003-2004	October, 13-17	October, 20
2004-2005	October, 11-15	October, 18
2005-2006	October, 10-14	October, 17

Notes: Eligibility cutoff date is defined as the first day of full academic year 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 02/05/2010.

Table 2 - Elementary School Grade Distribution: 2002-2003 to 2005-2006

	School Year			
School Grade	Summer 2002	Summer 2003	Summer 2004	Summer 2005
A	611	923	1016	991
B	362	352	334	355
C	431	288	289	313
D	123	62	69	91
F	37	17	8	29
Total	1564	1642	1667	1779

Notes: Author's calculations from state data. Elementary schools used in the analysis include all schools serving any of the tested elementary grades in Florida, namely 3rd, 4th and 5th grades.

Table 3 - FAY-Eligibility and Student Characteristics: 3rd, 4th and 5th Graders

	FAY-Eligible Students	FAY-Ineligible Students	FAY-Ineligible Students in 'D' and 'F' Schools
Proficient in reading (prior year)	0.673	0.510***	0.358***
	(0.469)	(0.500)	(0.479)
	[1,012,735]	[57,317]	[3,360]
Proficient in math (prior year)	0.641	0.473***	0.310***
	(0.480)	(0.499)	(0.463)
	[1,013,587]	[57,365]	[3,369]
FRL Eligible	0.519	0.705***	0.902***
	(0.499)	(0.456)	(0.297)
	[2,058,715]	[160,880]	[10,367]
White	0.491	0.391***	0.136***
	(0.499)	(0.488)	(0.343)
	[2,058,715]	[160,880]	[10,367]

Notes: Standard deviations and number of students with non-missing values are given in parentheses and brackets respectively. *, ** and *** indicate that the sample mean is statistically different than the mean for the entire student population in grades 3, 4 and 5 at significance levels of 10, 5 and 1 percent respectively.

Table 4
FAY-Eligibility and Student Achievement: Non-Parametric Estimates

	'A', 'B' and 'C' Schools			'D' and 'F' Schools		
Bandwidth	2	5	10	2	5	10
FCAT-SSS Reading	0.026	-0.005	0.009	-0.464**	-0.466**	-0.248*
	(0.069)	(0.066)	(0.039)	(0.225)	(0.195)	(0.131)
FCAT-SSS Math	-0.002	-0.024	-0.011	-0.388**	-0.366**	-0.182
	(0.061)	(0.06)	(0.035)	(0.192)	(0.182)	(0.121)
FCAT-NRT Reading	-0.084	-0.103*	-0.049	-0.240*	-0.228*	-0.193**
	(0.061)	(0.061)	(0.035)	(0.149)	(0.143)	(0.094)
FCAT-NRT Math	-0.024	-0.08	-0.039	-0.356*	-0.337*	-0.209*
	(0.058)	(0.057)	(0.033)	(0.212)	(0.204)	(0.108)
N(non-zero weight)	7,100	16,601	32,395	453	1,160	2,263

Notes: Standard errors, which are calculated using 2,000 bootstrapping samples clustered at the school level, are given in parentheses. Discontinuity estimates are obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten school days where the outcome is the current year standardized test. The last row gives the number of observations that received non-zero weights in the estimation. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 5 -FAY-Eligibility and Student Achievement: Parametric Estimates

	‘A’, ‘B’ and ‘C’ Schools				‘D’ and ‘F’ Schools			
	Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic
Entry date range	20	20	All	All	20	20	All	All
FCAT-SSS Reading	-0.033	-0.06	-0.071	-0.15	-0.28**	-0.438**	-0.37***	-0.409**
	(0.05)	(0.055)	(0.095)	(0.117)	(0.127)	(0.218)	(0.116)	(0.182)
FCAT-SSS Math	-0.034	-0.037	0.167	-0.148	-0.067	-0.008	0.013	-0.234
	(0.042)	(0.057)	(0.134)	(0.129)	(0.097)	(0.152)	(0.133)	(0.148)
FCAT-NRT Reading	-0.04	-0.102*	-0.106	-0.134	-0.205***	-0.363***	-0.305***	-0.359***
	(0.053)	(0.057)	(0.106)	(0.135)	(0.065)	(0.09)	(0.076)	(0.097)
FCAT-NRT Math	-0.033	-0.063	0.037	-0.103	-0.099	-0.092	-0.044	-0.206*
	(0.044)	(0.051)	(0.123)	(0.137)	(0.081)	(0.123)	(0.107)	(0.11)
				With Student Covariates				
FCAT-SSS Reading	-0.011	0.012	-0.012	-0.035	-0.233**	-0.366*	-0.207*	-0.272
	(0.02)	(0.028)	(0.026)	(0.039)	(0.101)	(0.2)	(0.108)	(0.169)
FCAT-SSS Math	-0.042***	-0.019	0.05*	-0.056*	-0.052	-0.001	-0.017	-0.144
	(0.016)	(0.026)	(0.03)	(0.033)	(0.066)	(0.104)	(0.085)	(0.111)
FCAT-NRT Reading	-0.027	-0.017	-0.041	-0.028	-0.158***	-0.306***	-0.183***	-0.224***
	(0.023)	(0.033)	(0.03)	(0.042)	(0.053)	(0.095)	(0.057)	(0.075)
FCAT-NRT Math	-0.053**	-0.033	-0.003	-0.072	-0.153**	-0.133	-0.115	-0.156*
	(0.025)	(0.042)	(0.034)	(0.048)	(0.062)	(0.113)	(0.073)	(0.09)
				With Student and School Covariates				
FCAT-SSS Reading	-0.016	0.002	-0.022	-0.036	-0.256***	-0.401**	-0.222*	-0.306*
	(0.021)	(0.029)	(0.026)	(0.033)	(0.095)	(0.184)	(0.114)	(0.175)
FCAT-SSS Math	-0.048***	-0.031	0.02	-0.061**	0.012	0.041	0.007	-0.083
	(0.018)	(0.026)	(0.029)	(0.031)	(0.062)	(0.089)	(0.076)	(0.105)
FCAT-NRT Reading	-0.033	-0.021	-0.051*	-0.031	-0.149***	-0.309***	-0.123**	-0.241***
	(0.024)	(0.035)	(0.028)	(0.034)	(0.049)	(0.097)	(0.061)	(0.072)
FCAT-NRT Math	-0.057**	-0.04	-0.03	-0.073	-0.136**	-0.137	-0.091	-0.165*
	(0.025)	(0.044)	(0.034)	(0.046)	(0.056)	(0.119)	(0.073)	(0.092)
N	23,802	23,802	998,202	998,202	1,351	1,351	33,360	33,360

Notes: Robust standard errors, which are two-way clustered at the school and entry day level, are given in parentheses. Discontinuity estimates are obtained parametrically using the specified polynomial order. The first panel presents the estimates from the base specification in equation (3), the second panel incorporates student covariates, and the third panel adds the school covariates. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 6 - FAY-Eligibility and Other Outcomes

	'A', 'B' and 'C' Schools			'D' and 'F' Schools		
Bandwidth	2	5	10	2	5	10
Current Year						
Disciplinary incident	0.014	0.016	-0.004	0.056	-0.021	0.042
	(0.013)	(0.015)	(0.009)	(0.063)	(0.084)	(0.046)
Suspended	0.014	0.016	0.0003	0.067	-0.009	0.044
	(0.013)	(0.014)	(0.008)	(0.06)	(0.079)	(0.045)
% Absent days	-0.009***	-0.011***	-0.008***	-0.02*	-0.02	-0.011
	(0.003)	(0.003)	(0.002)	(0.011)	(0.012)	(0.008)
Promoted	-0.005	0.001	-0.014	-0.028	-0.002	-0.014
	(0.024)	(0.024)	(0.013)	(0.103)	(0.118)	(0.061)
N(non-zero weight)	7,100	16,601	32,395	453	1,160	2,263
Following Year						
FCAT-SSS Reading	0.011	0.029	0.009	-0.276*	-0.292**	-0.165*
	(0.067)	(0.068)	(0.037)	(0.163)	(0.147)	(0.103)
FCAT-SSS Math	0.033	0.068	0.047	-0.371*	-0.387*	-0.147
	(0.062)	(0.061)	(0.034)	(0.224)	(0.208)	(0.149)
FCAT-NRT Reading	0.024	0.041	-0.007	-0.351**	-0.368**	-0.192*
	(0.067)	(0.066)	(0.037)	(0.163)	(0.175)	(0.098)
FCAT-NRT Math	0.02	0.046	-0.004	-0.428***	-0.544***	-0.266***
	(0.057)	(0.057)	(0.033)	(0.153)	(0.174)	(0.098)
Stayed in the same school	0.031	0.021	0.035*	-0.122	-0.169	-0.084
	(0.037)	(0.039)	(0.02)	(0.151)	(0.175)	(0.097)
Disciplinary incident	0.0004	-0.015	-0.011	0.078	0.073	0.114
	(0.024)	(0.026)	(0.015)	(0.133)	(0.127)	(0.069)
Suspended	0.014*	0.011	0.006	0.033	0.028	0.06**
	(0.007)	(0.009)	(0.005)	(0.05)	(0.05)	(0.03)
N(non-zero weight)	6,554	15,209	29,751	432	1,082	2,098

Notes: Standard errors, which were calculated using 2,000 bootstrapping samples clustered at the school level, are given in parentheses. Discontinuity estimates were obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten school days. The last row in each panel gives the number of observations that received non-zero weights in the estimation. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 7
FAY-Eligibility and Student Background Characteristics: Non-Parametric Estimates

	'A', 'B' and 'C' Schools			'D' and 'F' Schools		
Bandwidth	2	5	10	2	5	10
Previous Year						
FCAT-SSS Reading	-0.177*	-0.051	-0.101*	-0.04	-0.047	-0.071
	(0.095)	(0.106)	(0.06)	(0.226)	(0.254)	(0.166)
FCAT-SSS Math	-0.167**	-0.091	-0.067	0.004	-0.078	0.004
	(0.084)	(0.092)	(0.053)	(0.491)	(0.537)	(0.238)
FCAT-NRT Reading	-0.108	-0.065	-0.097*	-0.055	-0.121	-0.068
	(0.098)	(0.1)	(0.055)	(0.187)	(0.217)	(0.135)
FCAT-NRT Math	-0.164*	-0.137	-0.075	0.022	-0.06	0.001
	(0.088)	(0.095)	(0.053)	(0.274)	(0.277)	(0.144)
Disciplinary incident	0.006	-0.015	-0.01	0.078	0.073	0.114*
	(0.024)	(0.025)	(0.015)	(0.127)	(0.123)	(0.065)
Suspended	0.014*	0.011	0.006	0.033	0.028	0.061**
	(0.008)	(0.009)	(0.006)	(0.051)	(0.05)	(0.03)
Ineligible	0.019	-0.035	-0.006	-0.084	-0.11	-0.046
	(0.043)	(0.045)	(0.025)	(0.144)	(0.127)	(0.081)
Promoted	-0.166***	-0.067	-0.106***	0.172	0.18	0.05
	(0.045)	(0.047)	(0.028)	(0.182)	(0.199)	(0.125)
N(non-zero weight)	3,110	7,002	12,928	173	457	817
Withdrawn school						
Disciplinary incident	0.001	-0.006	-0.009*	0.044	0.023	0.004
	(0.007)	(0.008)	(0.005)	(0.039)	(0.047)	(0.031)
Suspended	0.001	-0.006	-0.009*	0.044	0.029	0.009
	(0.007)	(0.008)	(0.005)	(0.041)	(0.046)	(0.028)
N(non-zero weight)	7,100	16,601	32,395	453	1,160	2,263
Other Characteristics						
FRPL eligible	-0.01	-0.003	-0.027*	0.044	0.071	0.04
	(0.028)	(0.028)	(0.016)	(0.042)	(0.072)	(0.044)
Limited English proficiency	-0.035	-0.008	0.003	-0.012	0.029	0.023

	(0.024)	(0.023)	(0.013)	(0.076)	(0.09)	(0.047)
			(Table 7 continued)			
Special education	-0.011	-0.052**	-0.021*	0.022	0.006	0.04
	(0.022)	(0.023)	(0.012)	(0.083)	(0.088)	(0.046)
Gifted	-0.008	-0.014	-0.005	-0.026	-0.042	-0.015
	(0.009)	(0.009)	(0.005)	(0.03)	(0.034)	(0.015)
LEP Ineligible	0.003	0.005	0.001	0.04	0.051*	0.037
	(0.009)	(0.01)	(0.006)	(0.027)	(0.032)	(0.025)
ESE Ineligible	-0.026	-0.047***	-0.028***	-0.04	-0.056	-0.025
	(0.016)	(0.017)	(0.009)	(0.096)	(0.102)	(0.046)
On track	-0.001	0.027	0.01	0.002	-0.02	-0.039
	(0.018)	(0.019)	(0.011)	(0.073)	(0.09)	(0.048)
US born	0.068***	0.049**	0.007	-0.007	-0.025	-0.017
	(0.022)	(0.021)	(0.011)	(0.051)	(0.053)	(0.032)
Male	-0.015	-0.011	-0.01	0.085	0.052	0.021
	(0.025)	(0.026)	(0.014)	(0.072)	(0.083)	(0.049)
English native language	0.061**	0.024	-0.007	0.066	0.024	0.026
	(0.029)	(0.028)	(0.015)	(0.088)	(0.1)	(0.056)
Hispanic	-0.064**	-0.03	-0.004	-0.019	0.01	0.035
	(0.03)	(0.029)	(0.014)	(0.096)	(0.108)	(0.061)
Black	0.085***	0.091***	0.056***	0.123	0.134	0.033
	(0.029)	(0.029)	(0.017)	(0.113)	(0.119)	(0.073)
N(non-zero weight)	7,100	16,601	32,395	453	1,160	2,263

Notes: Standard errors, which were calculated using 2,000 bootstrapping samples clustered at the school level, are given in parentheses. Discontinuity estimates were obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten school days. The last row in each panel gives the number of observations that received non-zero weights in the estimation. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 8 – FAY-Eligibility and Student Achievement: Pseudo Cutoffs, 'D' and 'F' Schools

	Two weeks earlier			One week earlier		
Bandwidth	2	5	10	2	5	10
FCAT-SSS Reading	0.131	0.096	0.02	0.141	0.147	0.098
	(0.129)	(0.152)	(0.091)	(0.172)	(0.174)	(0.107)
FCAT-SSS Math	0.06	0.031	0.101	0.202	0.095	0.01
	(0.147)	(0.16)	(0.109)	(0.196)	(0.189)	(0.117)
FCAT-NRT Reading	0.265**	0.196	0.074	0.168	0.158	0.131
	(0.134)	(0.157)	(0.095)	(0.146)	(0.151)	(0.098)
FCAT-NRT Math	0.133	0.059	0.034	0.191	0.123	0.11
	(0.122)	(0.141)	(0.088)	(0.131)	(0.126)	(0.089)
N(non-zero weight)	549	1,167	2,295	495	1,263	2,346
	One week after			Two weeks after		
FCAT-SSS Reading	0.215	0.333	0.183	0.247*	0.275*	0.125
	(0.23)	(0.22)	(0.121)	(0.133)	(0.148)	(0.09)
FCAT-SSS Math	0.209	0.293	0.137	-0.131	-0.02	-0.065
	(0.237)	(0.225)	(0.123)	(0.123)	(0.154)	(0.105)
FCAT-NRT Reading	0.25	0.334*	0.237**	0.169	0.202	0.057
	(0.177)	(0.192)	(0.109)	(0.112)	(0.128)	(0.085)
FCAT-NRT Math	0.184	0.168	0.097	0.167	0.218	0.086
	(0.178)	(0.181)	(0.109)	(0.135)	(0.158)	(0.09)
N(non-zero weight)	406	1,039	2,241	482	1,213	2,229

Notes: Standard errors, which were calculated using 2,000 bootstrapping samples clustered at the school level, are given in parentheses. Discontinuity estimates were obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten weekdays with the specified pseudo cutoffs where the outcome is the current year standardized test score. The last row gives the number of observations that received non-zero weights in the estimation. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 9 – FAY-Eligibility and School Characteristics

	'A', 'B' and 'C' Schools			'D' and 'F' Schools		
Bandwidth	2	5	10	2	5	10
Prior Year Performance						
% meeting high standards in reading	0.007	0.008	0.006	0.026	0.02	0.021*
	(0.01)	(0.011)	(0.006)	(0.02)	(0.02)	(0.012)
% meeting high standards in math	0.017	0.02*	0.012*	-0.04	-0.046	-0.016
	(0.013)	(0.012)	(0.007)	(0.043)	(0.05)	(0.025)
% making learning gains in reading	0.007	0.002	0.005	0.014	0.013	0.011
	(0.007)	(0.007)	(0.004)	(0.014)	(0.015)	(0.008)
% making learning gains in math	0.001	-0.006	0.005*	-0.02	-0.013	-0.015
	(0.004)	(0.004)	(0.003)	(0.015)	(0.017)	(0.011)
% low-performers making gains in reading	0.002	-0.017**	-0.002	0.026	0.031	0.023
	(0.007)	(0.008)	(0.006)	(0.055)	(0.062)	(0.031)
Other Characteristics						
'A' or 'B' school in the prior year	0.117***	0.148***	0.069***	0.005	0.013	0.015
	(0.040)	(0.040)	(0.022)	(0.005)	(0.009)	(0.010)
'D' or 'F' school in the prior year	-0.005	-0.0001	-0.001	0.099	0.125	0.060
	(0.012)	(0.012)	(0.007)	(0.112)	(0.135)	(0.083)
Average teacher experience	-0.187	-0.1	-0.21	-0.756	-0.99	-0.187
	(0.219)	(0.221)	(0.135)	(1.236)	(1.435)	(0.758)
% teachers with advanced degrees	0.008	0.011	0.008*	0.034	0.055	0.014
	(0.007)	(0.008)	(0.004)	(0.037)	(0.04)	(0.021)
% FRPL eligible students	-0.016	-0.007	-0.013	0.036	0.042	0.013
	(0.017)	(0.017)	(0.011)	(0.023)	(0.027)	(0.02)
% chronically absent students	-0.001	-0.001	0.001	-0.009	-0.016	-0.003
	(0.003)	(0.003)	(0.002)	(0.014)	(0.015)	(0.01)
% special education students	0.002	0.0001	0.0003	0.012	0.018	0.011
	(0.003)	(0.003)	(0.002)	(0.017)	(0.018)	(0.01)
% stable students	0.0004	-0.001	0.001	0.005	0.005	0.001
	(0.002)	(0.002)	(0.001)	(0.007)	(0.008)	(0.006)
Disciplinary incident rate	-0.002	0.001	-0.001	0.02	0.017	0.019
	(0.003)	(0.004)	(0.003)	(0.014)	(0.016)	(0.012)

Notes: Standard errors, which were calculated using 2,000 bootstrapping samples clustered at the school level, are given in parentheses. Discontinuity estimates were obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten school days. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 10 – FAY-Eligibility and Student Achievement: Within-school Comparisons

	‘A’, ‘B’ and ‘C’ Schools				‘D’ and ‘F’ Schools			
	Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic
Entry date range	20	20	All	All	20	20	All	All
FCAT-SSS Reading	-0.011	-0.027	-0.015	-0.043	-0.221***	-0.164***	-0.2***	-0.208***
	(0.022)	(0.019)	(0.026)	(0.027)	(0.04)	(0.024)	(0.052)	(0.032)
FCAT-SSS Math	0.047**	0.018	-0.067	- 0.064	-0.385***	-0.217***	-0.271***	-0.188***
	(0.019)	(0.021)	(0.062)	(0.053)	(0.042)	(0.048)	(0.053)	(0.057)
FCAT-NRT Reading	-0.04***	-0.047**	0.053	0.001	-0.074**	-0.147**	0.004	-0.089**
	(0.006)	(0.014)	(0.037)	(0.021)	(0.031)	(0.048)	(0.054)	(0.039)
FCAT-NRT Math	-0.036**	-0.035	-0.059*	-0.086*	-0.02	0.043	-0.153***	-0.074
	(0.013)	(0.025)	(0.032)	(0.047)	(0.067)	(0.068)	(0.053)	(0.062)
				With School Fixed-Effects				
FCAT-SSS Reading	-0.014	-0.027	-0.087*	-0.085**	-0.199***	-0.205***	-0.226***	-0.206***
	(0.019)	(0.015)	(0.05)	(0.04)	(0.031)	(0.044)	(0.032)	(0.044)
FCAT-SSS Math	0.049**	0.008	-0.008	-0.018	-0.186*	-0.259**	-0.194***	-0.109*
	(0.02)	(0.017)	(0.032)	(0.03)	(0.086)	(0.075)	(0.044)	(0.06)
FCAT-NRT Reading	-0.048***	-0.044***	-0.068***	-0.096**	-0.109**	-0.184**	-0.051	-0.112*
	(0.007)	(0.008)	(0.02)	(0.036)	(0.038)	(0.055)	(0.054)	(0.056)
FCAT-NRT Math	-0.074***	-0.049**	-0.032	-0.053	0.009	-0.037	0.002	0.001
	(0.021)	(0.019)	(0.022)	(0.032)	(0.089)	(0.114)	(0.056)	(0.076)
N	23,802	23,802	998,202	998,202	1,351	1,351	33,360	33,360

Notes: Robust standard errors, which are two-way clustered at the school and entry week level, are given in parentheses. Discontinuity estimates are obtained parametrically using the specified polynomial order. The upper panel presents the estimates from the base specification in equation (3) with student covariates where the selection variable is the relative entry week of the student, and the lower panel adds school fixed-effects to the estimation. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 11 – FAY-Eligibility and 'Out-of-Sample' Classroom Performance

	'A', 'B' and 'C' Schools			'D' and 'F' Schools		
Bandwidth	2	5	10	2	5	10
FCAT-SSS Reading	0.014	0.003	0.006	-0.152**	-0.121*	-0.007
	(0.028)	(0.028)	(0.015)	(0.066)	(0.071)	(0.054)
FCAT-SSS Math	0.023	0.017	0.019	-0.145**	-0.093	-0.003
	(0.029)	(0.029)	(0.015)	(0.07)	(0.085)	(0.053)
FCAT-NRT Reading	0.004	0.002	0.004	-0.18***	-0.151**	-0.026
	(0.031)	(0.031)	(0.016)	(0.062)	(0.072)	(0.046)
FCAT-NRT Math	0.01	0.008	0.002	-0.137**	-0.122*	-0.002
	(0.029)	(0.028)	(0.016)	(0.063)	(0.073)	(0.045)
N(non-zero weight)	5,532	13,335	26,067	352	907	1,744

Notes: Standard errors, which were calculated using 2,000 bootstrapping samples clustered at the classroom level, are given in parentheses. Discontinuity estimates were obtained using kernel-weighted local linear smoothing with bandwidths of two, five and ten weekdays where the outcome is the average classroom test score calculated by excluding the students within the specified bandwidth. The last row gives the number of observations that received non-zero weights in the estimation. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

Table 12 - FAY-Eligibility and Student Achievement: 'Just-Failing' versus 'Just-Safe' Schools

	'Just-Safe' Schools					
	Linear	Quartic	Linear	Quartic	Linear	Quartic
Entry date range	20	All	20	All	20	All
FCAT-SSS Reading	0.119	0.071	0.036	-0.096	0.017	-0.118
	(0.142)	(0.173)	(0.088)	(0.109)	(0.091)	(0.098)
FCAT-SSS Math	0.032	-0.06	-0.133*	-0.245**	-0.099*	-0.235**
	(0.129)	(0.186)	(0.07)	(0.119)	(0.058)	(0.103)
FCAT-NRT Reading	0.134	0.076	0.005	-0.083	0.016	-0.108
	(0.11)	(0.149)	(0.086)	(0.11)	(0.087)	(0.109)
FCAT-NRT Math	0.075	0.022	-0.04	-0.094	-0.045	-0.094
	(0.1)	(0.139)	(0.058)	(0.064)	(0.065)	(0.073)
N	1,184	29,541	1,184	29,541	1,184	29,541
	'Just-Failing Schools'					
FCAT-SSS Reading	-0.405***	-0.473**	-0.297**	-0.372*	-0.303**	-0.437**
	(0.157)	(0.19)	(0.14)	(0.219)	(0.151)	(0.216)
FCAT-SSS Math	-0.136	-0.177	0.001	-0.056	0.091	-0.08
	(0.098)	(0.126)	(0.058)	(0.137)	(0.073)	(0.113)
FCAT-NRT Reading	-0.248**	-0.283***	-0.181***	-0.178**	-0.23***	-0.269***
	(0.099)	(0.109)	(0.066)	(0.084)	(0.089)	(0.09)
FCAT-NRT Math	-0.116	-0.109	-0.185*	-0.114	-0.187*	-0.193*
	(0.102)	(0.126)	(0.096)	(0.11)	(0.111)	(0.115)
Student covariates	No	No	Yes	Yes	Yes	Yes
School covariates	No	No	No	No	Yes	Yes
N	775	19,561	775	19,561	775	19,561

Notes: Robust standard errors, which are two-way clustered at the school and entry day level, are given in parentheses. Discontinuity estimates are obtained parametrically using the specified polynomial order. The first two columns present the estimates from the base specification in equation (3), third and fourth columns incorporate student covariates, and the last two columns add school covariates. The upper panel reports the estimates for just-safe report the findings for 'just-safe schools' that received accountability scores between 320 and 340, and the lower panels present the findings for 'just-failing schools' whose scores fall in the 300-320 band. *, ** and *** represent statistical significance at 10, 5 and 1 percent respectively.

APPENDIX A

Full Academic Year Definitions in 50 States and the District of Columbia as of 2009

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 concluded)	
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 02/01/2010. For more information, visit http://www2.ed.gov/admins/lead/account/stateplans03/index.html.

APPENDIX B

Grading Formula in Florida: 2008-2009 School Year

FCAT-SSS performance in various subjects including reading, math, writing and science is the main determinant of school grades in Florida. For each subject, student scores are classified into five achievement levels, with 1 being the lowest and 5 being the highest. Given these levels, there are three main components that determine schools grades in Florida:

1. Percentage of eligible students meeting high standards in reading, math (both given in grades three through ten), science (given in grades 5, 8 and 10) and writing (given in grades 4, 8 and 10). The proficiency threshold is achievement level 3 in reading, math and science, and 3.5 in writing, and students are considered proficient in a given subject if they perform at or above the threshold in that subject.
2. Percentage of eligible students making learning gains in reading and math. Students can demonstrate learning gains by improving the prior achievement level, maintaining the 'proficient' level or demonstrating more than 'one year's growth' within achievement levels 1 and 2.
3. Percentage of eligible students in the lowest quartile in reading or math in the schools with adequate progress in reading or math.

For the first component, eligibility is defined as the intersection of LEP eligibility, ESE eligibility and FAY eligibility. For the last two components, on the other hand, only FAY-ineligible students are excluded.

These eight quantities are then added to calculate the aggregate grade points for each school. School grades are determined using the following scale:

Table B1 - School Grading Scale in Florida: 2008-2009

School Grade	Requirements
A	• 525 points or more • Percentages of eligible students in the lowest quartile in reading or math with adequate progress are at least 50 percent each • At least 95% of eligible students are tested
B	• 495-524 points • Percentages of eligible students in the lowest quartile in reading or math with adequate progress are at least 50 percent each within two years • At least 90% of eligible students are tested
C	• 435-494 points • Percentages of eligible students in the lowest quartile in reading or math with adequate progress are at least 50 percent each within two years • At least 90% of eligible students are tested
D	• 395-434 points • At least 90% of eligible students are tested
F	• 395 or lower points OR • Less than 90% of eligible students are tested

Notes: Compiled from Appendix A in Florida Department of Education (FLDOE) publication titled "2009 Guide to Calculating School Grades: Technical Assistance Paper" posted on FLDOE website: http://schoolgrades.fldoe.org/pdf/0809/2009SchoolGradesTAP.pdf accessed on 02/05/2010.

Schools that fail to make adequate progress with their lowest performing students in reading and math need to develop a School Improvement Plan component that addresses this need. If a school, otherwise graded "C" or "B", does not demonstrate adequate progress in either the current or prior year, the final grade is reduced by one letter grade.