Technical Report

THE EFFECTS OF A POSITIVE BEHAVIOR INTERVENTIONS AND SUPPORTS (PBIS) DATA PLATFORM ON STUDENT ACADEMIC AND DISCIPLINARY OUTCOMES

EDUCATION RESEARCH ALLIANCE FOR NEW ORLEANS

Nathan Barrett, Tulane University
Douglas N. Harris, Tulane University

October 23, 2018

EducationResearchAllianceNOLA.org

The Effects of a Positive Behavior Interventions and Supports (PBIS) Data Platform on Student Academic and Disciplinary Outcomes

Nathan Barrett
Douglas N. Harris

October 23, 2018

Abstract: There is growing concern about suspension and expulsion rates and disparities by race and income, and interest in focusing schools on student outcomes, such as socio-emotional learning (SEL) that go beyond test scores. Improved behavioral management has the potential to accomplish both, by reducing negative behaviors and encouraging non-academic behaviors that are important to long-term life success. Positive Behavior Interventions and Supports (PBIS) is one increasingly common approach being used in thousands of schools across the country. We study a PBIS data platform called Kickboard. Using difference-in-difference analysis, with a matched comparison group, we find that Kickboard reduces the number of suspensions by 0.14-0.38 per student per year (26-72 percent from baseline) and the number of suspension days by 0.7-1.5 (at least 52 percent). We also see some evidence of reductions in suspensions for violent behavior. The effects are concentrated in schools that had at least modest implementation. Overall, these results are in line with prior rigorous research on the PBIS, which has shown positive effects on a range of student outcomes.

Acknowledgements: This work was carried out at the Education Research Alliance for New Orleans at Tulane University, which is funded by Tulane as well as the Laura and John Arnold Foundation, the William T. Grant Foundation, and the Spencer Foundation. We thank Tom Hierck, Robert Horner, Kent Peterson, and Keith Smolkowski for their comments on the draft and for sharing their expertise on PBIS. Sara Slaughter provided invaluable support in editing this report and the accompanying policy brief. This specific project was also supported in part by funding from Kickboard. The views expressed in this report reflect only those of the authors.

Author Information: Nathan Barrett, nbarret@tulane.edu, 1555 Poydras, Suite 700, New Orleans, LA. 70112. Douglas N. Harris, dharri5@tulane.edu, 302 Tilton Hall, 6823 St. Charles Ave., New Orleans, LA 70118.

I. Introduction

One widely held purpose of schooling, in addition to teaching academic skills, is to develop the whole child, including developing positive social behaviors. Perseverance, grit, cooperation, and other aspects of social and emotional learning (SEL) can help students as they become adults. In the short term, these behaviors can also help build school environments that are more conducive to academic learning. Students are unlikely to learn math, reading, and other skills if their schools are not safe and orderly and teachers often report that classroom management is one aspect of teaching they are least prepared for (Freeman, Simonsen, Briere, & MacSuga-Gage, 2014).

School policies, such as student discipline, can also affect student behavior. Over the past two decades, many schools have adopted zero-tolerance policies that harshly punish even small infractions and on the first offense. Part of the theory, following on the "broken windows" approach to policing (Wilson & Kelling, 1982), is that strict discipline prevents negative incidents and therefore encourages more positive behavior. The large number of security guards and police officers in schools is symbolic of this focus on reducing negative behaviors (Barnum, 2016).

More recently, however, attention has shifted from zero-tolerance policies to more "positive" school climate strategies intended to reduce the number of suspensions and expulsions (U.S. Department of Justice & U.S. Department of Education, 2014). The main goal is to focus less on punishing bad behavior and more on teaching, recognizing, and reinforcing positive behaviors. By reducing exclusionary discipline, students will spend more time learning or getting the help they may need, especially where their behaviors are manifestations of learning disabilities, trauma, mental health disorders, or other factors that

school services may be able to help address. Being out of school leaves these students further behind academically and more likely to act out, and may place some students in unhealthy or dangerous environments in their homes and communities. There is some evidence that being suspended or expelled is correlated with negative academic outcomes and growing concern that these policies may be related to what is sometimes called the school-to-prison pipeline (Skiba & Rausch, 2004; Skiba & Williams, 2014). That suspensions and expulsions are much more common among racial minorities, low-income, and special education students reinforces these concerns (Losen et al., 2014; Loveless, 2017).[1]

These disciplinary trends and other factors have led to interest in Positive Behavior Interventions and Supports (PBIS), "a prevention-oriented, problem-solving approach that emphasizes: (a) a continuum of supports to meet the needs of all students, (b) regular monitoring of implementation and outcomes, and (c) the use of data to guide decisions" (Mitchell et al. 2018, p.1). PBIS include broad-based school-level activities ("Tier I") and targeting services to specific groups of students (Tiers II and III) (Mitchell et al., 2018). Previously referred to as Effective Behavior Support (Sprague et al., 2001), more than 25,000 schools nationally, or roughly twenty-five percent of all schools, report having adopted some version of PBIS (Sugai, 2018).

Prior research suggests that consistently positive effects of PBIS on the number of referrals to the principal's office, number of suspensions, and academic achievement

[1] In 2014, 5.3% of all students nationally were suspended out-of-school and 0.2% were expelled (Hernandez, 2018). These same rates were 2 to 3 times higher among black students, 13.5% of whom were suspended out-of-school and 0.4% of whom were expelled. Research has also found that Black students are suspended at higher rates than White students in fights between Black and White students (Barrett et al., 2018).

(Horner et al., 2009; Ward & Gersten, 2013; Smolkowski, Stryker, & Ward, 2016; Mitchell, Hatton, & Lewis, 2018). Moreover, there is evidence that additional PBIS training improves various measures of school climate; however, there have been few experimental and quasi-experiments, and these have focused on PBIS training in four sites (Mitchell, Hatton, & Lewis, 2018).[2]

On the other hand, concern about PBIS arises because of its heavy reliance on extrinsic incentives. For this reason, the effects may be short-lived and reinforce a transactional way of thinking—that "positive behavior" is worthwhile only because of the immediate rewards that it brings. The potential to reward outcomes over behaviors is also a point of concern for promoting intrinsic motivation. If two students put forth equal effort but one does better on a test and receives an award, the other student may view that reward as negative (Deci, 1975; Lepper et al., 1973). Deci et al. (2001) conducted a meta-analysis of 128 experiments and concluded that expected tangible rewards significantly undermine intrinsic motivation (though this only applies when students are already intrinsically motivated to carry out the behavior). Their findings were particularly strong for school-aged children. However, there is some scholarly debate about this. Some suggest that, when properly used, extrinsic rewards might not reduce intrinsic motivation (Reiss, 2005; Cerasoli et al., 2014; Hidi, 2016).

We study the effectiveness of a customizable PBIS data and professional development platform called Kickboard. The software provides a menu of positive behavioral categories,

[2] Horner et al. (2009) conducted an RCT in 63 schools in Hawaii and Illinois. Bradshaw et al. (2012) conducted an RCT with 42 schools in Maryland. Ward and Gersten (2013) carried out an RCT with 33 schools in a large urban district. The most recent study of 37 schools in Maryland comes from Bradshaw, Waasdorp, & Leaf (2015). We omit a quasi-experiment by Sprague et al. (2001), because it did not account for baseline nonequivalence.

accompanied by varying levels of training with teachers and administrators. Teachers can input behavior data in real time using cell phones and tablet computers and allocate rewards to students on points-based system. The data are summarized in data dashboards that allow teachers to see data patterns for individual students and whole classes, including trends over time, in ways that are meant to shape student-teacher interactions. School administrators can track behavioral outcomes at the school level, and by classroom and teacher. Kickboard sells the software and PBIS/SEL-focused professional development for-profit to schools on a per-student basis and captures all inputted data. Founded in New Orleans after the city's intense market-based school reform, many of the schools using the software are charter schools.

We study the effects of adopting Kickboard in publicly funded schools in Louisiana using quasi-experimental methods. Specifically, we used the difference-in-differences method in which we first compare students in treatment schools before and after the schools adopted Kickboard (the first difference) and then subtract this from the change in a comparison group (the second difference), while also controlling for other student and school characteristics. The comparison group is matched to the treatment group at the individual student and school levels and trends using pre-treatment values of the dependent variable.

We find that the average treatment effect of Kickboard reduces the number of suspensions by 0.14-0.38 per student per year (26-72 percent from baseline) and the number of suspension days by 0.7-1.5 (at least 52 percent). A key potential threat to validity in the analysis is that schools may have simultaneously adopted other changes in policy and practice that might influence these same outcomes; for example, schools may have

relaxed their suspension policies at the same time they adopted Kickboard, as part of a concerted effort to reduce suspensions. While it is difficult to test this directly, we carried out two additional analyses that suggest the above effects are at least partially caused by Kickboard. First, we see some evidence of reductions in suspensions for violent behavior. Such effects on violent behavior are unlikely to be due to changes in school discipline policy, as compared with less violent behavior where the appropriate use of suspension is less clear cut. Second, we estimated effects separately for schools that had very little implementation. If schools adopting Kickboard were also relaxing suspension policies, then we would have expected to see effects regardless of implementation. These validity checks are not definitive, and the results are sensitive to sample restrictions, but they provide at least suggestive evidence that the effects might be due to Kickboard adoption.

This study adds to the existing literature in several ways: adding the first rigorous quasi-experimental research of PBIS; using a larger sample that allows for more extensive effect heterogeneity analysis and validity checks. An additional distinguishing feature is that all prior studies have focused on the PBIS professional development, but the majority of Kickboard schools only receive the software and perhaps minimal professional development. Below, we introduce the intervention, describe our methods and data, present our results, and provide concluding thoughts.

II. The Intervention: Kickboard

The leaders of Kickboard describe it as being rooted in both PBIS and socio-emotional learning (SEL). To support PBIS in schools, Kickboard focuses on collecting data about student behavior and teacher responses to behavior, tracking outcomes over

time, and building a positive school culture and climate.[3] The company began operating in 2009. Competitors include Class Dojo (https://www.classdojo.com/).

Schools that purchase the Kickboard platform receive two short sessions of virtual assistance for setting up and using the software, and customizing the behavior categories that schools use. The initial configuration of the software includes behaviors that have been typically measured by teachers and included on report cards under what is sometimes called "citizenship." PBIS, however, emphasizes the importance of measuring and rewarding specific behaviors. Kickboard includes, for example, "collaboration," "kindness," and "caring" in its menu of options.[4] Schools can add in additional behavioral codes of their choice.

Roughly 20 percent of schools also purchase from Kickboard 1-4 additional in-person services. These include 1-2 six-hour workshops with teachers and school administrators, entitled, "Envisioning Excellence and Strategic Planning for Culture" and "Empowering Staff for Success: Mindsets, Tools, and Strategies for a Positive School Culture." In the first of these, for example, Kickboard staff "guide the school leadership team to collaboratively create a vision for schoolwide culture excellence based on their shared beliefs."

[3] The Kickboard web site indicates that, at PBIS schools, "student expectations and character development with the same intentionality as academic content . . . schoolwide expectations are created by a multidisciplinary team after an assessment of the school's culture needs is conducted…behavior expectations are positively stated, and then are further defined across various settings within the school… [and] positive behaviors are reinforced through individual and group incentives.

[4] The full list of pre-programed "positive student behaviors" includes: "showing pride in school, collaboration, kindness, takes pride in one's work, leadership, helps others, uses time wisely, being prepared, love of learning, makes good choices, active listening/engaged, cooperation, uses appropriate communication, caring, self-reliant, perseverance/resilience, making an insightful comment, organization, above and beyond."

Schools can also purchase 1-2 follow-up sessions roughly six months after the initial implementation that focus on "Progress Monitoring For PBIS: Systems to Ensure School Culture Goals are Achieved" and "Culture Professional Learning Communities: A Framework for Using Data to Drive Improvement," which are meant to sustain and embed the software and PBIS in the ongoing work of educators. Kickboard provides surveys to schools and observes the school in operation; these data are then used to develop plans. While we are not able to distinguish between schools that receive the additional services and follows-ups, the analysis that follows does describe various aspects of teacher and administrator implementation and analyzes effects by intensity of implementation.

III. Data and Sample

Kickboard provided anonymized data to us at the detailed mark level, i.e., we can see each time a teacher decided to mark a student behavior. Each mark is coded to a specific behavior, selected by the teachers who can also add more detailed comments about the behaviors. The date and time are automatically captured. User data also include the date of teacher and administrator logins. Each teacher login is usually associated with many different student marks from throughout a class period or an entire day. Since we cannot link them to other data at the student level, we aggregate the marks to the school level to measure and understand implementation.[5]

Additional school and student-level data come from the Louisiana Department of Education (LDOE) and include student enrollments, demographics, test scores, and

[5] Further, while we can distinguish the marks made by different teachers (see descriptive analysis below), we cannot link the teacher IDs in the Kickboard data to the teacher IDs in the LDOE data.

disciplinary records. We use years 2009-10 through 2015-16, i.e., starting two years prior to the first year of Kickboard adoption to establish baseline equivalence and parallel trends. Enrollment data allow us to determine the school and grade in which a student is enrolled and to track students if they transfer schools. Demographic data include students' race/ethnicity, free or reduced lunch status (FRL), gender, and if the student has an identified individual education program (IEP), traditionally referred to as special education. Student outcome data include test scores for math, English Language Arts (ELA), science, and social studies, which we standardize by subject, grade, and year. We also have school performance scores (SPS) that are calculated by the state from which they assign school letter grades that summarize performance.

Given the focus on (non-academic) behavior, perhaps the most important outcomes are disciplinary infractions. As in most school data systems, we observe suspensions, the number of days suspended, and the type of infraction that led to the suspension. This data structure implies that we only observe information about behaviors that resulted in a suspension.[6] Behaviors that do not result in suspensions therefore are not observed.

With the above data, we are able to study 70 Louisiana schools that used Kickboard and have sufficient years of data over the period of study. These are almost entirely elementary and middle schools. For this reason, we limit ourselves to standardized test scores and discipline incidents, the only relevant measures that are available during those grades.

[6] We can also observe expulsions, but these occur too infrequently to be included in the analyses.

Forty of the 70 schools using Kickboard in our sample are in New Orleans (NOLA). New Orleans provides a unique context in which to study the effects of Kickboard because of the intense test- and market-based accountability put in place after Hurricane Katrina (Harris, 2015). Schools can be closed for poor performance with new schools opening in their place, and more than 40 schools have been taken over this way (Bross & Harris, 2016), so that the set of open schools has been constantly in flux. Some schools (comparison and treatment) exit the sample during the panel period, while other schools open during the panel period and start using Kickboard from the first day they are opened. This affects our matching process and requires additional assumptions about the pre-trends of the treatment group, as discussed later.

We do not have a balanced panel of schools. Table 1 summarizes the sample of schools and the years they were opened (in relation to the first year of Kickboard treatment (t=0). While the starting year with Kickboard varies by school, we align the first treatment year to be t, the two years prior to adoption to be t-2 and t-1, and the two years post adoption to be t+1 and t+2. Twenty-eight of our 70 (12 NOLA and 16 non-NOLA) schools were open for the entire period of study (t-2 through t+2). Another 20 schools (16 NOLA and 4 non-NOLA) were not open in either the first or last year of the study. Fifteen schools (10 NOLA and five non-NOLA) were not open in either the first two or last two years of the study. Finally, five schools are open for only two years during the period of the study. This is primarily because four schools opened in the 2014-15 school year and data are only available through 2015-16.[7]

[7] While these schools have only post-treatment LDOE data, we can include them by tracking individual students in those schools back to their prior outcomes in other schools.

Students attending a Kickboard school, on average, score lower on all standardized tests than students in the donor pool and are more likely to be suspended and serve more total days of suspension than the average student attending a donor pool school (Table 2, Columns 1 and 2). Kickboard schools also disproportionally serve students of color (93 percent) and students eligible for FRL (87 percent) as compared to schools in the donor pool. Accordingly, we conduct a matching process (described below) to create a comparison group of students that more closely resembles the group of students attending a Kickboard school (Column 3 of Table 2). The last two columns in Table 2 show that the matching process greatly reduces the differences between the Kickboard schools and the donor pool of schools.

IV. Methods

VI.A. Difference in Differences

We combine propensity score matching (PSM) with a difference-in-differences (DD) estimation model to determine the effects of Kickboard adoption on various student outcomes including test-scores and discipline. Specifically, we estimate the effects of the adoption of Kickboard starting with standard two-period DD estimation (Angrist & Pischke, 2009):

$$Y_{ikt} = \alpha_0 + \delta_0 K_i + \gamma_t d_t + \beta_t (K_i \cdot d_t) + X_{it} + \theta_k + \varepsilon_{it} \quad (1)$$

where Y_{ikt} is the outcome of interest for student i in school k at time t. K_i is the Kickboard indicator variable and is equal to one of student i attended a Kickboard school in that school's first year of adoption and zero otherwise. The term d_t is a vector of year indicators which measures the relative distance from the first treatment year with γ_t measuring the difference between year t and the baseline year. The model also controls

for school fixed-effects (θ_k) and student-level characteristics (X_{it}).[8] Ordinary least squares estimation of β_t provides a plausibly unbiased estimate of the average treatment effect (ATE).[9]We cluster the student-level error term (ε_{it}) at the school level.

This type of dynamic DD model, or event study (Granger, 1969; Angrist & Pischke, 2009) provides information about the trajectory of effects without imposing restrictive assumptions of two-period DD and related types of models. We test for parallel trends using the estimates of β_{t-1}which compares the immediate pre-treatment value of the dependent variable with the (omitted) twice-lagged value. Later, we discuss the assumptions of DD more formally and consider threats to validity.

IV.B. Matching Process

We follow a two-stage matching process. First, for each Kickboard school in the year they first adopted Kickboard, we identified comparison schools within the same district that had a similar SPS, grade level range (e.g., K-5) and that never used Kickboard, which we call our donor pool. Second, within the donor pool, we matched each student attending a Kickboard school to another student in a comparison school based on their propensity score as follows:

$$K_{ikt} = \beta_0 + \beta_1 Y_{it-1} + \beta_2 Y_{it-2} \quad (2)$$

where K_{ikt} is a treatment indicator for student i in school k in time t and is equal to one if school k is a Kickboard school and zero otherwise. Y_{it-1} and Y_{it-2} are lagged values of the outcome of interest. We construct a separate comparison group for each outcome. For example, if we are examining the effect of attending a Kickboard school on math

[8] These include race, free/reduced price lunch status, special education status, limited English proficiency, and grade repetition. In addition, we include bin indicators for each stratum in the matching process.
[9] Athey and Imbens (2002) discuss additional linearity assumptions used in DD estimation.

achievement, Y_{it-1} and Y_{it-2} will be students' prior math scores. The estimation of equation (3) yields a predicted probability of being in a treatment school for each student. We match treatment students to control students using this propensity score using a nearest neighbor approach (without replacement). The matching is blocked by grade and the same number of school moves in the prior two years, the latter of which accounts for possible disruption effects as students may be entering a new school as it adopts kickboard.

The matching process greatly improved the baseline equivalence of the comparison and treatment groups. Columns 1 through 3 of Table 2 report the average student's characteristics for Kickboard schools, donor group schools, and schools in the matched comparison group, respectively. As noted above, with matching, the comparison group is at a similar level across all demographic and outcome measures.

The standardized difference provides a measure of that overlap by comparing the means of two groups in units of the pooled standard deviation. The standardized difference is measured for continuous variables using:

$$d = \frac{(|\bar{x}_{tr} - \bar{x}_c|)}{\sqrt{\frac{s_{tr}^2 + s_c^2}{2}}} \quad (3)$$

where $\bar{x}_{tr}$ and $\bar{x}_c$ represent the sample mean for a given variable in the treatment group and comparison group, respectively, and s_{tr}^2 and s_c^2 represent the sample variance for a given variable in the treatment group and comparison group, respectively. The standardized difference for dichotomous variables is measured using:

$$d = \frac{|\hat{p}_{tr} - \hat{p}_c|}{\sqrt{\frac{\hat{p}_{tr}(1-\hat{p}_{tr}) + \hat{p}_c(1-\hat{p}_c)}{2}}} \quad (4)$$

where $\hat{p}_{tr}$ and $\hat{p}_c$ represent the sample mean of a given variable in the treatment and comparison groups, respectively. The standardized difference is not sensitive to the units of measure and can be compared across different outcomes. Following Normand et al. (2001), we consider standardized differences less than 10 to indicate a negligible difference between the two groups for the respective pre-treatment outcome.[10] Table 2 Columns 4 and 5 report the standardized differences for both comparisons. There are notable differences between Kickboard students and students in the donor pool in all pre-treatment outcomes except for days of discipline and total suspensions in the year prior to treatment. Matching greatly reduces these differences suggesting that the treatment and comparison groups are similar in pre-treatment outcomes and demographics.

IV.C. Threats to Validity

The main assumption of DD analysis is that the treatment group would have followed the same path as the comparison group in the absence of treatment. The matching process we utilize is meant to address this concern since we match on the two prior years of the pre-treatment outcome. We present the formal test of the parallel trends assumption in our main analyses.

The main threat to validity of a causal interpretation in this analysis is that schools adopted Kickboard simultaneously with other policies and practices, as part of broader school improvement efforts. For example, if schools were trying to improve student behavior, they might also have provided additional professional development to teachers unrelated to Kickboard, or even to PBIS. We cannot observe these other programmatic

[10] A standardized difference of 0.1 suggests that the percentage of non-overlap between the distributions of a variable for the two groups being compared is 7.7%. This value also suggests that, when the populations are equally sized, group membership explains .25% of the variance of the variable (Austin, 2009).

changes. Alternatively, schools may have decided to become less strict at the same time as adopting Kickboard, suspending students less frequently. We can test for this in two ways: (a) comparing the results for violent offenses, which are likely to result in suspensions or expulsions under any policy, and non-violent offenses, which might be more sensitive to a change in disciplinary strictness; and (b) comparing the results for schools that technically adopted Kickboard, but rarely used it, and other schools that used Kickboard more intensively. If the effects are due to Kickboard then we should expect the effects to be roughly proportional to the intensity of implementation.

Additional threats include regression to the mean (Daw & Hatfield, 2018) and functional form assumptions implicit in the DD methodology (Kahn-Lang & Lang, 2018). These are relatively easy to address with our data and we see no evidence that these latter threats affect the results.

V. Implementation

We start by describing how schools use Kickboard, including both the intensity (number of teacher logins) and type of use (types of behavioral codes used and whether teachers added comments) using the first year of implementation for each school (t=0). Figure 1 shows the average number of logins per teacher per year, by school. For example, the first bar on the far left shows that 10 percent of schools had an average of 1-10 logins per teacher annually (low intensity). At the high end, we have schools with more than 400 logins per teacher, meaning that each teacher logged in an average of about twice each day. The average number of logins is 89 per teacher per year, or about one every other school day.

The distribution of logins is similar for school administrators (average number of logins: 91). Note that administrators' use is fundamentally different in that they are not generally inputting data, but rather monitoring the data that teachers are entering. Therefore, even though the distribution of logins is similar, administrator use can be viewed as more intense.

Another way to measure intensity of use is through the number of marks per student. Panel C reports the percentage of schools with different numbers of marks per student. The right-hand bar indicates that one school averaged almost 400 marks per student annually, or roughly two per school day.

Schools can select the menu of student behaviors that teachers can choose from and direct teachers to focus on specific behaviors. To measure the degree to which schools focus on specific behaviors, and how this varies across schools, Panel D shows the number of unique codes schools use to comprise 80 percent of all marks. So, a school that only marks one kind of behavior, such as student "caring," would be in the leftmost bar with 0-10 codes. Conversely, a school that marks a large number of codes/behaviors, all with similar frequency, would show up in the right-hand bars. We find that the vast majority of schools use between one and 15 codes frequently. While not shown, we also calculated the total number of codes teachers used. Only 19 codes are pre-programed into Kickboard and part of the initial training is to show teachers how to add their own codes. Teachers apparently value this feature since we calculated more than 400,000 different codes. Many of these closely overlap one another, but it is noteworthy that teachers go to the trouble of creating their own coding schemes.

Teachers can also add nuance to the marks and codes with text comments. For example, a teacher could indicate that a student displayed "caring" and stop there (no comment) or explain exactly what the student did; teachers might use this additional information to include in report cards or IEPs, report informally back to parents, or other purposes. Panel E reports the percentage of behavior marks that also include text comments. The first two bars in the panel indicate that about 19 percent of schools almost never used comments. At the other end of the spectrum, four percent of schools had text comments with almost every student mark.

Every behavior code is associated with a positive or negative indicator. Kickboard recommends that teachers have a ratio of three positive marks to every one negative mark. Panel F shows the ratio of positive to negative behaviors.[11] Since the approach is *Positive* Behavior Interventions and Supports (emphasis added), and Kickboard only includes positive behaviors in its initial menu of options, we expect the ratio of positive-to-negative marks to be generally greater than one. This is true for a majority of schools with about 30 percent of schools having a ratio of less than one. These schools focusing on negative behaviors might be using Kickboard for discipline purposes; suspensions and expulsions are usually the result of specific negative behaviors rather than the absence of positive ones.

VI. Average Treatment Effects

Tables 3 and 4 provide average treatment effect estimates from the coefficients on *KBS*Treatment Year* for achievement and discipline, respectively. Table 3 provides some

[11] Examples of negative behaviors include not following directions, missing homework, disruptive behavior, talking, and unprepared for class.

evidence of positive effects on achievement with five of the 12 coefficients being positive and statistically different from zero; there are two negative point estimates but they are not statistically significant. All six coefficients on suspensions and suspension days for all infractions (Panel A of Table 4) indicate statistically significant reductions on those measures. For all outcomes, the coefficient on *KBS*1 Year Prior* are all small and statistically insignificant, meaning that they pass a parallel trends test.

When we restrict the sample to only include infractions for violent behaviors the coefficients on *KBS*Treatment Year* are significant and negative in the third post-treatment year. This provides mixed evidence on the potential influence of reporting effects and discipline strictness. If we focus only on the third post-treatment year, then we might conclude that the results are not driven by reporting effects, but the first two post-treatment years do suggest reporting effects. An alternative interpretation is that our estimated effects reflect actual improvements in student non-violent behavior, but that PBIS does not reduce violent behavior in the short term.

In Table 1, we showed that different schools have different number of years of data, which means that the coefficients on the various treatment years in Panel A of Tables 3 and 4 are based on different samples of schools. Panel B in both tables therefore provides results for only the balanced panel (schools with all five years of data). The effects on all suspensions and suspension days are robust to this sample restriction, but the effects on violent offenses and achievement effects become less positive and more erratic. (Note that the Panel B estimates for violent infractions involve both very few infractions and a small sample of schools, which likely explains their volatility.) Given

their sensitivity to sample restrictions, we have somewhat less confidence in the achievement results we showed in Panel A in Tables 3 and 4.

In Panel C we focus only on New Orleans schools, which comprise a bit more than half of the treatment group and which operate in a unique charter-based schooling environment. These results largely mirror Panel A for the entire sample; however, we often reject the parallel trends assumption with the New Orleans sample. We also tested whether the intensity of implementation differed between the two sectors, but found no difference. The appendix shows the results using figures (for the unbalanced panel).

VII. Effect Heterogeneity

VII.A. Effects by Implementation

One of our threats to validity is schools that adopt Kickboard may simultaneously change the way in which they report student behaviors through disciplinary actions. While we addressed this partially in the analysis above through the analysis of violent behaviors, we go further here using implementation heterogeneity. If Kickboard adoption were driving our results, as opposed to actual Kickboard implementation, then we would expect to see the earlier average treatment effects replicated when using the schools that were adopters but not active implementers. We identify a school as a non-active implementer if it had less than 20 entries per teacher in the first year of Kickboard adoption. (We limit to the first year to maintain a consistent sample of schools, and because schools can move in and out of implementation categories.)

Table 5 presents the results for both achievement and discipline outcomes. We see some evidence of decreases in student discipline rates and days in the second and third year as well as an increase in ELA and science achievement in the third year. For all six

outcomes, we pass the parallel trends test and there are no effects in the first year of Kickboard. This suggests that the results from our main analyses are driven by the use of Kickboard as opposed to changes in discipline reporting associated with Kickboard adoption.

As a further robustness check, we also estimated effects by relative level of implementation, comparing low implementers to high implementers. If more use of Kickboard yields larger effects then we might to see larger effects in treatment schools that used it relatively intensely. We measure implementation using an index of the implementation measures in Figure 1, a simple average of the relative ranking within each of the following: the number of teacher logins, administrator logins, average number of incidents logged per student, unique behavior codes used, incidents with comments, and positive/negative ratio. After placing schools into quartiles based on this index, we define "low implementers" as those in the bottom quartile and "high implementers" as those in the top quartile (omitting the middle two quartiles). We focus on the full, unbalanced panel of students and schools since the balanced panel, combined with the effect heterogeneity analysis would otherwise create samples that are too small. We see no evidence that the effects differ between the low and high implementers. Even where the effects for the low or high groups are statistically different from zero, they are never close to begin statistically different from one another (see the appendix for corresponding tables A3-A4 and figures A5-A6). Also, note that all the subgroups pass their respective parallel trends tests.

VII.B. Effects by Student Subgroup

Given disparities in discipline across student groups (Barrett, McEachin, Mills & Valant, 2017), effect heterogeneity by race, income, and gender are particularly salient, but we also considered effects by achievement and prior discipline records. These analyses are complicated by the general homogeneity of the Kickboard schools. For example, the vast majority of Kickboard schools are overwhelmingly black and low-income so that there are few students from their counterpart groups (white and higher-income) within Kickboard schools with which to compare.

Initial student performance/outcomes are the only student characteristics where we observe clear differences across subgroups. Low-performing students experienced less positive effects on student achievement than high-performing students. This suggests that Kickboard may have the effect of reducing negative peer spillovers; that is, reducing the amount of disruption in classrooms that affect high-performing students.

When we separate students based on the number of pre-Kickboard suspensions, the results are erratic across years. Additional evidence on effect heterogeneity by students' special education (IEP) status are shown in the appendix, though these analyses usually fail the parallel trends tests.

VIII. Conclusion

Promoting positive student behavior in schools, and preventing negative behavior, is one of the most important tasks of educators. Establishing good habits can improve the learning environment and allow teachers to spend more time on the core tasks of instruction, as opposed to classroom management. If students display fewer negative

behavior, they will be suspended less, and perhaps learn more. Student behavior is also an outcome itself, as reflected in the current national interest in socio-emotional learning.

PBIS attempts to improve student behavior by capturing detailed data on behavior, providing rewards for positive behavior, and using the data, through professional development and school leadership, to build effective classroom management and strong school cultures and learning environments. The small number of rigorous studies to date on PBIS suggest that the strategy improves a variety of student outcomes.

Our research on Kickboard is largely consistent with this prior evidence. The program appears to have reduced the number of suspensions by 0.14-0.38 per student per year (26-72 percent from baseline) and the number of suspension days by 0.7-1.5 (at least 52 percent). A causal interpretation is partially reinforced by additional evidence of reductions in suspensions for violent behavior and that we only see effects in schools that implemented the program to some degree.

This study adds to the literature in several ways. First, it uses a rigorous quasi-experimental design that provides plausibly causal effects. Second, prior studies have not considered the possibility that effects on measured incidents may reflect reporting effects rather than behavior effects. Future research should attempt to address the remaining limitations by carrying out qualitative analysis to better understand the mechanisms through they operate (e.g., by increasing time on task) and using larger samples of schools to allow for more convincing effect heterogeneity analysis.

Based on what we know today, and when viewed in conjunction with other studies, PBIS programs and data platforms such as Kickboard appear promising and warrant further study.

References

Angrist, J. D., Pischke, J.S. (2009). *Mostly harmless econometrics: An empiricist's companion*.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, *28*(25), 3083-3107.

Barnum, M. (2016). Data Shows 3 of the 5 Biggest School Districts Hire More Security Officers Than Counselors. *The 74 Million*. Downloaded Ocobter 15, 2018 from: https://www.the74million.org/article/exclusive-data-shows-3-of-the-5-biggest-school-districts-hire-more-security-officers-than-counselors/.

Barrett, N., McEachin, A., Mills, J. N., & Valant, J. (2017). Disparities in Student Discipline by Race and Family Income. New Orleans, LA: Education Research Alliance for New Orleans, Tulane University.

Barrett, S., Bradshaw, C., & Lewis-Palmer, T. (2008). Maryland statewide PBIS initiative: Systems, evaluation, and next steps. *Journal of Positive Behavior Interventions*, 10, 105–114.

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly*, *23*(4), 462.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, *12*(3), 133-148.

Bradshaw, C. P., Pas, E. T., Goldweber, A., Leaf, P. J., & Rosenberg, M. S. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams: The PBISplus model. *Advances in School Mental Health Promotion*, 5, 177–193.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2015). Examining variation in the impact of school-wide positive behavioral interventions and supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology*, 107, 546.

Bross, W. & Harris, D.N. (2016). *The Effects of Performance-Based School Closure and Charter Takeover*. New Orleans, LA: Education Research Alliance for New Orleans, Tulane University.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological bulletin*, *140*(4), 980.

Daw J. R., Hatfield L.A. (2018) Matching and regression to the mean in difference-in-differences analysis. *Health Services Research*. In Press.

Deci, E. L. (1975). Intrinsic motivation. New York, NY, US.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of educational research*, *71*(1), 1-27.

Eber, L. (2006). Illinois PBIS evaluation report. LaGrange Park: Illinois State Board of Education, PBIS/EBD Network.

Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. Teacher Education and Special Education, 37(2), 106-120.

Gottfredson, D. C., Gottfredson, G. D., & Hybl, L. G. (1993). Managing adolescent behavior: A multiyear, multischool study. *American Educational Research Journal*, 30, 179–215.

Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica, 37*(3), 424-438. doi:10.2307/1912791

Hidi, S. (2016). Revisiting the role of rewards in motivation and learning: Implications of neuroscientific research. *Educational Psychology Review*, *28*(1), 61-93.

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2005). School-wide positive behavior support. In L. Bambara & L. Kern (Eds.), *Individualized supports for students with problem behaviors: Designing positive behavior plans*. New York: Guilford.

Horner, R.H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A.W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions* 11(3), 133-144.

Kahn-Lang, A., Lang, K. (2018) The Promise and Pitfalls of Differences-in-Differences: Reflections on '16 and Pregnant' and Other Applications. National Bureau of Economic Research Working Paper Series. Retrieved from http://www.nber.org/papers/w24857

Kellam, S. G., Mayer, L. S., Rebok, G. W., & Hawkins, W. E. (1998). The effects of improving achievement on aggressive behavior and of improving aggressive behavior on achievement through two prevention interventions: An investigation of causal paths. In B. Dohrenwend (Ed.), Adversity, stress, and psychopathology (pp. 486–505). Oxford, UK: Oxford University Press.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the" overjustification" hypothesis. *Journal of Personality and social Psychology*, *28*(1), 129.

Lohrman-O'Rourke, S., Knoster, T., Sabatine, K., Smith, D., Horvath, B., & Llewellyn, G. (2000). School-wide application of PBS in the Bangor Area school district. *Journal of Positive Behavior Interventions*, 2(4), 238–240.

Losen, D. J., Martinez, T. E., & Okelola, V. (2014). *Keeping California's kids in school: Fewer students of color missing school for minor misbehavior.* Los Angeles, CA: The Center for Civil Rights Remedies, University of California.

Loveless, T. (2017). *The 2017 Brown Center report on American education: How well are American students learning?* Washington, DC: Brookings Institution.

Luiselli, J. K., Putnam, R. F., & Sunderland, M. (2002). Longitudinal evaluation of behavior support intervention in a public middle school. *Journal of Positive Behavior Interventions*, 4(3), 182–188.

McIntosh, K., Mercer, S. H., Nese, R. N., Strickland-Cohen, M. K., & Hoselton, R. (2016). Predictors of sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, 18, 209–218. doi:10.1177/1098300715599737

Mitchell, B.S., Hatton, H., & Lewis, T.J. (2018). An examination of the evidence-base of school-wide positive behavior intervention and supports through two quality appraisal processes. *Journal of Positive Behavior Interventions*, p. 1-12.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning: The effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorder*, 10(3) 136–148.

Normand, S. L. T., Landrum, M. B., Guadagnoli, E., Ayanian, J. Z., Ryan, T. J., Cleary, P. D., & McNeil, B. J. (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *Journal of clinical epidemiology*, *54*(4), 387-398.

OSEP Center on Positive Behavioral Interventions, Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., ... & Turnbull, A. P. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of positive behavior interventions*, *2*(3), 131-143.

Putnam, R. F., Handler, M. W., Ramirez-Platt, C. M., & Luiselli, J. K. (2003). Improving student bus-riding behavior through a whole-school intervention. *Journal of Applied Behavior Analysis*, 36, 583–590.

Reiss, S. (2005). Extrinsic and intrinsic motivation at 30: Unresolved scientific issues. *The Behavior Analyst*, *28*(1), 1-14.

Skiba, R., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of actors predicting ISTEP scores. Children Left Behind Policy Briefs, Supplementary analysis 2-D.* Bloomington, IN: Center for Evaluation and Education Policy, Indiana University. Retrieved from http://files.eric.ed.gov/fulltext/ED488899.pdf

Skiba, R. J., & Williams, N. T. (2014). *Are black kids worse? Myths and facts about racial differences in behavior: A summary of the literature.* Bloomington, IN: The Equity Project at Indiana University. Retrieved from http://www.indiana.edu/~atlantic/wp-content/uploads/2014/03/African-American-Differential-Behavior_031214.pdf

Sprague, J., Walker, H., Golly, A., White, K., Myers, D. R., & Shannon, T. (2001). Translating research into effective practice: The effects of a universal staff and student intervention on indicators of discipline and school safety. *Education and Treatment of Children*, 24, 495–511.

Sugai, G. (2018). *Positive Behavioral Interventions and Supports: Questions and Answers*. Downloaded October 15, 2018 from: https://www2.ed.gov/documents/press-releases/fssc-field-visit-sugai.pdf.

Taylor-Greene, S., Brown, D., Nelson, L., Longton, J., Gassman, T., Cohen, J., et al. (1997). School-wide behavioral support: Starting the year off right. *Journal of Behavioral Education*, 7, 99–112.

Taylor-Greene, S. J., & Kartub, D. T. (2000). Durable implementation of school-wide behavior support: The high five program. *Journal of Positive Behavior Interventions*, 2(4), 233–235.

U. S. Department of Justice, & U.S. Department of Education. (2014). *Dear Colleague Letter on the nondiscriminatory administration of school discipline.* Retrieved from https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html

Ward, B., & Gersten, R. (2013). A randomized evaluation of the safe and civil schools model for positive behavioral interventions and supports at elementary schools in a large urban school district. *School Psychology Review*, 42, 317–333.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic Monthly*, 249(3), 29-38.

Table 1: Number of schools by years in the sample

Number of Schools			Treatment Period				
Total	NOLA	Non-NOLA	t-2	t-1	t	t+1	t+2
28	12	16	x	x	x	x	x
12	11	1		x	x	x	x
8	5	3	x	x	x	x	
8	7	1			x	x	x
7	3	4	x	x	x		
2	0	2		x	x	x	
1	1	0			x	x	
4	1	3		x	x		
0	0	0			x		

Notes: The first three columns report the number of Kickboard schools in total and by location with New Orleans (NOLA) and others within Louisiana (non-NOLA). Each row indicates the number of years the schools were opened and for which we have data. For example, the first row indicates that 28 schools in total had data two years pre-treatment (t-2) all the way through two years post-treatment (t+2), where t denotes the first year of Kickboard implementation. The second row indicates that 12 schools had data only starting one year prior to treatment (t-1), and so on.

Table 2: Baseline Equivalence of Mean Student Characteristics in Year Prior to Kickboard Adoption

	Year Prior to Kickboard Adoption			Standardized Differences	
	(1)	(2)	(3)	(4)	(5)
	Kickboard	Comparison (All)	Comparison (Matched)	Balance Pre-Match	Balance Post-Match
Outcome Variables					
Math	-0.335	-0.075	-0.350	**24.82**	1.47
ELA	-0.385	-0.055	-0.390	**32.25**	0.51
Science	-0.439	-0.119	-0.438	**30.98**	0.10
Social Studies	-0.323	-0.079	-0.322	**23.27**	0.10
Total Suspensions	0.525	0.351	0.488	**13.46**	2.75
Total Days of Discipline	1.350	0.887	1.285	**12.01**	1.57
Control Variables					
Male	0.512	0.510	0.516	0.33	0.65
Black	0.835	0.620	0.841	**42.59**	1.11
White	0.074	0.265	0.077	**46.74**	0.62
Other Race	0.091	0.115	0.098	6.80	1.68
FRL	0.869	0.693	0.864	**37.51**	1.20
IEP	0.119	0.114	0.115	1.27	1.01
Gifted	0.025	0.047	0.026	**10.17**	0.52
ELL	0.036	0.048	0.033	5.00	1.33

Notes. The reported means are from the pre-treatment period for all groups. The Kickboard group includes all students in Kickboard schools in year t. The sample size for this group ranges from 10,035 to 9,977 for test outcomes and is 13,725 for all other variables. The donor pool (comparison-all) column includes all students in districts that had at least one school participate in Kickboard in year t, whose sample size ranges from 70,895 to 70,432 for test outcomes and is 96,065 for all other variables. The comparison-matched group includes students matched to a Kickboard student in year t, whose sample size matches the sample size of the Kickboard group.

Table 3: Effects of Kickboard Treatment on Student Outcomes

	Math	ELA	Science	Social Studies
Panel A: All Schools				
Kickboard Student (KBS)	0.004	-0.008	-0.010	-0.019
	(0.014)	(0.014)	(0.015)	(0.016)
KBS*1 Year Prior	0.019	0.011	0.011	0.025
	(0.018)	(0.018)	(0.019)	(0.021)
KBS*1st Treatment Year	0.041**	0.030*	0.027	-0.023
	(0.019)	(0.018)	(0.019)	(0.021)
KBS*2nd Treatment Year	0.045**	0.029	0.010	0.001
	(0.019)	(0.018)	(0.020)	(0.021)
KBS*3rd Treatment Year	-0.004	0.036*	0.039*	0.029
	(0.021)	(0.020)	(0.022)	(0.023)
N	70,218	70,226	69,146	69,014
Panel B: Schools with all Five Years of Data				
Kickboard Student (KBS)	0.003	0.079	0.130**	0.021
	(0.051)	(0.048)	(0.052)	(0.057)
KBS*1 Year Prior	0.084	-0.001	-0.040	0.066
	(0.062)	(0.059)	(0.064)	(0.070)
KBS*1st Treatment Year	0.024	-0.070	-0.095	-0.100
	(0.056)	(0.053)	(0.058)	(0.064)
KBS*2nd Treatment Year	0.058	-0.057	-0.165***	-0.032
	(0.055)	(0.053)	(0.057)	(0.063)
KBS*3rd Treatment Year	0.077	-0.046	-0.094	0.031
	(0.059)	(0.056)	(0.061)	(0.066)
N	20,220	20,236	20,008	19,984
Panel C: NOLA Schools				
Kickboard Student (KBS)	-0.007	-0.024	-0.052**	-0.042*
	(0.020)	(0.019)	(0.021)	(0.023)
KBS*1 Year Prior	0.030	0.046*	0.097***	0.090***
	(0.026)	(0.025)	(0.027)	(0.030)
KBS*1st Treatment Year	0.060**	0.045*	0.064**	-0.004
	(0.027)	(0.026)	(0.028)	(0.030)
KBS*2nd Treatment Year	0.078***	0.055**	0.060**	0.034
	(0.027)	(0.026)	(0.028)	(0.030)
KBS*3rd Treatment Year	0.021	0.036	0.096***	0.059*
	(0.028)	(0.027)	(0.028)	(0.031)
N	38,672	38,670	37,752	37,674

Notes. Full results are reported in Appendix Table A1. All models include additional student-level controls for prior discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.

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

Table 4: Effects of Kickboard Treatment on Student Discipline

	All Infractions		Violent Infractions	
	Number of Suspensions	Days of Suspension	Number of Suspensions	Days of Suspension
Panel A: All Schools				
Kickboard Student (KBS)	0.173***	0.608***	0.004	0.061*
	(0.029)	(0.098)	(0.009)	(0.033)
KBS*1 Year Prior	-0.049	-0.173	0.006	0.006
	(0.037)	(0.117)	(0.012)	(0.044)
KBS*1st Treatment Year	-0.140***	-0.702***	0.003	0.009
	(0.033)	(0.111)	(0.012)	(0.044)
KBS*2nd Treatment Year	-0.312***	-0.937***	0.006	-0.029
	(0.033)	(0.113)	(0.013)	(0.046)
KBS*3rd Treatment Year	-0.206***	-0.778***	-0.033**	-0.130***
	(0.034)	(0.116)	(0.014)	(0.050)
N	101,868	101,868	101,868	101,868
Panel B: Schools with all Five Years of Data				
Kickboard Student (KBS)	0.141	0.939***	-0.079**	-0.128
	(0.119)	(0.351)	(0.039)	(0.133)
KBS*1 Year Prior	0.069	-0.427	0.020	-0.121
	(0.141)	(0.418)	(0.049)	(0.165)
KBS*1st Treatment Year	-0.149	-0.961**	-0.013	-0.089
	(0.126)	(0.374)	(0.045)	(0.154)
KBS*2nd Treatment Year	-0.381***	-1.527***	0.105**	0.225
	(0.123)	(0.365)	(0.045)	(0.154)
KBS*3rd Treatment Year	-0.244**	-1.320***	0.070	0.156
	(0.123)	(0.365)	(0.047)	(0.160)
N	28,150	28,150	28,150	28,150
Panel C: NOLA Schools				
Kickboard Student (KBS)	0.196***	0.635***	0.007	0.058
	(0.039)	(0.138)	(0.012)	(0.046)
KBS*1 Year Prior	-0.081	-0.377**	0.017	0.106*
	(0.050)	(0.176)	(0.016)	(0.060)
KBS*1st Treatment Year	-0.246***	-0.745***	0.019	0.112*
	(0.045)	(0.160)	(0.016)	(0.060)
KBS*2nd Treatment Year	-0.258***	-0.759***	0.013	0.003
	(0.048)	(0.171)	(0.016)	(0.060)
KBS*3rd Treatment Year	-0.202***	-0.642***	-0.010	-0.054
	(0.050)	(0.178)	(0.017)	(0.062)
N	52,128	52,128	52,128	52,128

Notes. Full results are reported in Appendix Table A2. All models include additional student-level controls for prior discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Effects of Kickboard Treatment on Student Achievement and Discipline by No Implementation

	Math	ELA	Science	Social Studies	Number of Suspensions	Days of Suspension
Kickboard Student (KBS)	-0.009	-0.033	-0.056	-0.046	0.280***	0.315*
	(0.033)	(0.032)	(0.035)	(0.038)	(0.056)	(0.187)
1 Year Prior	-0.026	-0.067***	-0.030*	-0.017	-0.048**	-0.175***
	(0.017)	(0.017)	(0.018)	(0.019)	(0.020)	(0.066)
1st Treatment Year	-0.009	0.010	0.000	0.049**	-0.091***	-0.284***
	(0.018)	(0.018)	(0.019)	(0.020)	(0.019)	(0.065)
2nd Treatment Year	0.050***	0.057***	0.002	0.038*	-0.121***	-0.348***
	(0.018)	(0.018)	(0.019)	(0.020)	(0.020)	(0.066)
3rd Treatment Year	0.055***	0.065***	0.027	0.034	-0.135***	-0.384***
	(0.020)	(0.020)	(0.021)	(0.023)	(0.020)	(0.069)
KBS*1 Year Prior	0.046	0.043	0.060	0.033	-0.084	-0.250
	(0.045)	(0.044)	(0.047)	(0.051)	(0.074)	(0.248)
KBS*1st Treatment Year	0.008	0.041	0.008	-0.014	-0.032	-0.269
	(0.046)	(0.045)	(0.047)	(0.051)	(0.063)	(0.211)
KBS*2nd Treatment Year	0.052	0.043	0.077	0.055	-0.468***	-0.836***
	(0.046)	(0.045)	(0.048)	(0.052)	(0.062)	(0.208)
KBS*3rd Treatment Year	0.041	0.083*	0.119**	0.079	-0.303***	-0.595***
	(0.047)	(0.045)	(0.048)	(0.052)	(0.063)	(0.212)
N	20,319	20,315	20,815	20,771	29,637	29,637

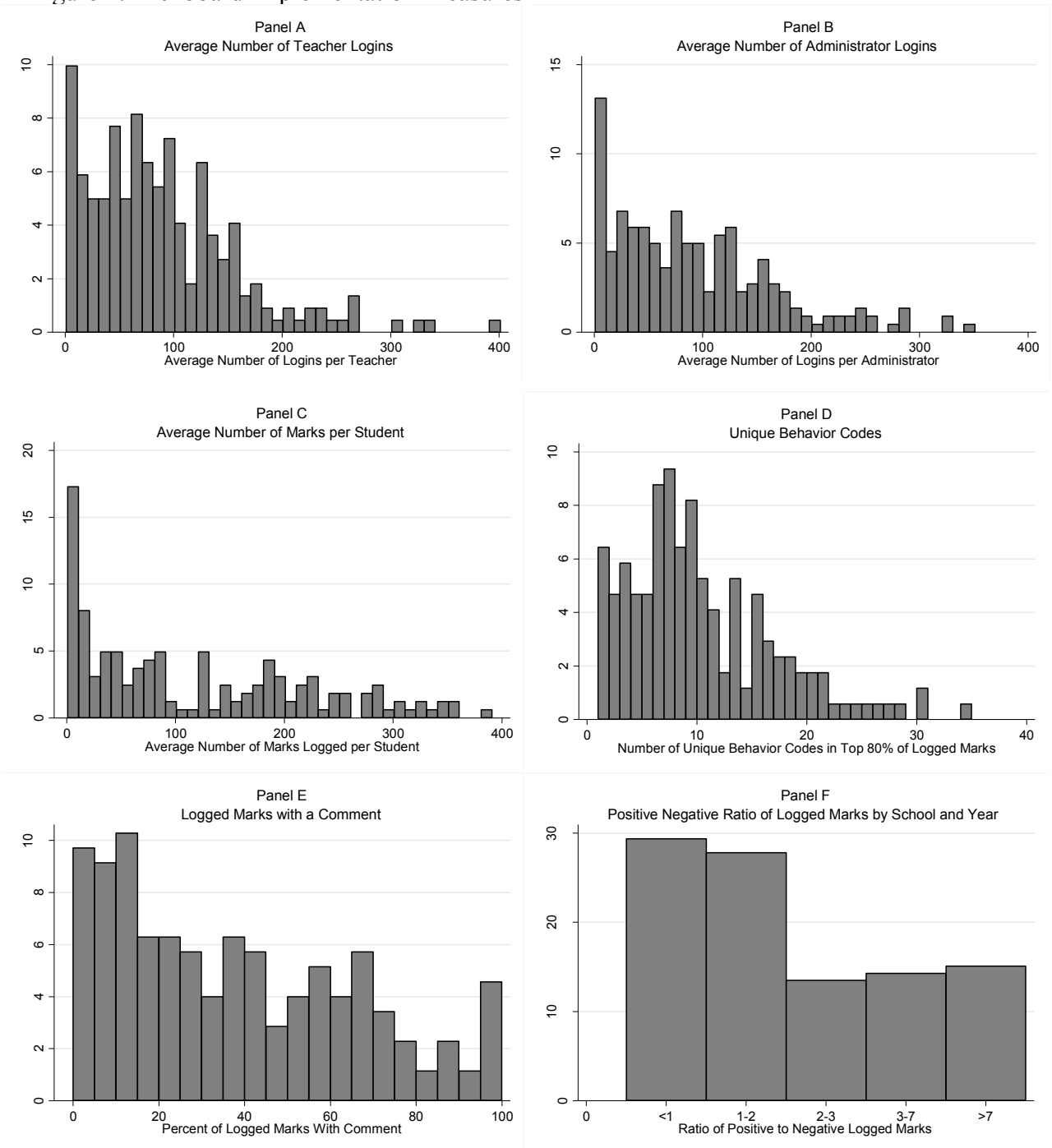
Notes. All models include additional student-level controls for prior achievement, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6: Effects of Kickboard Treatment on Student Outcomes – Low-Performing Students

	Math	ELA	Science	Social Studies	Number of Suspensions	Days of Suspension
Kickboard Student (KBS)	-0.003	-0.017	-0.017	-0.001	0.129***	0.390***
	(0.017)	(0.017)	(0.018)	(0.019)	(0.034)	(0.117)
Low-Performing Student	-0.805***	-0.819***	-0.849***	-0.858***	0.220***	0.740***
	(0.019)	(0.018)	(0.020)	(0.022)	(0.029)	(0.098)
KBS*Low	0.125***	0.131***	0.138***	0.081***	0.113**	0.520***
	(0.026)	(0.025)	(0.027)	(0.029)	(0.056)	(0.191)
1 Year Prior	-0.066***	-0.076***	-0.045***	-0.038**	-0.005	-0.020
	(0.016)	(0.016)	(0.017)	(0.018)	(0.022)	(0.076)
1st Treatment Year	-0.224***	-0.199***	-0.244***	-0.193***	-0.031	-0.083
	(0.015)	(0.015)	(0.016)	(0.017)	(0.020)	(0.069)
2nd Treatment Year	-0.168***	-0.167***	-0.226***	-0.207***	-0.063***	-0.172**
	(0.016)	(0.015)	(0.016)	(0.018)	(0.021)	(0.070)
3rd Treatment Year	-0.161***	-0.138***	-0.220***	-0.224***	-0.064***	-0.196***
	(0.017)	(0.016)	(0.017)	(0.019)	(0.021)	(0.072)
KBS*1 Year Prior	-0.006	-0.009	0.006	-0.007	-0.012	-0.185
	(0.023)	(0.022)	(0.024)	(0.026)	(0.045)	(0.153)
KBS*1st Treatment Year	0.090***	0.085***	0.095***	0.023	-0.107***	-0.467***
	(0.022)	(0.021)	(0.022)	(0.024)	(0.038)	(0.129)
KBS*2nd Treatment Year	0.104***	0.079***	0.071***	0.032	-0.262***	-0.678***
	(0.022)	(0.021)	(0.022)	(0.024)	(0.038)	(0.131)
KBS*3rd Treatment Year	0.033	0.079***	0.073***	0.033	-0.159***	-0.536***
	(0.023)	(0.023)	(0.024)	(0.026)	(0.039)	(0.133)
Low*1 Year Prior	0.148***	0.113***	0.053**	0.053*	-0.081**	-0.339**
	(0.026)	(0.025)	(0.027)	(0.029)	(0.039)	(0.132)
Low*1st Treatment Year	0.567***	0.497***	0.559***	0.506***	-0.138***	-0.526***
	(0.029)	(0.029)	(0.031)	(0.033)	(0.044)	(0.151)
Low*2nd Treatment Year	0.524***	0.493***	0.548***	0.539***	-0.059	-0.357**
	(0.032)	(0.031)	(0.033)	(0.036)	(0.048)	(0.164)
Low*3rd Treatment Year	0.657***	0.546***	0.638***	0.520***	-0.034	-0.157
	(0.038)	(0.037)	(0.039)	(0.043)	(0.052)	(0.177)
KBS*Low*1 Year Prior	0.021	0.018	-0.016	0.037	-0.094	-0.053
	(0.035)	(0.034)	(0.037)	(0.040)	(0.077)	(0.261)
KBS*Low *1st Treatment Year	-0.169***	-0.141***	-0.205***	-0.139***	-0.013	-0.764***
	(0.038)	(0.037)	(0.040)	(0.043)	(0.076)	(0.258)
KBS*Low*2nd Treatment Year	-0.200***	-0.144***	-0.182***	-0.110**	-0.282***	-1.363***
	(0.041)	(0.040)	(0.043)	(0.046)	(0.099)	(0.335)
KBS*Low*3rd Treatment Year	-0.216***	-0.204***	-0.177***	-0.055	0.369	1.732**
	(0.050)	(0.048)	(0.051)	(0.055)	(0.231)	(0.785)
Constant	0.160***	0.055**	0.282***	0.277***	0.139***	0.250**
	(0.023)	(0.022)	(0.024)	(0.026)	(0.036)	(0.123)
N	70,218	70,226	69,146	69,014	101,868	101,868

Notes. All models include additional student-level controls for prior achievement and discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 1: Kickboard Implementation Measures



Appendix Tables and Figures

Figure A1: Effects of Kickboard Treatment of Student Achievement

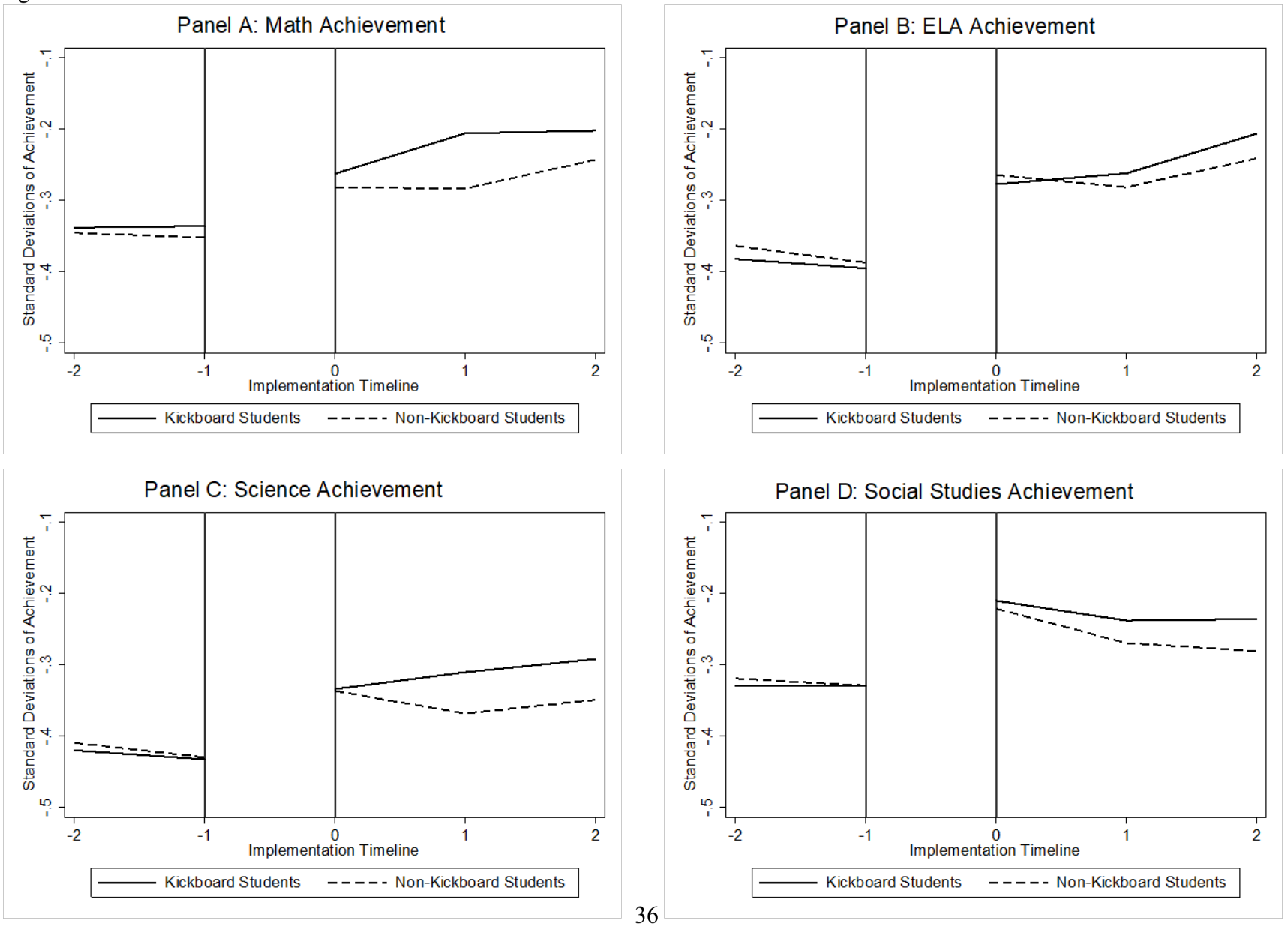


Figure A2: Effects of Kickboard Treatment on Student Discipline

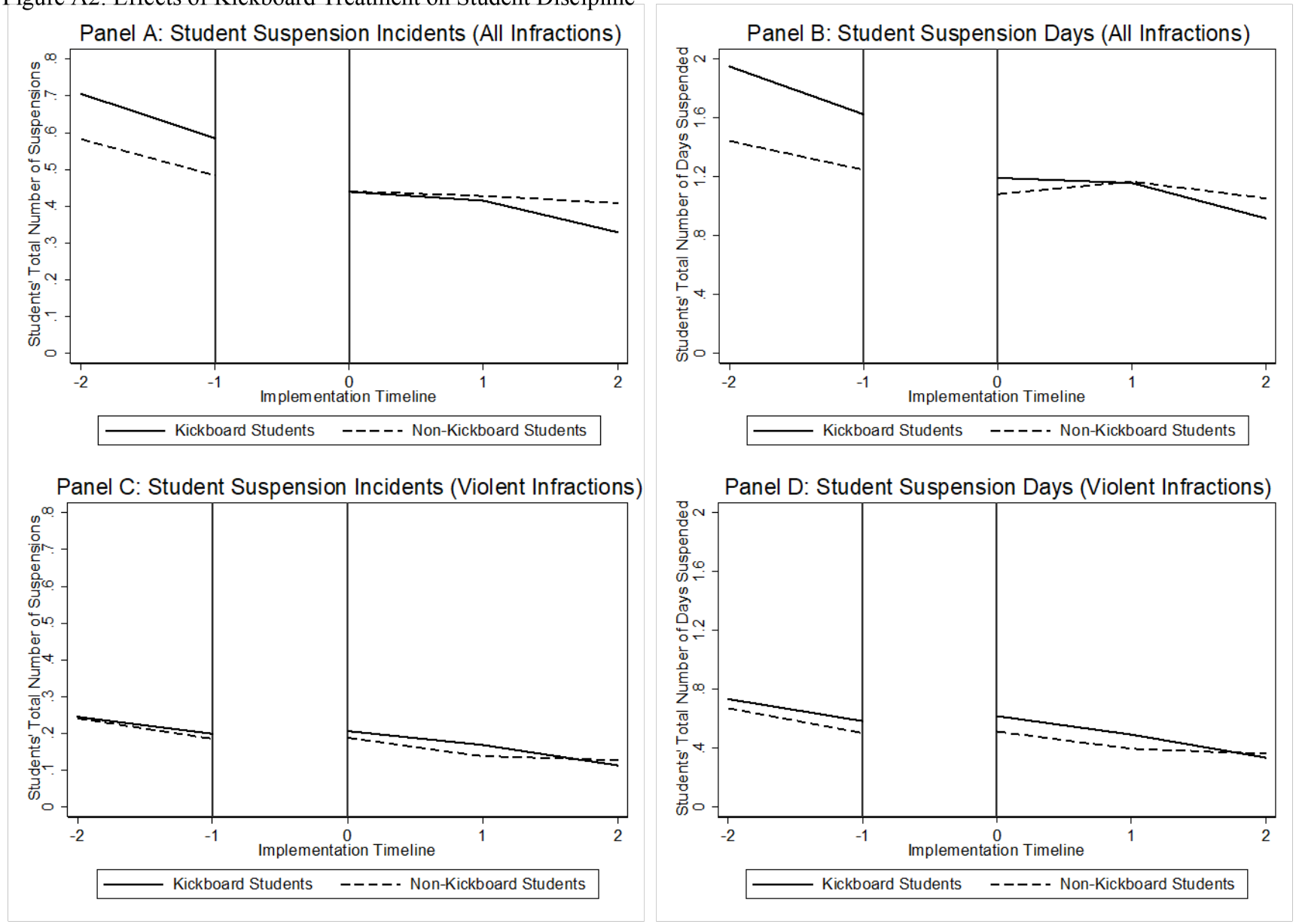


Figure A3: Effects of Kickboard Treatment on Student Achievement by No Implementation

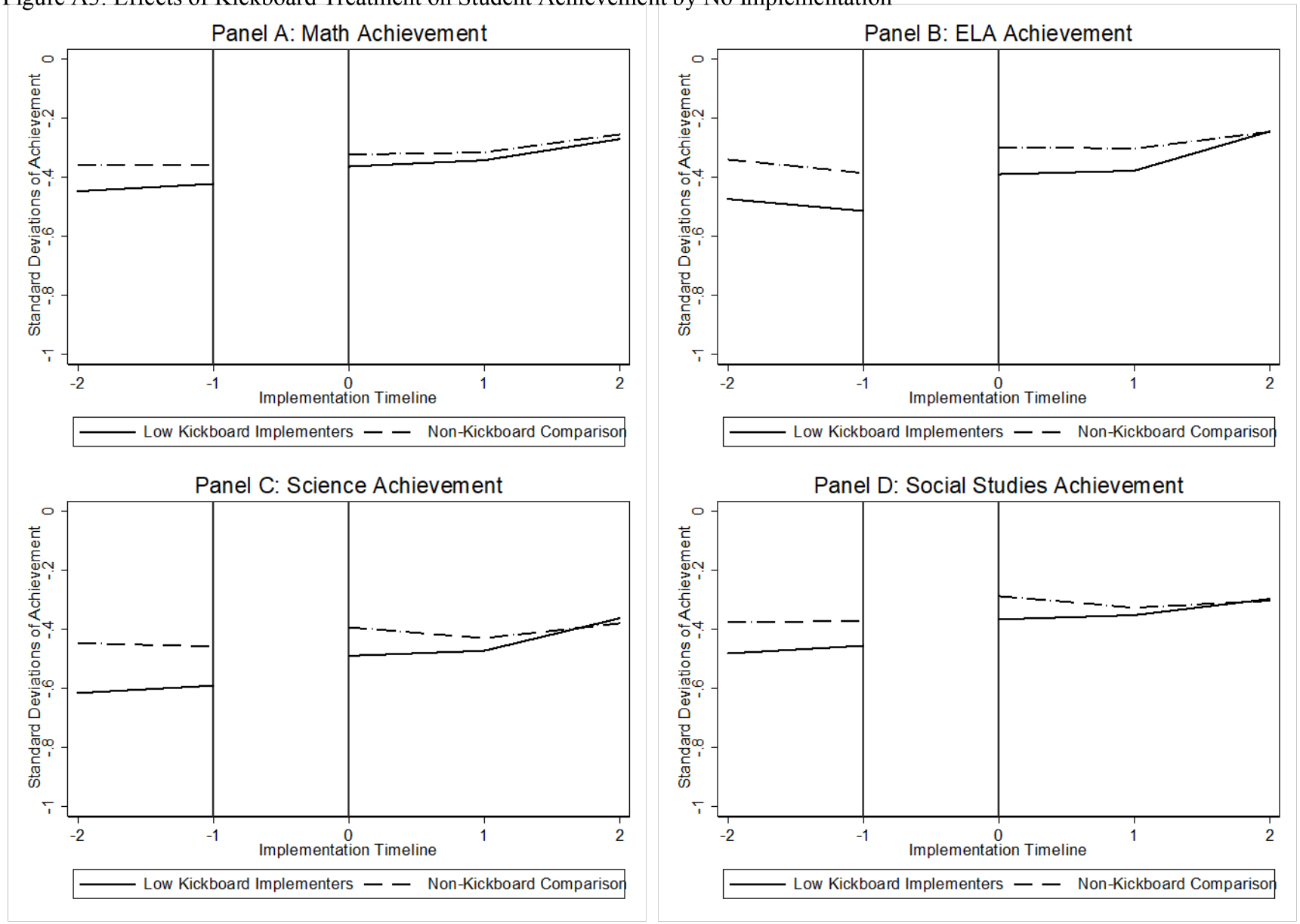


Figure A4: Effects of Kickboard Treatment on Student Discipline by No Implementation

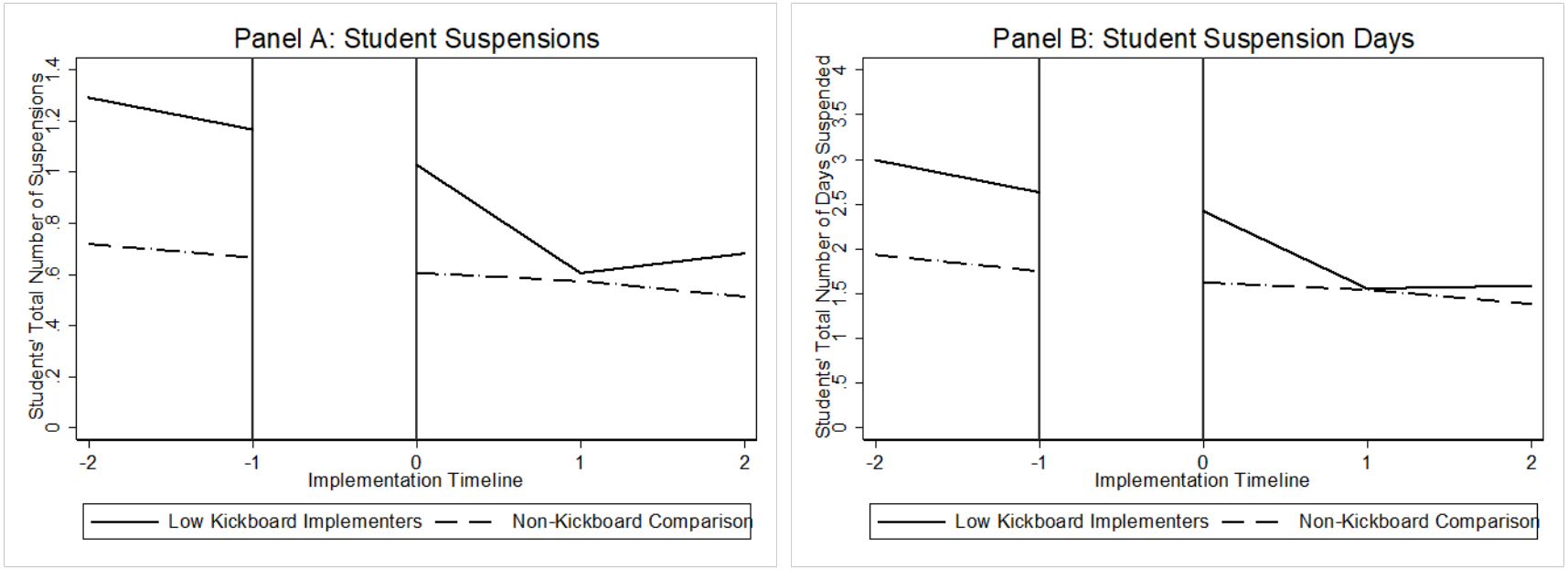


Figure A5: Effects of Kickboard Treatment on Student Achievement by High and Low Implementation

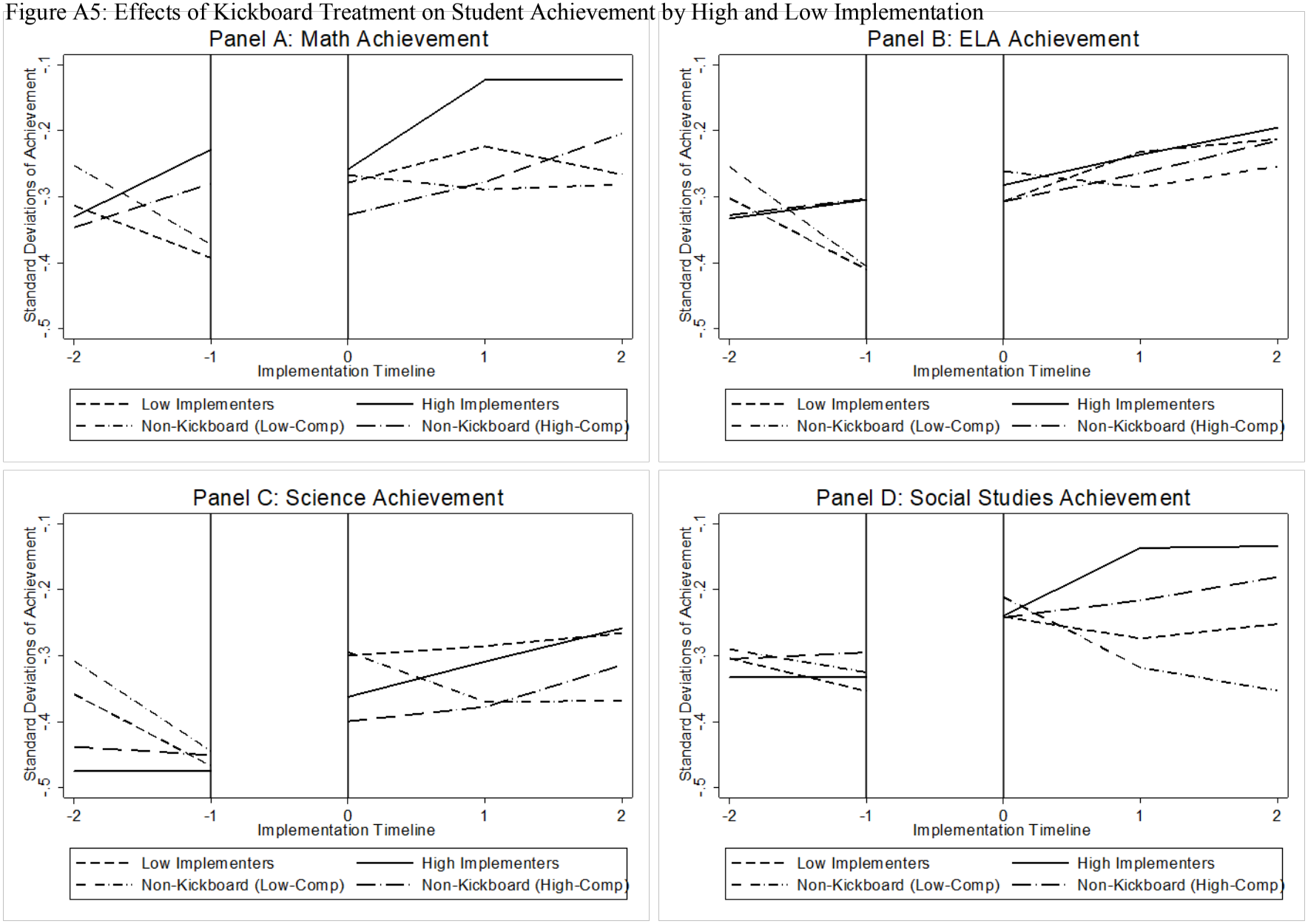


Figure A6: Effects of Kickboard Treatment on Student Discipline by High and Low Implementation

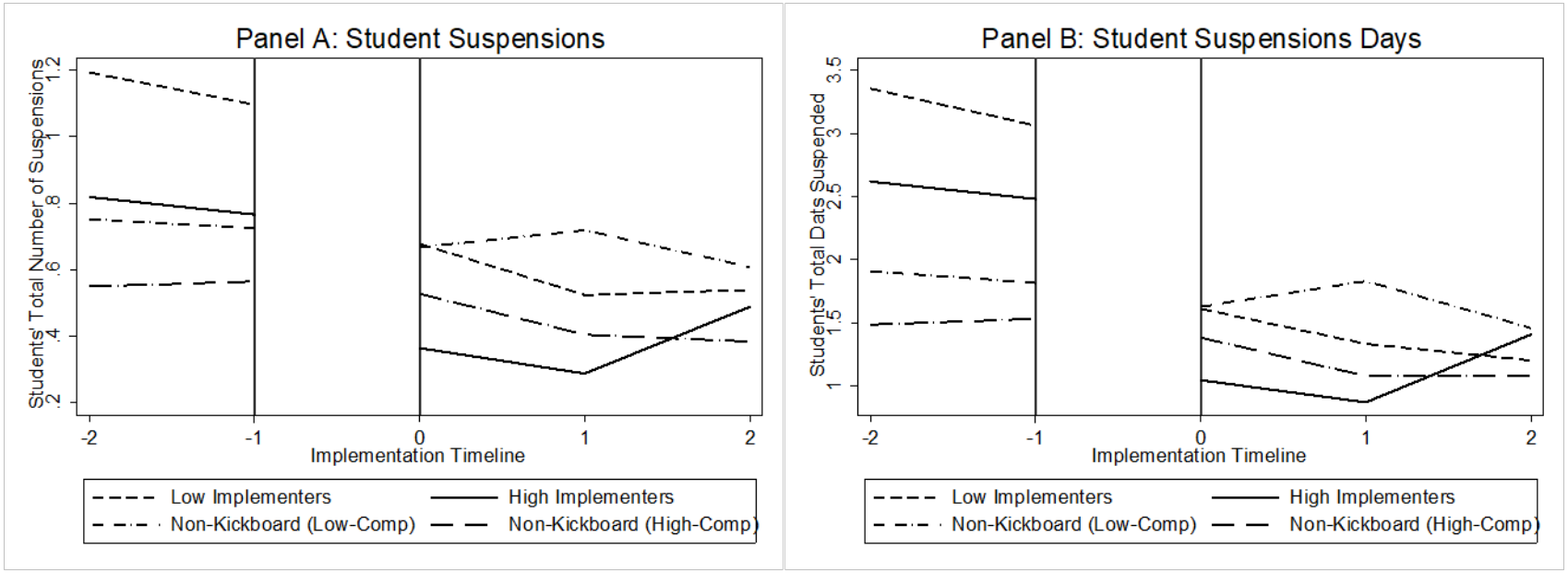


Figure A7: Effects of Kickboard Treatment on Student Achievement – Low-Performing Students

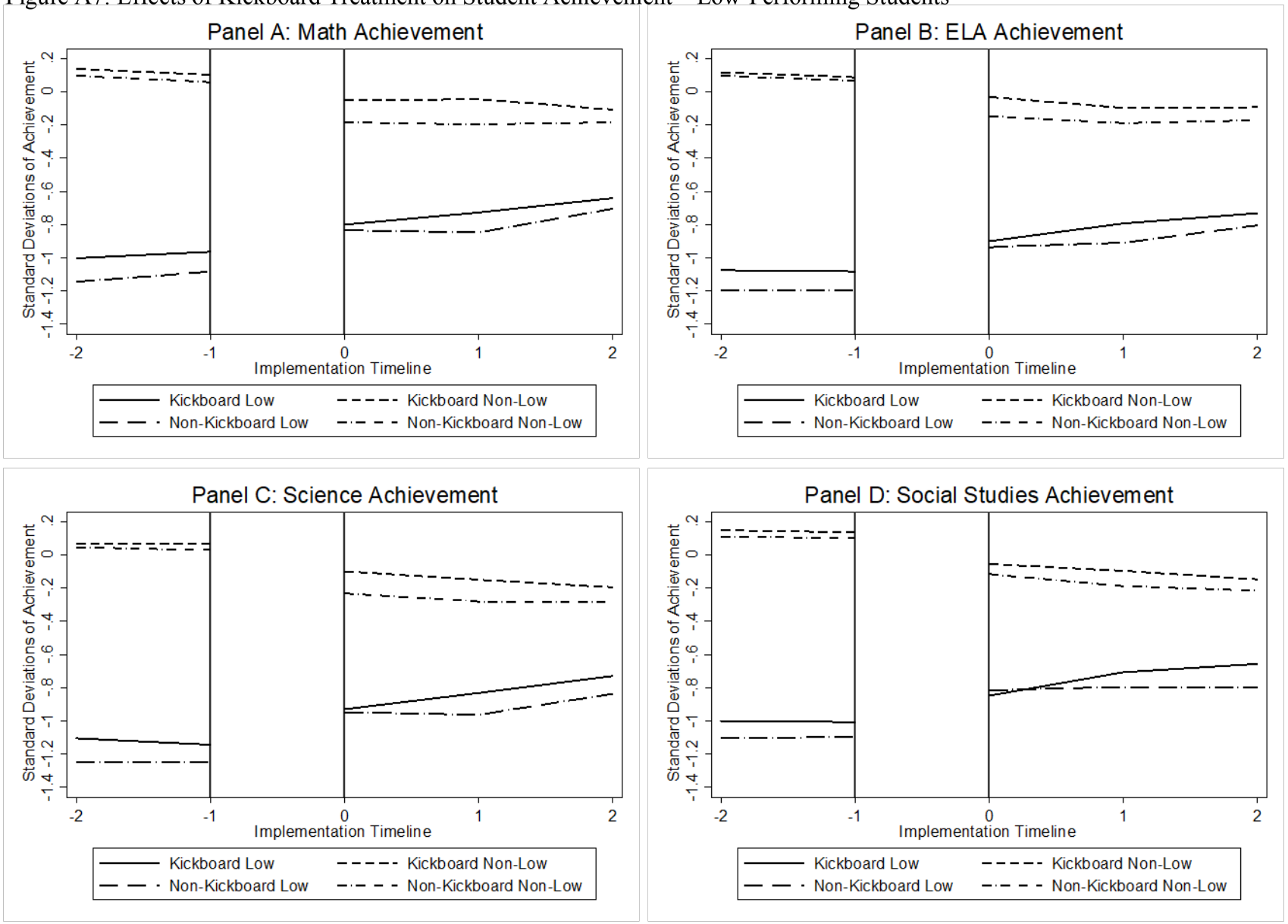


Figure A8: Effect of Kickboard Treatment on Student Discipline – Low-Performing Students

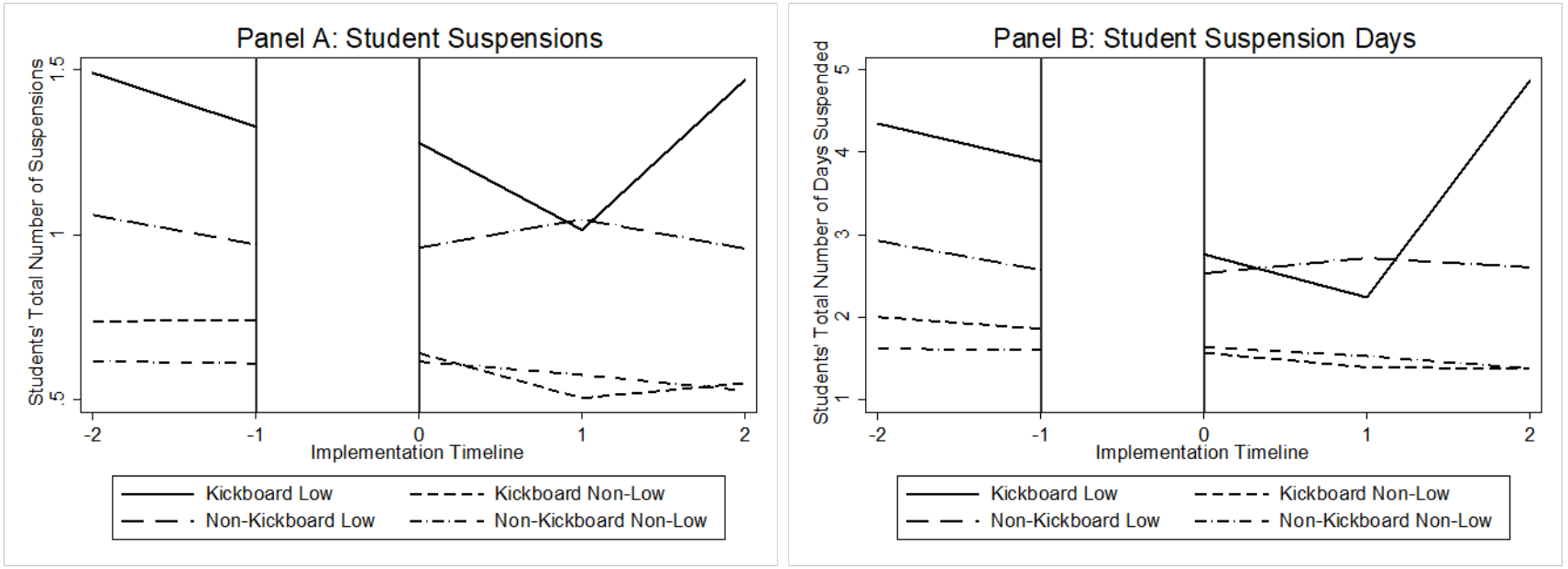


Table A1: Effects of Kickboard Treatment on Student Outcomes

	Math	ELA	Science	Social Studies
Panel A: All Schools				
Kickboard Student (KBS)	0.004	-0.008	-0.010	-0.019
	(0.014)	(0.014)	(0.015)	(0.016)
1 Year Prior	-0.011	-0.031**	-0.025*	-0.017
	(0.013)	(0.013)	(0.014)	(0.015)
1st Treatment Year	0.019	0.041***	0.005	0.050***
	(0.014)	(0.014)	(0.014)	(0.016)
2nd Treatment Year	0.098***	0.105***	0.051***	0.074***
	(0.014)	(0.014)	(0.015)	(0.016)
3rd Treatment Year	0.122***	0.145***	0.071***	0.062***
	(0.016)	(0.015)	(0.016)	(0.018)
KBS*1 Year Prior	0.019	0.011	0.011	0.025
	(0.018)	(0.018)	(0.019)	(0.021)
KBS*1st Treatment Year	0.041**	0.030*	0.027	-0.023
	(0.019)	(0.018)	(0.019)	(0.021)
KBS*2nd Treatment Year	0.045**	0.029	0.010	0.001
	(0.019)	(0.018)	(0.020)	(0.021)
KBS*3rd Treatment Year	-0.004	0.036*	0.039*	0.029
	(0.021)	(0.020)	(0.022)	(0.023)
Constant	-0.036	-0.138***	0.100***	0.099***
	(0.023)	(0.022)	(0.024)	(0.026)
N	70,218	70,226	69,146	69,014
Panel B: Schools with all Five Years of Data				
Kickboard Student (KBS)	0.003	0.079	0.130**	0.021
	(0.051)	(0.048)	(0.052)	(0.057)
1 Year Prior	0.004	-0.034	0.200***	0.106*
	(0.057)	(0.054)	(0.058)	(0.064)
1st Treatment Year	0.094*	0.163***	0.272***	0.265***
	(0.052)	(0.049)	(0.054)	(0.059)
2nd Treatment Year	0.116**	0.191***	0.364***	0.240***
	(0.051)	(0.048)	(0.053)	(0.057)
3rd Treatment Year	0.066	0.186***	0.263***	0.125**
	(0.054)	(0.051)	(0.055)	(0.060)
KBS*1 Year Prior	0.084	-0.001	-0.040	0.066
	(0.062)	(0.059)	(0.064)	(0.070)
KBS*1st Treatment Year	0.024	-0.070	-0.095	-0.100
	(0.056)	(0.053)	(0.058)	(0.064)
KBS*2nd Treatment Year	0.058	-0.057	-0.165***	-0.032
	(0.055)	(0.053)	(0.057)	(0.063)
KBS*3rd Treatment Year	0.077	-0.046	-0.094	0.031
	(0.059)	(0.056)	(0.061)	(0.066)
Constant	-0.469***	-0.770***	-0.197***	-0.094
	(0.060)	(0.057)	(0.062)	(0.067)
N	20,220	20,236	20,008	19,984
Panel C: NOLA Schools				
Kickboard Student (KBS)	-0.007	-0.024	-0.052**	-0.042*
	(0.020)	(0.019)	(0.021)	(0.023)
1 Year Prior	0.006	-0.029	-0.036*	0.012
	(0.019)	(0.018)	(0.020)	(0.021)
1st Treatment Year	-0.044**	-0.026	-0.026	0.046**
	(0.019)	(0.018)	(0.020)	(0.021)
2nd Treatment Year	0.060***	0.037**	0.017	0.081***
	(0.020)	(0.019)	(0.020)	(0.022)
3rd Treatment Year	0.085***	0.082***	0.064***	0.099***
	(0.021)	(0.020)	(0.021)	(0.023)

KBS*1 Year Prior	0.030	0.046*	0.097***	0.090***
	(0.026)	(0.025)	(0.027)	(0.030)
KBS*1st Treatment Year	0.060**	0.045*	0.064**	-0.004
	(0.027)	(0.026)	(0.028)	(0.030)
KBS*2nd Treatment Year	0.078***	0.055**	0.060**	0.034
	(0.027)	(0.026)	(0.028)	(0.030)
KBS*3rd Treatment Year	0.021	0.036	0.096***	0.059*
	(0.028)	(0.027)	(0.028)	(0.031)
Constant	0.116**	0.051	0.161***	0.142**
	(0.049)	(0.047)	(0.051)	(0.055)
N	38,672	38,670	37,752	37,674

Notes. All models include additional student-level controls for prior achievement, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.

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

Table A2: Effects of Kickboard Treatment on Student Discipline

	All Infractions		Violent Infractions	
	Number of Suspensions	Days of Suspension	Number of Suspensions	Days of Suspension
Panel A: All Schools				
Kickboard Student (KBS)	0.173***	0.608***	0.004	0.061*
	(0.029)	(0.098)	(0.009)	(0.033)
1 Year Prior	-0.030	-0.121*	-0.060***	-0.189***
	(0.019)	(0.064)	(0.009)	(0.032)
1st Treatment Year	-0.084***	-0.271***	-0.052***	-0.170***
	(0.018)	(0.062)	(0.009)	(0.032)
2nd Treatment Year	-0.112***	-0.336***	-0.102***	-0.271***
	(0.019)	(0.063)	(0.010)	(0.034)
3rd Treatment Year	-0.112***	-0.346***	-0.100***	-0.249***
	(0.019)	(0.065)	(0.010)	(0.037)
KBS*1 Year Prior	-0.049	-0.173	0.006	0.006
	(0.037)	(0.117)	(0.012)	(0.044)
KBS*1st Treatment Year	-0.140***	-0.702***	0.003	0.009
	(0.033)	(0.111)	(0.012)	(0.044)
KBS*2nd Treatment Year	-0.312***	-0.937***	0.006	-0.029
	(0.033)	(0.113)	(0.013)	(0.046)
KBS*3rd Treatment Year	-0.206***	-0.778***	-0.033**	-0.130***
	(0.034)	(0.116)	(0.014)	(0.050)
Constant	0.183***	0.407***	0.056***	0.046
	(0.035)	(0.119)	(0.016)	(0.056)
N	101,868	101,868	101,868	101,868
Panel B: Schools with all Five Years of Data				
Kickboard Student (KBS)	0.141	0.939***	-0.079**	-0.128
	(0.119)	(0.351)	(0.039)	(0.133)
1 Year Prior	0.023	0.412	0.052	0.392***
	(0.092)	(0.272)	(0.044)	(0.150)
1st Treatment Year	-0.046	0.058	0.119***	0.399***
	(0.086)	(0.254)	(0.042)	(0.143)
2nd Treatment Year	-0.015	0.221	-0.115***	-0.272*
	(0.081)	(0.241)	(0.041)	(0.140)
3rd Treatment Year	-0.085	0.040	-0.097**	-0.234
	(0.081)	(0.240)	(0.042)	(0.144)
KBS*1 Year Prior	0.069	-0.427	0.020	-0.121
	(0.141)	(0.418)	(0.049)	(0.165)
KBS*1st Treatment Year	-0.149	-0.961**	-0.013	-0.089
	(0.126)	(0.374)	(0.045)	(0.154)
KBS*2nd Treatment Year	-0.381***	-1.527***	0.105**	0.225
	(0.123)	(0.365)	(0.045)	(0.154)
KBS*3rd Treatment Year	-0.244**	-1.320***	0.070	0.156
	(0.123)	(0.365)	(0.047)	(0.160)
Constant	0.461***	0.572	0.116**	0.029
	(0.122)	(0.361)	(0.050)	(0.169)
N	28,150	28,150	28,150	28,150
Panel C: NOLA Schools				
Kickboard Student (KBS)	0.196***	0.635***	0.007	0.058
	(0.039)	(0.138)	(0.012)	(0.046)
1 Year Prior	-0.021	-0.081	0.024**	0.029
	(0.019)	(0.069)	(0.012)	(0.043)
1st Treatment Year	-0.096***	-0.361***	0.006	-0.024
	(0.019)	(0.068)	(0.012)	(0.043)
2nd Treatment Year	-0.107***	-0.293***	-0.009	-0.009
	(0.019)	(0.069)	(0.012)	(0.044)
3rd Treatment Year	-0.120***	-0.305***	-0.014	-0.020
	(0.020)	(0.072)	(0.012)	(0.046)
KBS*1 Year Prior	-0.081	-0.377**	0.017	0.106*

	(0.050)	(0.176)	(0.016)	(0.060)
KBS*1st Treatment Year	-0.246***	-0.745***	0.019	0.112*
	(0.045)	(0.160)	(0.016)	(0.060)
KBS*2nd Treatment Year	-0.258***	-0.759***	0.013	0.003
	(0.048)	(0.171)	(0.016)	(0.060)
KBS*3rd Treatment Year	-0.202***	-0.642***	-0.010	-0.054
	(0.050)	(0.178)	(0.017)	(0.062)
Constant	-0.004	-0.029	-0.042	-0.275**
	(0.050)	(0.176)	(0.029)	(0.107)
N	52,128	52,128	52,128	52,128

Notes. All models include additional student-level controls for prior discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A3: Effects of Kickboard Treatment on Student Achievement by High and Low Implementation

	Math (High)	Math (Low)	ELA (High)	ELA (Low)	Science (High)	Science (Low)	Social Studies (High)	Social Studies (Low)
Kickboard Student (KBS)	0.020	-0.045	-0.007	-0.012	-0.054	-0.021	-0.042	-0.015
	(0.029)	(0.029)	(0.028)	(0.029)	(0.033)	(0.034)	(0.033)	(0.034)
1 Year Prior	0.059**	-0.134***	0.014	-0.168***	-0.028	-0.135***	0.013	-0.025
	(0.026)	(0.028)	(0.025)	(0.028)	(0.027)	(0.030)	(0.030)	(0.032)
1st Treatment Year	-0.051*	-0.024	-0.036	-0.018	-0.023	0.002	0.021	0.078**
	(0.027)	(0.028)	(0.026)	(0.028)	(0.028)	(0.030)	(0.030)	(0.032)
2nd Treatment Year	0.053**	0.057**	0.020	0.091***	0.021	0.040	0.063**	0.107***
	(0.027)	(0.029)	(0.026)	(0.029)	(0.027)	(0.031)	(0.030)	(0.033)
3rd Treatment Year	0.089***	0.079**	0.077***	0.119***	0.070**	0.042	0.100***	0.068*
	(0.028)	(0.032)	(0.027)	(0.032)	(0.028)	(0.034)	(0.031)	(0.036)
KBS*1 Year Prior	0.040	0.040	0.030	0.023	0.047	0.013	0.011	-0.009
	(0.037)	(0.039)	(0.035)	(0.038)	(0.038)	(0.041)	(0.041)	(0.044)
KBS*1st Treatment Year	0.048	0.067*	0.050	-0.000	0.064*	0.024	0.011	0.010
	(0.037)	(0.038)	(0.036)	(0.037)	(0.038)	(0.040)	(0.042)	(0.042)
KBS*2nd Treatment Year	0.101***	0.094**	0.062*	0.056	0.064*	0.042	0.057	-0.009
	(0.037)	(0.038)	(0.035)	(0.037)	(0.038)	(0.040)	(0.041)	(0.043)
KBS*3rd Treatment Year	0.018	-0.008	0.042	0.009	0.077**	0.083*	0.040	0.061
	(0.037)	(0.044)	(0.036)	(0.043)	(0.038)	(0.046)	(0.042)	(0.049)
Constant	0.179***	0.007	0.137**	-0.136***	0.176***	0.174***	0.169**	0.066
	(0.058)	(0.044)	(0.056)	(0.043)	(0.060)	(0.046)	(0.066)	(0.049)
N	20,978	19,344	20,986	19,339	20,633	19,339	20,596	18,926

Notes. All models include additional student-level controls for prior achievement and discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.

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

Table A4: Effects of Kickboard Treatment on Student Discipline by High and Low Implementation

	Total Suspensions (High)	Total Suspensions (Low)	Days of Suspension (High)	Days of Suspension (Low)
Kickboard Student (KBS)	0.199**	0.190**	0.895**	0.736***
	(0.099)	(0.087)	(0.375)	(0.280)
1 Year Prior	0.014	-0.015	0.052	-0.059
	(0.030)	(0.037)	(0.113)	(0.119)
1st Treatment Year	-0.027	-0.096***	-0.112	-0.331***
	(0.030)	(0.035)	(0.113)	(0.112)
2nd Treatment Year	-0.093***	-0.082**	-0.268**	-0.197*
	(0.030)	(0.036)	(0.113)	(0.114)
3rd Treatment Year	-0.099***	-0.111***	-0.215*	-0.370***
	(0.031)	(0.036)	(0.117)	(0.117)
KBS*1 Year Prior	-0.077	-0.053	-0.102	-0.135
	(0.120)	(0.111)	(0.453)	(0.357)
KBS*1st Treatment Year	-0.263**	-0.198**	-0.957**	-0.800***
	(0.108)	(0.092)	(0.409)	(0.295)
KBS*2nd Treatment Year	-0.274**	-0.369***	-0.971**	-1.204***
	(0.109)	(0.091)	(0.412)	(0.293)
KBS*3rd Treatment Year	-0.098	-0.254***	-0.580	-0.993***
	(0.110)	(0.092)	(0.417)	(0.294)
Constant	-0.250***	0.178***	-0.773***	0.315
	(0.068)	(0.067)	(0.257)	(0.216)
N	36,987	37,723	36,987	37,723

Notes. All models include additional student-level controls for prior achievement and discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.

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

Table A3: Effects of Kickboard Treatment on Student Outcomes (IEP Students)

	Math	ELA	Science	Social Studies	Number of Suspensions	Days of Suspension
Kickboard Student (KBS)	0.003	-0.008	-0.010	-0.017	0.180***	0.646***
	(0.014)	(0.014)	(0.015)	(0.016)	(0.029)	(0.100)
IEP Student	-0.331***	-0.362***	-0.268***	-0.181**	-0.026	-0.401
	(0.072)	(0.070)	(0.074)	(0.080)	(0.085)	(0.289)
KBS*IEP	0.078	0.002	-0.007	-0.093	-0.077	-0.492
	(0.098)	(0.095)	(0.100)	(0.108)	(0.125)	(0.424)
1 Year Prior	-0.013	-0.036***	-0.023	-0.015	-0.038**	-0.142**
	(0.014)	(0.013)	(0.014)	(0.015)	(0.019)	(0.065)
1st Treatment Year	0.018	0.042***	0.005	0.052***	-0.089***	-0.287***
	(0.014)	(0.014)	(0.015)	(0.016)	(0.018)	(0.062)
2nd Treatment Year	0.100***	0.103***	0.047***	0.076***	-0.106***	-0.324***
	(0.015)	(0.014)	(0.015)	(0.016)	(0.019)	(0.065)
3rd Treatment Year	0.120***	0.145***	0.070***	0.062***	-0.116***	-0.363***
	(0.016)	(0.016)	(0.017)	(0.018)	(0.020)	(0.067)
KBS*1 Year Prior	0.022	0.015	0.015	0.027	-0.034	-0.184
	(0.019)	(0.018)	(0.020)	(0.021)	(0.038)	(0.130)
KBS*1st Treatment Year	0.042**	0.030	0.030	-0.024	-0.140***	-0.728***
	(0.019)	(0.018)	(0.020)	(0.021)	(0.033)	(0.114)
KBS*2nd Treatment Year	0.041**	0.026	0.012	-0.003	-0.333***	-1.004***
	(0.019)	(0.019)	(0.020)	(0.022)	(0.034)	(0.117)
KBS*3rd Treatment Year	-0.007	0.037*	0.041*	0.036	-0.200***	-0.788***
	(0.021)	(0.021)	(0.022)	(0.024)	(0.035)	(0.120)
IEP*1 Year Prior	0.109	0.141*	-0.042	-0.135	0.173*	0.492
	(0.083)	(0.081)	(0.085)	(0.092)	(0.101)	(0.342)
IEP*1st Treatment Year	0.081	-0.014	0.001	-0.123	0.159	0.528
	(0.089)	(0.086)	(0.091)	(0.099)	(0.097)	(0.330)
IEP*2nd Treatment Year	0.034	0.072	0.066	-0.126	-0.032	0.009
	(0.081)	(0.079)	(0.083)	(0.090)	(0.091)	(0.309)
IEP*3rd Treatment Year	0.105	0.046	0.018	-0.091	0.070	0.287
	(0.084)	(0.082)	(0.086)	(0.093)	(0.091)	(0.308)
KBS*IEP *1 Year Prior	-0.128	-0.079	-0.089	0.005	-0.291*	-0.784
	(0.115)	(0.112)	(0.117)	(0.127)	(0.161)	(0.546)
KBS*IEP *1st Treatment Year	-0.055	0.020	-0.091	0.099	-0.093	0.168
	(0.120)	(0.117)	(0.123)	(0.133)	(0.144)	(0.490)
KBS*IEP *2nd Treatment Year	0.006	0.040	-0.010	0.141	0.238*	0.808*
	(0.110)	(0.107)	(0.112)	(0.122)	(0.136)	(0.463)
KBS*IEP *3rd Treatment Year	-0.019	-0.009	-0.006	-0.022	-0.041	0.253
	(0.114)	(0.111)	(0.116)	(0.125)	(0.137)	(0.467)
Constant	-0.035	-0.136***	0.099***	0.097***	0.185***	0.417***
	(0.023)	(0.022)	(0.024)	(0.026)	(0.035)	(0.119)
N	70,218	70,226	69,146	69,014	101,868	101,868

Notes. All models include additional student-level controls for prior achievement and discipline, gender, race, IEP and gifted status, ELL status, and FRL status. Standard errors are clustered at the school level.

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