

University of Arkansas, Fayetteville

ScholarWorks@UARK

Education Reform Faculty and Graduate
Students Publications

Education Reform

4-29-2017

Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?

Kaitlin P. Anderson

Michigan State University, ande2018@msu.edu

Gary W. Ritter

University of Arkansas, Fayetteville

Gema Zamarro

University of Arkansas, Fayetteville

Follow this and additional works at: <https://scholarworks.uark.edu/edrepub>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Leadership Commons](#), and the [Other Educational Administration and Supervision Commons](#)

Citation

Anderson, K. P., Ritter, G. W., & Zamarro, G. (2017). Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?. *Education Reform Faculty and Graduate Students Publications*. Retrieved from <https://scholarworks.uark.edu/edrepub/11>

This Article is brought to you for free and open access by the Education Reform at ScholarWorks@UARK. It has been accepted for inclusion in Education Reform Faculty and Graduate Students Publications by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.



UNIVERSITY OF
ARKANSAS

College of Education & Health Professions
Education Reform

WORKING PAPER SERIES

Understanding a Vicious Cycle:

Do Out-of-School Suspensions Impact Student Test Scores?

Kaitlin P. Anderson

Gary W. Ritter, Ph.D.

Gema Zamarro, Ph.D.

Original Version: March 30, 2017

This Version: April 29, 2017

EDRE Working Paper 2017-09

The University of Arkansas, Department of Education Reform (EDRE) working paper series is intended to widely disseminate and make easily accessible the results of EDRE faculty and students' latest findings. The Working Papers in this series have not undergone peer review or been edited by the University of Arkansas. The working papers are widely available, to encourage discussion and input from the research community before publication in a formal, peer reviewed journal. Unless otherwise indicated, working papers can be cited without permission of the author so long as the source is clearly referred to as an EDRE working paper.

Abstract

A vast body of research has proven the correlation between exclusionary discipline (out-of-school suspensions and expulsions) and student outcomes such as lower test scores, dropout, grade retention, and involvement in the juvenile justice system, but there is no consensus on the causal impacts of exclusionary discipline. This study uses six years of de-identified demographic, achievement, and disciplinary data from all K-12 public schools in Arkansas to estimate the causal relationship. We conduct dynamic panel data models incorporating student fixed effects using Anderson-Hsiao (1981) estimation. We find, counter-intuitively, a null to positive impact of out-of-school suspensions on test scores. Therefore, while policymakers may have other reasons to limit exclusionary discipline, we should not expect academic gains to follow.

Keywords: school discipline, exclusionary discipline, academic impacts, causal inference

JEL Codes: I20, I24

Acknowledgements

We would like to thank Daniel Leeds and Matthew Steinberg for their helpful comments on earlier versions of this work. In addition, we would also like to thank participants at the Association for Public Policy and Management Fall 2016 Conference, the Society for Research on Educational Effectiveness Spring 2017 Conference, and the Association for Education Finance and Policy Spring 2017 Conference for their questions and comments.

I. Introduction

There is much discussion in the United States education community about high rates of exclusionary discipline such as suspensions and expulsions for students in elementary and secondary schools (Bowditch, 1993; Marchbanks, Blake, Smith, Seibert, & Carmichael, 2014, Rausch & Skiba, 2005; Skiba, Peterson, & Williams, 1997). Moreover, there is concern about substantial disparities in rates of suspension or expulsion between white students and students of color (Anderson & Ritter, 2015; Anderson & Ritter, 2016; Anyon et al., 2014; Losen, Hodson, Keith, Morrison, & Belway, 2015; Losen & Skiba, 2010; Sartain et al., 2015; Skiba et al., 2014; Skiba et al., 2011; Skiba, Michael, Nardo, & Peterson, 2002; Welch & Payne, 2010).

Exclusionary discipline is associated with several negative student outcomes including lower academic achievement (Arcia, 2006; Beck & Muschkin, 2012; Cobb-Clark, Kassenboehmer, Le, McVicar, & Zhang, 2015; Raffaele-Mendez, Knoff, & Ferrer, 2002; Raffaele-Mendez, 2003; Rausch & Skiba, 2005; Skiba & Rausch, 2004;), school drop-out and grade retention (American Academy of Pediatrics, 2013; American Psychological Association, 2008; Balfanz et al., 2014; Cobb-Clark et al., 2015; Ekstrom, Goertz, Pollack, & Rock, 1986; Fabelo et al., 2011; Gregory and Weinstein, 2008; Krezmien, Leone, & Achilles, 2006; Marchbanks et al., 2014; Raffaele-Mendez, 2003; Raffaele-Mendez and Sanders, 1981; Rodney, Crafter, Rodney, & Mupier, 1999; Stearns and Glennie, 2006; Wald and Kurlaendar, 2003), and involvement in the juvenile justice system (American Academy of Pediatrics, 2013; Balfanz et al., 2003; Fabelo et al., 2011; Nicholson-Crotty, Birchmeier, & Valentine, 2009). If these relationships are causal rather than simply correlational, the economic impact could be great. Marchbanks et al. (2014), for example, using data on three cohorts of Texas seventh grade students in 2000-01 to 2002-03, estimated that grade retentions associated with discipline cost

the state of Texas about \$76 million per year. Further, school suspension may predict higher rates of misbehavior, anti-social behavior, and subsequent suspensions (Balfanz et al., 2014; Costenbader & Markson, 1998; Hemphill, Toumbourou, Herrenkohl, McMorris, & Catalonao, 2006; Raffaele-Mendez, 2003; Tobin, Sugai, & Colvin, 1996).

Lower academic achievement could be a result of suspensions and other learning time lost (Davis & Jordan, 1994; Public Agenda, 2004; Scott & Barrett, 2004), which is consistent with findings that increased opportunity for learning is associated with high achievement and large achievement gains (Brophy, 1988; Brophy & Good, 1986; Carter, 1984; Cooley & Leinhardt, 1980; Fisher et al., 1981; Greenwood, Horton, & Utley, 2002; Hattie, 2002; Reynolds & Walberg, 1991; Stalling, Cory, Fairweather, & Needels, 1978; Wang et al., 1997). This argument is consistent with studies that find suspensions precede lower performance (Balfanz et al., 2014; Cobb-Clark et al., 2015, McIntosh et al., 2008; Rausch & Skiba, 2005).

For example, Balfanz et al. (2014) examined the connection between receiving an out-of-school suspension in ninth grade and later high school and post-secondary outcomes in Florida. In this descriptive work, even after controlling for demographics, attendance, and course performance, suspensions in ninth grade were associated with suspension in the future, later course failures, and chronic absenteeism. Suspensions may predict future suspensions if certain students are viewed by school employees as “frequent flyers” (Greene, 2008; Kennedy-Lewis, Murphy, & Grosland, 2014), “problem students” or “bad kids” (Collins, 2011; Pifer, 2000; Weismann, 2015), and this presumption of an inherent discipline issue harms interactions between students and teachers (Kennedy-Lewis et al., 2014).

However, misbehavior and suspensions do not always precede lower academic achievement. Several studies have found that low academic performance predicts a variety of

undesirable behaviors in the future (Arcia, 2006; Choi 2007; Miles & Stipek 2006; McIntosh et al., 2008). For example, Miles and Stipek (2006) find that poor literacy achievement in the first and third grades predicted relatively high aggressive behavior in the third and fifth grades. Choi (2007) found that grade point averages predicted delinquent offenses, substance abuse, gang initiation, and sexual activity across all racial groups. This could be due to decreased engagement or bond with the school (Hawkins, Smith, & Catalano, 2004). Further, Arcia (2006) matched a group of suspended students to similar non-suspended peers and found the suspended students had lower pre-suspension achievement (Arcia, 2006).

The literature described thus far indicates many potential relationships between exclusionary discipline and student academic outcomes, but there is ambiguity about the actual causal link, and in what direction this link may occur. Therefore, the ability to direct public policy based on these correlational studies is limited. However, despite the ambiguity surrounding the causal relationship, many school districts and states are moving away from exclusionary discipline towards less punitive consequences. As of May 2015, 22 states and the District of Columbia had revised laws to “require or encourage schools to: limit the use of exclusionary discipline practices; implement supportive (that is, non-punitive) discipline strategies; and provide support services such as counseling, dropout prevention, and guidance services for at-risk students” (Steinberg & Lacoë, 2016, p. 44). Further, as of the 2015–16 school year, 23 of the nation’s 100 largest school districts had changed policies to require non-punitive discipline strategies and/or limit suspension use (Steinberg & Lacoë, 2016).

The move away from exclusionary discipline appears to presume a causal effect of exclusionary discipline on these student outcomes, yet prior work is only correlational. Policymakers and school leaders would benefit from more rigorous, causal research on the effect

of exclusionary discipline on student outcomes in order to make government and school policies more effective. This is no easy task, however, because of the potential for reverse causality. That is, it is unclear whether disciplinary issues precede and “cause” poor student achievement, or the declining achievement of a struggling student and the associated disengagement from school leads to disciplinary problems. Another plausible chain of events is that a negative shock outside of the school setting causes simultaneous problems with both behavior and academic achievement at school. Thus, while the data appear to show that academic achievement and school discipline are related, sorting out the causality is a far more complicated task.

In this study, we attempt to estimate the impact of out-of-school suspension on future academic achievement.

The main research questions guiding this study are:

1. *What is the impact of out-of-school suspension on academic achievement in reading and math in the following year?*
2. *Do out-of-school suspensions affect academic achievement of certain subgroups differently (e.g. males and females, white and non-white students, free and reduced price lunch (FRL) eligible and non-eligible students, special and regular education students, lower and higher performing students, and students in elementary or higher grades)?*

These research questions are limited but also an important first step toward identifying a causal impact of out-of-school suspensions on student outcomes. Academic achievement, in terms of performance on tests, is only one outcome that school disciplinary policies might affect. Suspensions are also associated with increased risk of drop-out and reduced on-time graduation rates (American Academy of Pediatrics, 2013; American Psychological Association, 2008; Balfanz et al., 2014; Cobb-Clark et al., 2015; Ekstrom, Goertz, Pollack, & Rock, 1986; Fabelo et

al., 2011; Gregory and Weinstein, 2008; Krezmien, Leone, & Achilles, 2006; Marchbanks et al., 2014; Raffaele-Mendez, 2003; Raffaele-Mendez and Sanders, 1981; Rodney, Crafter, Rodney, & Mupier, 1999; Stearns and Glennie, 2006; Wald and Kurlaendar, 2003). Therefore, while this study will not examine all possible impacts of exclusionary discipline, it will provide evidence on at least two measures of academic achievement; math and reading test scores.

We focus on the academic impacts on suspended students, ignoring systemic or school-wide impacts, but academic impacts may not stop with the suspended students themselves. One study found that high levels of suspensions are also associated with lower achievement gains on non-suspended students (Perry & Morris, 2014). Others suggest that strict disciplinary policies could improve school achievement through the removal of disruptive students (Burke & Herbert, 1996; Kinsler, 2013). Nevertheless, these studies are also limited by the potential for reverse causality or confounding effects of factors that influences both school achievement and behavior.

Regardless of these limitations, this work is an important first step to move beyond correlational studies and estimate the causal impact of out-of-school suspensions (OSS) on student test scores. Next, we describe the data utilized for this study and the analytic sample.

II. Data and Sample

This study uses six years of de-identified demographic, achievement (test score), and disciplinary data from all K-12 schools in Arkansas provided by the Arkansas Department of Education (2008-09 through 2013-14). Demographic data include race, gender, grade, special education status, limited English proficiency-status, and FRL status.

Academic achievement data include standardized scores on state tests in reading and mathematics for six school years from 2008-09 to 2013-14. For the school years from 2008-09 to 2013-14, state tests in reading and math were administered as part of the Arkansas

Comprehensive Testing, Assessment, and Accountability Program (ACTAAP). The Arkansas Benchmark exams in English Language Arts (ELA) were administered in grades 3-8, and End of Course (EOC) examinations were administered in Algebra I, Geometry, Algebra II, and 11th Grade Literacy.¹ We standardized test scores within grade, year, test subject, and testing group (e.g. with accommodations or without) to account for differences in test administrations.

Discipline data include indicators for 19 infraction types and 13 consequences, the date of the infraction, and the length of the consequence. To simplify the analysis, we grouped similar infraction types, resulting in only 12 groups.² The three most common types of infractions, disorderly conduct (31.0%), insubordination (25.9%), and other non-specified infractions (21.7%),³ and insubordination (24.7%), represent almost 80% of all infractions during the study period. Furthermore, 13 consequence categories were collapsed into 7 (in school suspension (ISS), out of school suspension (OSS), expulsion, referral to an alternative learning environment (ALE), corporal punishment, no action, and other). The most common consequence types during the study period were in-school suspension (39.2%), out-of-school suspension (24.1%), other non-specified consequences (18.8%),⁴ and corporal punishment (16.5%). Expulsions (0.1%), referrals to Alternative Learning Environments (0.4%), and no action (0.9%) are very rare.

Disciplinary data were aggregated to the student-by-school-year unit level, so the indicators for both infractions and consequences indicate the number of times within a given school year the student was cited for some particular type of infraction and received some particular type of consequence. These disciplinary data are merged with the student level demographic and achievement data using unique student identifiers.

The analytic samples vary by type of analysis, but, in our preferred models, we excluded from our analytic sample students who were expelled or received a referral to an alternative

learning environment (ALE) during the study period. We estimate the impact on students who are suspended out-of-school, relative to receiving some other, non-exclusionary consequence. Therefore, excluding from the analysis any students who received an expulsion or ALE referral makes the reference category clearly less exclusionary than OSS. Further, we estimate the impact of OSS on a more typical (not extremely misbehaving) student. Excluding students who were expelled or referred to ALE for disciplinary purposes removed 4,353 to 8,940 observations from our samples, depending on the sample, representing only about 0.008% of observations.⁵

III. Analytic Methods

Our preferred models exploit the panel nature of our dataset, but we first estimate a pooled ordinary least squares (OLS) model as a correlational benchmark to compare with our preferred dynamic panel data estimates. In the OLS model, the standard errors are clustered at the student level (Angrist & Pischke, 2009; Huber, 1967; Rogers, 1993; White, 1980). The pooled OLS model suffers from endogeneity as the error terms ε_{it} are likely correlated to other explanatory variables in the model due to reverse causality or confounding factors.

Our pooled OLS model (benchmark) is:

$$y_{it} = \beta_1 OSSdays_{it-1} + \beta_2 infractioncount_{it-1} + \beta_3 gradelevel_{it} + \beta_4 schoolyear_t + \beta_5 y_{it-1} + \beta_6 X_{it} + d_{it} + \varepsilon_{it} \quad (1)$$

The variable of interest, $OSSdays_{t-1}$, is defined as the number of days of out-of-school suspension student i receives in year $t - 1$. Although this specification assumes that each day of OSS has the same effect on student outcomes, we relax this assumption as part of our robustness checks. We account for a student's behavioral history using a vector of counts for individual types of infractions a student committed in the previous year (e.g. alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying,

or other). Thus, the data allow us to control for the observed behavior that instigated the disciplinary consequence. We account for district time-invariant characteristics with district fixed effects, d_{it} , and include a vector of grade level indicators, $gradelevel_{it}$, and school year indicators, $schoolyear_{it}$, with 2008-09 school year as the reference category. The error term, ε_{it} , contains student and district time variant unobserved characteristics. This model also includes student characteristics, X_{it} , including gender, FRL status, special education status, limited English proficiency, and race/ethnicity indicators (White, Black, Hispanic, Asian, and other).

Despite the ability to control for measures of student behavior and background characteristics in this OLS specification, this model does not account for other unobservable student characteristics. Unobservable characteristics such as family backgrounds and communities may relate to both the students' risk of OSS and their academic outcomes, so omitted variable bias is problematic for a causal interpretation of the OLS results. Therefore, our preferred model uses student fixed effects and is estimated using the approach for dynamic panel data models introduced by Anderson-Hsiao (1981). By adopting a dynamic panel data approach, we are able to relax strict exogeneity assumptions on our explanatory variables.

The fixed effects model relaxes the strict exogeneity assumption, allowing a limited form of endogeneity through time-invariant student characteristics. A limitation of the fixed effects approach is that it requires adequate variation (between states of OSS) within individual students across time, and even if there is "enough" variation within students, biases may remain if endogeneity is driven by time-varying shocks with effects that persist over time. To the extent that there are time-varying factors that are related both to the likelihood of being suspended out-of-school, and to future academic outcomes, we may be concerned that ε_{it} remains correlated with our variable of interest, $OSSdays_{it-1}$, even after allowing for student fixed effects.

However, by including the lagged number of days in OSS, as opposed to the current number, we aim to limit this potential endogeneity. It would only be problematic if shocks that affect OSS are more permanent in nature and, as a result, have an effect over multiple years. Transitory shocks would not be problematic in our specification. In fact, the Anderson-Hsiao (1981) method we employ is valid only under the assumption that these time varying shocks are temporary and not correlated across time periods. Although we recognize this assumption may be strong, we believe that it is more plausible and less strict than that imposed by simple pooled OLS.

The fixed effects are identified only for students who switch states of OSS (i.e. they must have variation in the days of OSS they receive in different school years). Thus, the estimates of interest are not identified using students never exposed to OSS, although all students remain in the analysis and help gain precision in other estimates.⁶ We can consider this a *selected average treatment effect* (SATE) rather than an overall treatment effect (Angrist & Imbens, 1991). Therefore, the results do not necessarily generalize to any randomly selected student in the state of Arkansas. Although this might seem to be a limitation, the students who have at least one day in OSS during the time of our study might be the most relevant group from a policy point of view as these would be the types of students most affected by proposed policy changes.

Our proposed student fixed effects specification includes the same covariates as in our benchmark OLS model, but rather than including student demographic variables, we include student fixed effects. A basic student fixed effects model would be represented by the following:

$$y_{it} = \beta_0 + \beta_1 OSSdays_{it-1} + \beta_2 infractioncount_{it-1} + \beta_3 gradelevel_{it} + \beta_4 schoolyear_{it} + \beta_5 y_{it-1} + d_{it} + a_i + \varepsilon_{it} \quad (2)$$

We account for student individual time-invariant heterogeneity with a_i , which is allowed to be correlated to our other regressors through relaxing the assumption of strict exogeneity.

With the inclusion of a_i , we exclude the vector of student characteristics X_{it} in Equation (1).

Other variables in (2) are the same as those included in Equation (1).

To estimate this student fixed effects model, one could transform the model using first differencing as:

$$\begin{aligned}
y_{it} - y_{it-1} = & \beta_1(OSSdays_{it-1} - OSSdays_{it-2}) + \beta_2(infractioncount_{it-1} - \\
& infractioncount_{it-2}) + \beta_3(gradelevel_{it} - gradelevel_{it-1}) + \\
& \beta_4(schoolyear_{it} - schoolyear_{it-1}) + \beta_5(y_{it-1} - y_{it-2}) + \beta_6(d_{it} - d_{it-1}) + \varepsilon_{it} - \\
& \varepsilon_{it-1}
\end{aligned} \tag{3}$$

Equation (3) above makes it clear that $(y_{it-1} - y_{it-2})$ will be mechanically correlated to $\varepsilon_{it} - \varepsilon_{it-1}$, introducing bias (Nickell, 1981). Similarly, $(OSSdays_{it-1} - OSSdays_{it-2})$, and each of our first-differenced infraction count measures in $(infractioncount_{it-1} - infractioncount_{it-2})$ would be mechanically correlated to $\varepsilon_{it} - \varepsilon_{it-1}$, as we argued above that this variables are potentially contemporaneously endogenous. Fortunately, the bias induced through this endogeneity can be corrected by using prior lags of these variables as instruments for the first differences. We use two-stage least squares (2SLS) to estimate our impact of out of school suspensions (Anderson & Hsiao, 1981). Our 2SLS models are given by:

First Stage:

$$\begin{aligned}
\Delta OSSdays_{it-1} = & \hat{\pi}_0 + \hat{\pi}_1 OSSdays_{it-2} + \sum_{j=1}^k \hat{\pi}_{2j} infractioncount_{ijt-2} + \\
& \hat{\pi}_3 \Delta gradelevel_{it} + \hat{\pi}_4 \Delta schoolyear_{it} + \hat{\pi}_5 y_{it-2} + \hat{\pi}_6 \Delta d_{it} + \eta_{it}^{OSS}
\end{aligned} \tag{4}$$

$$\begin{aligned}
\Delta y_{it-1} = & \hat{\pi}_0 + \hat{\pi}_1 OSSdays_{it-2} + \sum_{j=1}^k \hat{\pi}_{2j} infractioncount_{ijt-2} + \hat{\pi}_3 \Delta gradelevel_{it} + \\
& \hat{\pi}_4 \Delta schoolyear_{it} + \hat{\pi}_5 y_{it-2} + \hat{\pi}_6 \Delta d_{it} + \eta_{it}^{test_scores}
\end{aligned} \tag{5}$$

And k=12 equations for each infraction type

$$\Delta \text{infractioncount}_{ijt-1} = \hat{\pi}_0 + \hat{\pi}_1 \text{OSSdays}_{it-2} + \sum_{j=1}^k \hat{\pi}_{2j} \text{infractioncount}_{ijt-2} + \hat{\pi}_3 \Delta \text{gradelevel}_{it} + \hat{\pi}_4 \Delta \text{schoolyear}_{it} + \hat{\pi}_5 y_{it-2} + \hat{\pi}_6 \Delta d_{it} + \eta_{it}^{\text{infrac}}$$

for each $j = 1, \dots, 12$ (6)

Second Stage:

$$\Delta y_{it} = \beta_0 + \beta_1 \Delta \widehat{\text{OSSdays}}_{it-1} + \beta_2 \Delta \widehat{\text{infractioncount}}_{it-1} + \beta_3 \Delta \text{gradelevel}_{it} + \beta_4 \Delta \text{schoolyear}_{it} + \beta_5 \Delta \widehat{y}_{it-1} + \beta_6 \Delta d_{it} + \Delta \varepsilon_{it}$$
(7)

A valid instrumental variable requires two key assumptions: relevance (the instrument is correlated enough with the endogenous variable) and independence (the instrument does not directly affect the outcome, Δy_{it} , and the instrument is uncorrelated to the error term, $\Delta \varepsilon_{it}$). The relevance assumption is tested by looking for a clear relationship in the first stage results. In our case, our instruments are relevant by design. The independence assumption is based on the assumption that time varying shocks affecting OSS, infractions, or test scores are only temporary and so, they are not correlated over time with time varying unobservables that determine future test scores. Although this could be a strong assumption, we believe it is still weaker and more reasonable than the assumptions accompanying pooled OLS or other descriptive methods.

We describe our analytic samples in Table 1. The analytic samples are quite similar, regardless of the method used, and the samples are quite reflective of the state population.

IV. Results

Mathematics Results

The math results are in Columns 1 and 2 of Table 2. Column 1 presents the descriptive pooled OLS analysis results, which do not account for time-invariant student unobserved heterogeneity and cannot be interpreted as causal. The results in Column 1 of Table 2 indicate a

statistically significant (at the 99% confidence level) 0.006 standard deviation decrease in math test scores associated with each day of OSS in the prior year. Compared to prior literature showing large correlations between OSS and student outcomes, this estimate is small and reflects the robust set of controls for student behavior and background characteristics in the analysis.

In Column 2 of Table 2, we present our preferred student fixed effects models, instrumenting for the endogenous variables. The results of the Anderson-Hsiao model in Column 2 indicate a slight positive impact of OSS days in the prior year on math test scores (0.004 s.d. per day of OSS), significant at the 99% confidence level. The results of this model imply that, when we are more able to control for the endogeneity of our variable of interest and identify an arguably more causal impact, the effect of an additional day in exclusionary discipline on math test scores, among those who experience OSS at least once, if anything, is a very small positive.

ELA Results

The results of the ELA analysis are shown in the final two columns of Table 2. Based on the pooled OLS model, each day of OSS in the prior year is associated with a -0.006 standard deviation decrease in ELA test score. This is similar to the math test score estimate in Column 1. The preferred model in Column 4 indicates a slight positive impact of prior year OSS days on test scores (about 0.01 s.d.) among those receiving OSS at least once.

Within-student variation over time

One diagnostic test of whether fixed effects are appropriate is to investigate the amount of variation in the counts of OSS days between and within students over time. Fixed effects use only within variation to identify the impact of OSS. Appendix Table A1 indicates that there is almost as much variation within students as across students, so a fixed effects allow us to control for unobserved heterogeneity, without sacrificing much in terms of identifying variation.

The student fixed effect approach bases its identification only on those students who changed levels of exposure to OSS over time. Given that 82% to 84% of students in our analytic samples never received any OSS, we test whether the estimated relationship between OSS days in the prior year and test scores is different for students whose exposure to OSS varied over time. We estimate the same OLS models as in Table 2 but drop students who never received OSS. The estimated relationships are similar (negative, but smaller in magnitude) to those seen in columns 1 and 3 of Table 2 (-0.004 s.d. in math and -0.003 s.d. in ELA). Given that these results are of the same sign as the results from the broader sample, we do not have strong evidence that the results are driven by estimating effects only for students with variation in exposure to OSS days.

Testing for nonlinearities in impact of OSS days

While we find slight positive impacts on math and ELA test scores, it could be that the impact of OSS is not linear. For example, longer suspensions may have a greater impact than one or two day suspensions, which may be similar to a typical illness-related absence. Further, in Arkansas, out-of-school suspensions longer than 10 days are considered expulsions (Arkansas Code § 6-18-507), and for special education students, if a student receives more than 10 days of OSS in a school year, it may be considered a change of placement, requiring additional notification and services for that child (Arkansas Department of Education, 2008).

To test for non-linear relationships between prior year OSS days and student test scores, we transform our continuous variables, $OSSdays_{it-1}$ and $OSSdays_{it-2}$ into sets of indicator variables for whether the student received (in either the prior year, or the second prior year) zero days, 1-2 days, 3-4 days, 5-6 days, 7-10 days, or 11 or more days of OSS. In each of these models, student-by-school year observations with zero cumulative days of OSS during the year are treated as the reference group. See Table 3 for the frequencies of each of these groups, as a

percent of all student-by-year observations, for each of our four samples. About 95% of student-by-school year observations in each sample had zero days of OSS in the prior year, with about 1.5 to 2% in each of the 1-2 days and 3-4 day categories, and under 1% in each of the remaining categories. Instances of eleven or more days of OSS in the prior year are particularly rare.

The overall math results using these new explanatory variables are in the left two columns of Table 4, and the ELA results are in the right two columns. The negative relationships in our descriptive, pooled OLS models are, as expected, consistent with the results in Table 2, but our focus is on the student fixed effects models. The sign of the coefficients for the math student fixed effects model (Column 2) are consistent with the results in Table 2 (which included only a linear count of OSS days). However, not all are statistically significant. We find that 1-2 or 3-4 days of OSS, relative to none, leads to increases of about 0.02 s.d. in math test scores, while some longer suspensions (5-6 days or 11 or more days) lead to larger (yet still small) increases. Given the relative infrequency of high numbers of OSS days in Table 3, the lack of significance on the 3-4 OSS days and 7-10 OSS days could be due to low statistical power. The ELA impacts in Column 4 are similar to the results in Table 2 (positive in magnitude), with all but the impact of 1-2 OSS days statistically significant at the 95% confidence level or higher.

Subgroup Effects

We also assess whether the impact is different for certain groups of students. We present separate results for FRL and non-FRL eligible, white and non-white, male and female, and special education and regular education students. We also present results for students whose first test score was above (or below) average for their grade and school year and for observations recorded in grades 2-5 and grades 6-10.

Table 5 presents the subgroup impacts on math test scores using our preferred dynamic panel data methods. Recalling that the overall impact on math was about 0.004 s.d., per OSS day in the prior year, we see similar impacts (0.004 to 0.006 s.d.) in Table 5 for FRL students, non-white students, male students, regular education students, below average students, and students in grades 6-10 with no impacts on the remaining groups. None of the analyses in Table 5 indicate statistically significant heterogeneous impacts in math. For example, even though there is a positive impact on non-white students, we cannot reject (at the 95% confidence level) the null hypothesis that the impact for white and non-white students is the same.

Subgroup effects in ELA are in Table 6. Compared to the overall ELA impact of about 0.01 s.d. per OSS day the prior year, we find similar impacts on certain subgroups. As with the math impacts, we see very small positive impacts on FRL students, non-white students, male students, regular education students, below average students, and students in grades 6-10, but we also see some positive ELA impacts on female students as well. In addition, there is evidence that the students who initially scored below or above average are impacted differently, although these effects could just be reversion to the mean, if some of these students simply have idiosyncratically low or idiosyncratically high scores the first time we observe them.

We conduct similar subgroup analyses using the buckets for 1-2 days, 3-4 days, 5-6 days, 7-10 days, and 11 or more days of OSS, focusing, again, on our preferred dynamic panel data method. As in Table 5, regular education students' math scores appear to be impacted positively by OSS. Further, students whom we first observe with test scores below average consistently appear to benefit, but this result could just be due to mean reversion after an idiosyncratically low first test score. Otherwise, subgroup effects do not indicate clear and consistent stories, other than that there is only one coefficient (out of 60) in Table 7 that is statistically significant and

negative. Therefore, it is just as likely that this single negative impact is a result of chance, and we conclude that there are generally no negative impacts of OSS on math test scores.

Table 8 shows the same subgroup analyses, but predicting ELA test scores. As in Table 6, there are generally positive or null impacts of OSS on ELA test scores, with more consistently positive impacts on non-white students, regular education students, and students who scored below average the first time we observe their ELA score. This last result could be a result of reversion to the mean. While we do see two negative and significant impacts of students who were scoring above average the first time we observe their ELA score, this could be reversion to the mean for students who scored idiosyncratically high in their first observed year.

Robustness Checks

The results presented so far exclude students expelled or referred to an ALE for disciplinary reasons during the study period. We conduct similar analyses adding these students back in, and estimating the effect of total days of exclusion (including ALE and expulsion). The results reiterate that there is not a negative impact of exclusionary discipline on math or ELA test scores. Compared to the overall results in columns 2 and 4 of Table 2, (0.0039 s.d. in math and 0.0095 in ELA), the overall results using all three forms of arguably exclusionary discipline has impacts on both types of test scores that are null but positive in magnitude (see Appendix C).

In addition, while the subgroup impacts of OSS days in Tables 5 and 6 were always null or positive, we find a negative impact of exclusionary discipline more generally (including ALE and expulsions along with OSS) on white students in math (-0.002 s.d.) and in ELA (-0.003 s.d.). Otherwise, the results for all other subgroups are still null or positive (regression output tables available from the authors by request).

Further, we analyze, by subgroup, the impacts of 1-2 days, 3-4 days, 5-6 days, 7-10 days,

and 11 or more days of exclusionary discipline (as in Tables 7 and 8 but with all three types of exclusionary discipline, rather than just OSS). Regression output tables are available from the authors by request. No estimated math test score impacts were negative. For the ELA subgroup impacts of 1-2 days, 3-4 days, 5-6 days, 7-10 days, and 11 or more days of exclusionary discipline (OSS, expulsion, and ALE), we estimate negative impacts in only two cases and only for students who scored above average the first time we observe them, which could indicate simply reversion to the mean. Although we found tiny negative impacts of prior year exclusionary discipline on white students' math and ELA test scores overall (using the linear specification), none of the impacts of 1-2 days, 3-4 days, 5-6 days, 7-10 days, and 11 or more days of exclusionary discipline are statistically significant. Our estimates of the subgroup impacts of prior year exclusionary discipline on test scores tend to be positive or null, so we have confidence that the null to positive impacts of OSS are not driven by our sample restrictions.

V. Discussion and Conclusions

Summary of Results

We embarked on this study with the objective of generating a better understanding of the impact of out-of-school suspensions on academic achievement, in light of the growing concern that exclusionary discipline harms the academic progress of students. Our prior assumption was that students would learn less when they are not in school. However, it is also possible that the kinds of students who receive OSS – perhaps disaffected or disengaged students – are exactly the students who would already experience academic declines because of that disengagement. In this situation fraught with endogeneity concerns, estimating causal relationships is quite challenging.

Using dynamic panel data methods, we aimed to identify the causal impact of OSS days on a student's academic achievement in the following year. The use of student fixed effects, with

instruments for endogenous variables, produces an estimate that is closer to a causal impact than most previous work on this topic. The use of student fixed effects controls for the time-invariant characteristics of students, and predicting test scores in a future year allows us to avoid the likely impact of contemporary shocks on OSS and test scores in the same year. A remaining concern for a causal interpretation of our results is that there may be time-varying shocks to students that affect discipline and student test scores over time. Still, with the data available, the student fixed effect model utilizing Anderson-Hsiao (1981) estimation is our preferred approach to estimate the causal impact of OSS on student test scores, and is an important contribution to the field.

In general, we find that OSS days have a slight positive impact on the following year's test scores in math (about 0.004 s.d. per day of OSS) and in ELA (about 0.010 s.d. per day of OSS). When we test for nonlinearities in the impacts, we find null to positive effects, with no evidence of negative impacts on test scores. When we analyze the effects of OSS across various different models, there are only three negative and statistically significant impacts, (out of 156 different coefficients reported), so this could just be due to mere chance. The most consistent results from the subgroup analyses are that regular education students' and non-white students' test scores consistently get a boost from OSS. However, these impacts are generally quite small (about 1 percent of a standard deviation *at most*), so we interpret the results less as an indication of positive impacts and more as a rejection of negative impacts of OSS on test scores.

Our primary estimates are derived from a sample that is very representative of the state as a whole, but that excludes the most extreme disciplinary offenders. In addition, the estimation of the impact of OSS on student test scores is only estimated using variation in exposure to OSS of students who had at least one day of OSS. While this is a limitation, the students with at least one day of OSS during the study period are arguably the most relevant group from a policy point of

view as these would be the students most affected by proposed policy changes. While this is important to ensure comparison of OSS with non-exclusionary consequences such as in-school suspension, corporal punishment, no action, or other, these primary estimates refer to the impacts of OSS on a more typical, perhaps less high-risk type of student. Even in our robustness checks, where we include in the sample the most highly disciplined students (those who were expelled or referred to an ALE for disciplinary reasons), the general finding is again of null to positive impacts, with any negative impacts likely just being a chance occurrence.

Overall, the results were surprising to us. While our prior assumption was that OSS most likely depresses the academic achievement of suspended students, we find, at least in this one state, no evidence that OSS negatively impacts student test scores. Why might this be? One possibility is that disciplinary consequences are doing what they are, at least in part, intended to do: encourage students to get back on track. It could be that students with many suspension days receive additional supports at home to reinvest them in their education, and that this translates into positive growth in the next year. It could also be that other school-based interventions follow suspensions and precede academic gains. For example, students in Arkansas who receive exclusionary discipline in eighth grade are more likely to be retained in ninth grade, compared to similar peers who received no exclusionary discipline (Swanson, Erickson, & Ritter, 2017). Perhaps students test scores improve if they repeat a grade and as a result receive an extra year of math or ELA instruction in a course in which they were previously struggling.

Implications

These results are important given the trend toward reining in the use of OSS in schools. According to Steinberg and Lacoë (2016), as of May 2015, 22 states and the District of Columbia had revised their laws in order to “require or encourage schools to: limit the use of

exclusionary discipline practices; implement supportive (that is, nonpunitive) discipline strategies that rely on behavioral interventions; and provide support services such as counseling, dropout prevention, and guidance services for at-risk students.” In addition, as of the 2015–16 school year, 23 of the nation’s 100 largest school districts changed policies to require non-punitive discipline strategies and/or limit suspension use (Steinberg & Lacoë, 2016). Based on our results, if policymakers continue to push for reductions in exclusionary discipline, they should do so for reasons other than concern about negative impacts on student test scores.

In fact, recent experiences in states and cities implementing suspension reduction policies indicate that these policies might not work as planned and unintended negative consequences could occur. For example, Loveless (2017) documents efforts in California to reduce out-of-school suspensions. The California reforms were of two types: 1) outlawing suspensions in third grade and below for willful defiance (a.k.a. insubordination) and 2) incorporating restorative justice. While the report argues that this push to reduce OSS use was largely out of concern about racial disparities, the reforms have reduced the rate of suspensions overall without actually closing the gap between OSS usage for different racial groups. Further, using California school-level data over three years (2013-15), Loveless (2017) finds that middle schools and schools serving high proportions of poor or black students tended to have elevated suspension rates for Black students. In addition, Loveless (2017) argues that some educators are concerned about declines in safety and learning because more trouble-makers remain in school.

Max Eden (2017) reports on changes in school climate in NYC, using student and teacher surveys conducted over a ten year period in which two sets of discipline policy reforms occurred: one during the Bloomberg mayoral administration, and one under Mayor Bill de Blasio. Eden argues that school climate (based on survey responses) stayed relatively constant during

Bloomberg's reforms. Bloomberg's reforms were two-fold: 1) prohibiting the use of suspensions for first-time, low-level offenses such as 'uncooperative/noncompliant' behaviors or 'disorderly behavior' and 2) setting the maximum number of days (5) for kindergarten through third grade suspensions for mid-level offenses such as "disruptive behavior," shoving, using racial slurs, or engaging in inappropriate physical contact. Then, Eden, argues, school climate deteriorated following de Blasio's reform requiring principals to obtain written permission from the Office of Safety and Youth Development (OSYD) to suspend a student for "uncooperative/ noncompliant" and "disorderly" behavior. According to Eden, following de Blasio's reforms, teachers reported less order and discipline, and students reported more violence, drug use, alcohol use, and gang activity, and lower mutual respect among their peers. Echoing Loveless' (2017) concerns about differential impacts on certain types of schools, Eden (2017) finds that schools with high concentrations of non-white students experienced the worst declines in climate.

While issues highlighted in Loveless (2017) and Eden (2017) focus on systemic effects on school, it is clear from our work that even expected impacts on the suspended students themselves may be minimal, at least in terms of student test scores. Therefore, as some have argued, the case against the use of suspensions is weaker than advocates have often led themselves to believe (Griffith, 2017). We should not necessarily expect better test scores or overall improvements in student outcomes simply from reductions in OSS. Indeed, some have argued that school districts are changing discipline policies too quickly (Mathews, 2017), referring to the changes as "sickening rides on the out-of-school-suspension roller coaster."

Still, there could be valid reasons (beyond improving test scores) for school leaders to use exclusionary discipline sparingly. Exclusionary discipline disproportionately affects students of color (Anderson & Ritter, 2015; Anderson & Ritter, 2016; Anyon et al., 2014; Losen, Hodson,

Keith, Morrison, & Belway, 2015; Losen & Skiba, 2010; Sartain et al., 2015; Skiba et al., 2014; Skiba et al., 2011; Skiba, Michael, Nardo, & Peterson, 2002; Welch & Payne, 2010). These documented disparities suggest that the use of exclusionary discipline should at least be evaluated and researched further. Perhaps, regardless of the positive or null impacts on student test scores, if perceived as overly harsh or unfair, the use of exclusionary discipline could still lead to negative school climate or distrust in a school community.

Where does this leave us? On its own, reductions in OSS may not be effective, but there are non-trivial reasons to reevaluate the use of OSS in schools. However, if states or school districts seek to reduce reliance on exclusionary discipline, they should do so while improving access to preventative and supportive systems at the same time. School-Wide Positive Behavioral Interventions and Supports (SWPBIS a.k.a PBIS), for example, is a framework that implements three tiers of supports, with the top tier focusing on intensive supports for at risk students. There is experimental evidence indicating that implementation of a PBIS framework has a variety of positive impacts such as decreases in office referrals (Flannery et al., 2014) and improvement in test scores and student perceptions of school safety (Horner et al., 2009).

In this study, we find a counter intuitive result that OSS does not harm student test scores, which suggests that educators and policy makers should think careful about policies designed solely with the goal of reducing suspensions. Ultimately, large-scale policy changes focused on reducing suspensions may achieve that short-sighted goal, but without the resources or capacity to actually change the underlying behaviors of students, there very well could be negative unintended impacts on the school as a whole. While there may be some promising alternatives to OSS, it is not clear what we should expect from reductions in OSS, particularly if high-level policy changes are not supported by capacity building at the local level.

REFERENCES

- American Academy of Pediatrics. (2013). Out-of-School Suspension and Expulsion. *Pediatrics*, *131*(3).
- American Psychological Association. (2008). Are Zero Tolerance Policies Effective in Schools? An Evidentiary Review and Recommendations. *American Psychologist*, *63*(9), 852-862.
- Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*, 227-97.
- Anderson, T. W., and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, *76*, 581-606.
- Anderson, K. P. & Ritter, G. W. (2015). *Discipline disproportionalities in schools: The relationship between student characteristics and school disciplinary outcomes*. EDRE Working Paper 2015-08. Retrieved from:
<http://www.uaedreform.org/downloads/2015/11/edre-working-paper-2015-08.pdf>
- Anderson, K. P. & Ritter, G. W. (2016). *Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. State*. EDRE Working Paper 2016-14. Retrieved from: <http://www.uaedreform.org/downloads/2016/09/disparate-use-of-exclusionary-discipline-evidence-on-inequities-in-school-discipline-from-a-u-s-state.pdf>
- Angrist, J. D. & Imbens, G. W. (1991). *Sources of identifying information in evaluation models*. NBER Working Paper 117. Retrieved from: <http://www.nber.org/papers/t0117.pdf>
- Angrist, J. D., and Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

- Anyon, Y., Jenson, J. M., Altschul, I., Farrar, J., McQueen, J., Greer, E., Downing, B., & Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review, 44*, 379-386.
- Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society, 38*(3), 359-369.
- Arkansas Act 1329. (2013). An Act to Evaluate the Impact of School Discipline on Student Achievement; And for Other Purposes. Retrieved from:
<ftp://www.arkleg.state.ar.us/acts/2013/Public/ACT1329.pdf>
- Arkansas Code § 6-18-507. Title 6 – Education, Subtitle 2 – Elementary and Secondary Education. Retrieved from: <http://law.justia.com/codes/arkansas/2010/title-6/subtitle-2/chapter-18/subchapter-5/6-18-507>
- Arkansas Department of Education (2008). *Special Education and Related Services: 11.0 Discipline Procedures*. Retrieved from:
https://arksped.k12.ar.us/rules_regs_08/1.%20SPED%20PROCEDURAL%20REQUIREMENTS%20AND%20PROGRAM%20STANDARDS/11.00%20DISCIPLINE%20PROCEDURES.pdf
- Balfanz, R., Byrnes, V., & Fox, J. (2014). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. *Journal of Applied Research on Children: Informing Policy for Children at Risk, 5*(2).
- Balfanz, R., Spiridakis, K., Neild, R. C., & Legters, N. (2003). High-poverty secondary schools and the juvenile justice system: How neither helps the other and how that could change. *New Directions for Youth Development, No. 99*.

- Beck, A. N. & Muschkin, C. G. (2012). The Enduring Impact of Race: Understanding the Disparities in Student Disciplinary Infractions and Achievement. *Sociological Perspectives, 55*(4), 637-662.
- Bowditch, C. (1993). Getting rid of troublemakers: High school disciplinary procedures and the production of dropouts. *Social Problems, 40*(4), 493-509.
- Brophy, J. (1988). Research linking teacher behavior to student achievement: Potential implications for instruction of Chapter 1 students. *Educational Psychologist, 23*(3), 235-386.
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd Edition) (pp. 328-375). New York, NY: Macmillan Publishers.
- Burke, E., & Herbert, D. (1996). Zero tolerance policy: Combating violence in schools. *National Association of Secondary School Principals. NASSP Bulletin, 80*(579), 49. Retrieved from <http://0-search.proquest.com.library.uark.edu/docview/216029446?accountid=8361>
- Carter, L. (1984). The sustaining effects study of compensatory education. *Educational Researcher, 13*, 4-13.
- Cho, H., Hallfors, D. D., & Sánchez, V. (2005). Evaluation of a high school peer group intervention for at-risk youth. *Journal of Abnormal Child Psychology, 33*(3), 363-374.
- Choi, Y. (2007). Academic achievement and problem behaviors among Asian Pacific Islander American adolescents. *Journal of Youth Adolescence, 36*(4), 403-415.
- Cobb-Clark, D. A., Kassenboehmer, S. C., Le, T., McVicar, D., & Zhang, R. (2015). Is there an educational penalty for being suspended from school? *Education Economics, 23*(4), 376-395.

- Collins, K. (2011). Discursive positioning in fifth-grade writing lesson: The Making of a “Bad, Bad Boy.” *Urban Education, 46*, 741-785.
- Collins-Ricketts, J. E. & Rambo, A. (2015). The PROMISE program case examples: From get tough to solution building. *International Journal of Solution-Focused Practices, 3*(2), 17-22.
- Cooley, W., & Leinhardt, G. (1980). The instructional dimensions study. *Educational Evaluation and Policy Analysis, 2*, 7-25.
- Cornwell, P. (September 23, 2015). Seattle School Board halts suspensions for elementary students. *The Seattle Times*. Retrieved from: <http://www.seattletimes.com/seattle-news/education/seattle-school-board-halts-suspensions-for-elementary-students/>
- Costenbader, V. & Markson, S. (1998). School Suspension: A Study with Secondary School Students. *Journal of School Psychology, 36*(1), 59-82.
- Davis, J. E., & Jordan, W. J. (1994). The effects of school context, structure, and experiences on African American males in middle and high schools. *Journal of Negro Education, 63*, 570-587.
- Eden, M. (2017). *School discipline reform and disorder: Evidence from New York City Public Schools, 2012-16*. Manhattan Institute.
- Ekstrom, R. B., Goertz, M. E., Pollack, J. M., & Rock, D. A. (1986). Who Drops Out of High School and Why? Findings from a National Study. *The Teachers College Record, 87*(3), 356-373.
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to*

Students' Success and Juvenile Justice Involvement. The Council of State Governments Justice Center & Public Policy Research Institute.

- Fairbanks, S., Sugai, G., Guardino, D. & Lathrop, M. (2007). Response to intervention: Examining classroom behavior support in second grade. *Exceptional Children*, 73(3), 288-310.
- Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1981). Teaching behaviors, academic learning time, and student achievement: An overview. *Journal of Classroom Interaction*, 17, 2-15.
- Flannery, K. B., Fenning, P., McGrath Kato, M., & McIntosh, K. (2014). Effects of School-Wide Positive Behavioral Interventions and Supports and fidelity of implementation on problem behavior in high schools. *American Psychological Association*, 29(2), 111-124.
- Fronius, T., Persson, H., Guckenberger, S., Hurley, N., & Petrosino, A. (2016). *Restorative justice in U.S. schools: A research review*. West Ed. Retrieved from: http://jprc.wested.org/wp-content/uploads/2016/02/RJ_Literature-Review_20160217.pdf
- Greene, R.W. (2008). *Lost at School*. New York: Scribner.
- Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*, 31, 328-349.
- Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). The promise of a teacher professional development program in reducing racial disparity in classroom exclusionary discipline. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 166–179). New York: Teachers College Press.
- Gregory, A. & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology*, 46, 455 – 475.

Griffith, D. (March 22, 2017). Suspensions' questionable effect on graduation rates [Blog post].

Retrieved from <https://edexcellence.net/articles/suspensions-questionable-effect-on-graduation-rates>

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37, 449-482.

Hawkins, J. D., Smith, B. H., & Catalano, R. F. (2004). Social development and social and emotional learning. In Zins, J.E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.), *Building Academic Success on Social and Emotional Learning: What Does the Research Say?* (135-150). New York: Teachers College, Columbia University.

Hemphill, S. A., Toumbourou, J. W., Herrenkohl, T. I., McMorris, B. J., & Catalano, R. F. (2006). The effect of school suspensions and arrests on subsequent adolescent antisocial behavior in Australia and the United States. *Journal of Adolescent Health*, 39, 736-744.

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.

Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 221–233. Berkeley: University of California Press.

Kennedy-Lewis, B. L., Murphy, A. S., & Grosland, T. J. (2014). Using narrative inquiry to understand persistently disciplined middle school students. *International Journal of Qualitative Studies in Education*. Retrieved from: <http://dx.doi.org/10.1080/09518398.2014.974718>

- Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, 54(1), 355-383.
- Krezmien, M. P., Leone, P. E., & Achilles, G. M. (2006). Suspension, race, and disability: Analysis of statewide practices and reporting. *Journal of Emotional & Behavioral Disorders*, 14, 217 – 226.
- Lacoe, J. R., & Steinberg, M. P. (2016). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage, school climate, and student achievement. Manuscript in preparation.
- Losen, D., Hodson, C., Keith, M., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* The Civil Rights Project, UCLA.
- Losen, D. & Skiba, R. (2010). *Suspended Education: Urban Middle Schools in Crisis*. Report: The Civil Rights Project at UCLA, The Equity Project at the Indiana University Center for Evaluation and Education Policy, and the Southern Poverty Law Center.
- Loveless, T. (2017). *The 2017 Brown Center Report on American Education. How well are American Students Learning?* Washington, DC: Brookings Institution.
- Mader, N., Sartain, L., & Steinberg, M. (Forthcoming). When suspensions are shorter: The effects on school climate and student learning. Working paper.
- Marchbanks, M. P., Blake, J. J., Smith, D., Seibert, A. L., & Carmichael, D. (2014). More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 5(2).
- Mathews, J. (March 26, 2017). Stopping sickening rides on the out-of-school-suspension roller coaster. *The Washington Post*. Retrieved from:

https://www.washingtonpost.com/local/education/stopping-sickening-rides-on-the-out-of-school-suspension-roller-coaster/2017/03/26/acf50424-0fd8-11e7-9b0d-d27c98455440_story.html?utm_term=.178b3f5c7047

McIntosh, K., Flannery, K. B., Sugai, G., Braun, D. H., & Cochrane, K. L. (2008). Relationships between academics and problem behavior in the transition from middle school to high school. *Journal of Positive Behavior Interventions*, *10*(4), 243-255.

Miles, S. B. & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, *77*(1), 103-117.

Nicholson-Crotty, S., Birchmeier, Z., & Valentine, D. (2009). Exploring the Impact of School Discipline on Racial Disproportion in the Juvenile Justice System. *Social Science Quarterly*, *90*(4), 1003-1018.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417-1426.

O'Connor, J. (July 29, 2015). *Miami-Dade schools eliminating out-of-school suspensions*. State Impact, NPR. Retrieved from: <https://stateimpact.npr.org/florida/2015/07/29/miami-dade-schools-eliminating-out-of-school-suspensions/>

Perry, B. L., & Morris, E. W. (2014). Suspending progress: collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, *79*(6), 1067-1087.

Pifer, D. A. (2000). Getting in trouble: The meaning of school for “problem” students. *The Qualitative Report*, *5*(1 & 2).

- Public Agenda. (2004). *Teaching interrupted: Do discipline policies in today's public schools foster the common good?* Retrieved from:
http://www.publicagenda.org/files/teaching_interrupted.pdf
- Public Counsel. (September 27, 2014). *California enacts first-in-the-nation law to eliminate student suspensions for minor misbehavior* [Press Release]. Retrieved from:
http://www.publiccounsel.org/press_releases?id=0088
- Raffaele-Mendez L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New Directions for Youth Development*, No. 99.
- Raffaele-Mendez, L. M., Knoff, H. M., & Ferron, J. M. (2002). School Demographic Variables and Out-of-School Suspension Rates: A Quantitative and Qualitative Analysis of a Large, Ethnically Diverse School District. *Psychology in the Schools*, 39(3), 259-277.
- Raffaele-Mendez, L. M., and Sanders, S. G. (1981). An Examination of In-School Suspension: Panacea or Pandora's Box? *NASSP Bulletin*, 65(441): 65–69.
- Rausch, M. K., & Skiba, R. J. (2005). The academic cost of discipline: The contribution of school discipline to achievement. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec.
- Reynolds, A. J., & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology*, 83, 97-107.
- Rodney, L. W., Crafter, B., Rodney, H. E., & Mupier, R. M. (1999). Variables contributing to grade retention among African American adolescent males. *The Journal of Educational Research*, 92(3), 185-190.
- Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin 13*: 19–23. *Stata Technical Bulletin Reprints, Vol. 3*, 88–94.

- Sartain, L., Allensworth, E. M., & Porter, S. *with* Levenstein, R., Johnson, D. W., Huynh, M. H., Anderson, E., Mader, N., & Steinberg, M. P. (2015). *Suspending Chicago's Students: Differences in Discipline Practices across Schools*. The University of Chicago Consortium on Chicago School Research.
- Scott, T. M., & Barrett, S. B. (2004). Using staff and student time engaged in disciplinary procedures to evaluate the impact of school-wide PBS. *Journal of Positive Behavior Interventions, 6*, 21-27.
- Skiba, R. J. (2000). Zero tolerance, zero evidence: An analysis of school disciplinary practice. (Bloomington, IN: Education Policy Center, Indiana University).
- Skiba, R. J., Chung, C., Trachok, M., Baker, T., Sheya, A., & Hughes, R. (2014). Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion. *American Educational Research Journal, 51(4)*, 640-670.
- Skiba, R. J., Horner, R., Chung, C. G., Rausch, M. K., May, S., & Tobin, T. (2011). Race is Not Neutral: A National Investigation of Black and Latino Disproportionality in School Discipline. *School Psychology Review, 40(1)*, 85-107.
- Skiba, R. J., Michael, R., Nardo, A., & Peterson, R. (2002). The Color of Discipline: Source of Racial and Gender Disproportionality in School Punishment. *The Urban Review, 34(4)*, 317-342.
- Skiba, R. J., Peterson, R. L., & Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. *Education and Treatment of Children, 20*, 295-315.

- Skiba, R. J. & Rausch, M. K. (2004). *The Relationship between Achievement, Discipline, and Race: An Analysis of Factors Predicting ISTEP Scores*. Center for Evaluation and Education Policy, Indiana University.
- Skiba, R., Ritter, S., Simmons, A., Peterson, R., & Miller, C. (2005). The Safe and Responsive Schools Project: A school reform model for implementing best practices in violence prevention. *Handbook of school violence and school safety: From research to practice*, 631-650.
- Stallings, J., Cory, R., Fairweather, J., & Needels, M. (1978). *A study of basic reading skills taught in secondary schools*. Menlo Park, CA: SRI International.
- Stearns, E. & Glennie, E. J. (2006). When and why dropouts leave high school. *Youth & Society*, 38, 29 – 57.
- Steinberg, M. P. & Lacoë, J. (2016). What do we know about school discipline reform? *Education Next*. Retrieved from: <http://educationnext.org/what-do-we-know-about-school-discipline-reform-suspensions-expulsions/>
- Swanson, E., Erickson, H. H., & Ritter, G. W. (2017). *Examining the impacts of middle school disciplinary policies on 9th grade retention*. EDRE Working Paper 2017-11. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2952972
- Tobin, T., Sugai, G., & Colvin, G. (1996). Patterns in Middle School Discipline Records. *Journal of Emotional and Behavioral Disorders*, 4(2), 82-94.
- U.S. Department of Education & U.S. Department of Justice. (2014). *U.S. Departments of Education and Justice Release School Discipline Guidance Package to Enhance School Climate and Improve School Discipline Policies/Practices*. Retrieved from:

<http://www.ed.gov/news/press-releases/us-departments-education-and-justice-release-school-discipline-guidance-package->

- Wald, J. & Kurlaendar, M. (2003). Connected in Seattle? An exploratory study of student perceptions of discipline and attachments to teachers. *New Directions for Youth Development, 99*, 35-54.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1997). Learning influences. In H.J. Walberg & G.D. Haertel (Eds.), *Psychology and educational practice* (pp. 199-211). Berkeley: McCutchan.
- Welch, K. & Payne, A. N. (2010). Racial threat and punitive school discipline. *Social Problems, 57(1)*, 25-48.
- Weismann, M. (2015). *Prelude to Prison*. Syracuse, NY: Syracuse University Press.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*: 48, 817–838.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data. Second Edition*. Cambridge, MA and London, UK: The MIT Press.

Endnotes:

¹ Most but not all students take the Algebra I exam in ninth grade, Geometry in tenth, and Algebra II in eleventh, but we standardize within grade, year, test subject, and testing group to account for differences in scores that may be a function of the age at which a test is taken.

² We grouped all infractions involving weapons (handguns, rifles, shotguns, clubs, knives, or explosives) into one category. We grouped staff assault and student assault into one category. We grouped alcohol and tobacco into one category.

³ “Other” non-specified infractions were coded as a particular type of infraction at the school level, but when combined and reported by the ADE, they are grouped into an “other” category.

⁴ “Other” non-specified consequences were coded as a particular type of consequence at the school level, but when combined and reported by the ADE, they are grouped into an “other” category. Conversations with the Arkansas Department of Education Assistant Commissioner for Research and Technology, Eric Saunders, indicates that the majority of these other consequences are detentions, bus suspensions, parent/guardian conferences, Saturday school, or warnings.

⁵ As a robustness check, we also added back in the students who were referred to an ALE during the study period and/or expelled during the study period, and estimate the impacts of (in total) all three different kinds of exclusionary discipline on student test scores. Details are presented in Results section IV.

⁶ About 82% of students in our math analytical sample and about 84% of students in our ELA analytical sample never were exposed to OSS, so only about 16 to 18 percent of students in the state contribute to estimation of the impact of OSS.

Table 1: Descriptive statistics for state and analytic samples

	Entire State	Math Student FE		ELA Student FE	
		Math POLS Sample	(Anderson-Hsiao) Sample	ELA POLS Sample	(Anderson-Hsiao) Sample
N Observations	N/A	1,033,936	660,826	839,542	512,684
N Students	470,362	367,759	275,810	324,033	235,917
Male	51.0%	51.0%	50.9%	50.7%	50.6%
FRL	60.0%	61.0%	60.7%	61.0%	60.6%
Special Education	11.0%	11.2%	11.0%	11.0%	10.6%
Limited English Proficient	7.0%	6.9%	6.8%	6.9%	6.7%
White	64.6%	65.2%	65.0%	65.5%	65.6%
Black	21.2%	21.1%	21.4%	20.8%	20.8%
Hispanic	10.1%	10.0%	10.0%	9.9%	9.9%
Other Race	4.0%	3.7%	3.6%	3.8%	3.7%
Lagged Math Z-Score	0.00	0.00	0.01	0.03	0.04
Lagged ELA Z-Score	0.00	-0.01	0.00	0.02	0.03

Table 2: Relationship between OSS days and student test scores

	Dependent Variable: Math Z-Score		Dependent Variable: ELA Z-Score	
	(1)	(2)	(3)	(4)
	Pooled OLS	Student FE with Anderson-Hsiao	Pooled OLS	Student FE with Anderson-Hsiao
Prior Year (PY) OSS Days	-0.0060 ** (0.0006)	0.0039 ** (0.0013)	-0.0056 ** (0.0008)	0.0095 ** (0.0019)
PY Infraction Counts By Category ¹	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y
Student Fixed Effects		Y		Y
Student Demographic Controls ²	Y		Y	
Lagged Z-Score	0.714 ** (0.0009)	0.208 ** (0.0042)	0.686 ** (0.0010)	0.261 ** (0.0045)
Constant		0.330 (13.50)		0.403 ** (0.0145)
Observations	1,033,936	660,826	839,542	512,684
Number of Students	367,759	275,810	324,033	235,917

Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level.

** p<0.01, * p<0.05

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

²Student demographic controls include gender, FRL-status, special education status, limited English proficiency, and a vector of race/ethnicity indicators (White, Black, Hispanic, Asian, and Other).

Table 3: Frequency of OSS Days in Prior Year, by Sample.

	Math POLS		Math Student FE		ELA POLS		ELA Student FE	
0 OSS Days in PY	981,242	94.9%	623,586	94.4%	803,270	95.7%	487,989	95.2%
1-2 OSS Days in PY	15,789	1.5%	10,780	1.6%	12,396	1.5%	8,149	1.6%
3-4 OSS Days in PY	17,666	1.7%	12,612	1.9%	11,946	1.4%	8,216	1.6%
5-6 OSS Days in PY	8,141	0.8%	5,811	0.9%	5,150	0.6%	3,539	0.7%
7-10 OSS Days in PY	6,639	0.6%	4,763	0.7%	3,969	0.5%	2,760	0.5%
11+ OSS Days in PY	4,459	0.4%	3,274	0.5%	2,811	0.3%	2,031	0.4%
Total Observations	1,033,936	100%	660,826	100%	839,542	100%	512,684	100%

Note: POLS = Pooled ordinary least squares regression. Student FE = Student fixed effects models using Anderson-Hsiao estimation.

Table 4: Relationship between OSS days and student test scores

	Dependent Variable: Math Z-Score		Dependent Variable: ELA Z-Score	
	(1)	(2)	(3)	(4)
	Pooled OLS	Student FE with Anderson-Hsiao	Pooled OLS	Student FE with Anderson-Hsiao
1-2 OSS Days in PY	-0.047 ** (0.0057)	0.019 * (0.0088)	-0.0505 ** (0.0071)	0.0215 (0.0111)
3-4 OSS Days in PY	-0.0668 ** (0.0058)	0.0126 (0.0092)	-0.0551 ** (0.0075)	0.038 ** (0.0127)
5-6 OSS Days in PY	-0.0521 ** (0.0084)	0.0319 * (0.0137)	-0.0552 ** (0.0113)	0.0461 * (0.0199)
7-10 OSS Days in PY	-0.0695 ** (0.0100)	0.0228 (0.0162)	-0.0775 ** (0.0137)	0.0828 ** (0.0242)
11+ OSS Days in PY	-0.0842 ** (0.0128)	0.0852 ** (0.0233)	-0.0696 ** (0.0169)	0.128 ** (0.0352)
PY Infraction Counts By Category ¹	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y
Student Fixed Effects		Y		Y
Student Demographic Controls ²	Y		Y	
Lagged Z-Score	0.714 ** (0.0009)	0.208 ** (0.0042)	0.685 ** (0.0010)	0.261 ** (0.0045)
Constant		0.393 (12.00)		0.406 ** (0.0144)
Observations	1,033,936	660,826	839,542	512,684
Number of Students	367,759	275,810	324,033	235,917

Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level.

** p<0.01, * p<0.05

Note: Reference group is 0 days of OSS in prior year.

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

²Student demographic controls include gender, FRL-status, special education status, limited English proficiency, and a vector of race/ethnicity indicators (White, Black, Hispanic, Asian, and Other).

Table 5: Subgroup impacts of OSS days on standardized math test scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
Prior Year (PY) OSS Days	0.0040 ** (0.0014)	0.0029 (0.0032)	0.0051 ** (0.0016)	-0.0006 (0.0023)	0.0044 ** (0.0016)	0.0027 (0.0022)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y	Y	Y
Lagged Math Z-Score	0.243 ** (0.0051)	0.152 ** (0.0065)	0.213 ** (0.0070)	0.216 ** (0.0050)	0.245 ** (0.0060)	0.165 ** (0.0059)
Constant	-0.666 (22.24)	0.468 (3.445)	-0.441 (10.50)	0.436 ** (0.0165)	0.390 (8.370)	0.378 ** (0.0202)
Observations	404,859	255,967	230,981	429,845	336,029	324,797
Number of Students	168,096	107,714	95,494	180,316	140,556	135,254
Panel B:	Special Education	Regular Education	Below Avg. Math Score	Above Avg. Math Score	Grades 2-5	Grades 6-10
Prior Year (PY) OSS Days	-0.0033 (0.0046)	0.0050 ** (0.0012)	0.0064 ** (0.0016)	-0.0009 (0.0023)	0.0052 (0.0048)	0.0039 ** (0.0013)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y	Y	Y
Lagged Math Z-Score	0.367 ** (0.0113)	-0.0259 ** (0.0044)	0.283 ** (0.0057)	0.00912 (0.0055)	0.181 ** (0.0085)	0.214 ** (0.0049)
Constant	0.563 ** (0.0470)	0.312 ** (0.0115)	0.394 ** (0.0151)	0.646 (12.09)	0.726 ** (0.0275)	22.1 ** (4.225)
Observations	72,338	588,488	333,836	326,990	129,908	530,743
Number of Students	33,897	247,024	136,869	138,941	128,565	239,462

Robust standard errors in parentheses

** p<0.01, * p<0.05

†Subgroup effects are statistically different

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

Note: FRL, non-FRL, White, non-White, and above or below average test scores are based on the first available observation for that student. Grade-level subgroups and special education or regular education subgroups are based on the grade level associated with each particular observation.

Table 6: Subgroup impacts of OSS days on standardized ELA test scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
Prior Year (PY) OSS Days	0.0093 ** (0.0021)	0.0069 (0.0049)	0.0099 ** (0.0024)	0.0021 (0.0033)	0.0083 ** (0.0024)	0.0145 ** (0.0037)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y	Y	Y
Lagged ELA Z-Score	0.251 ** (0.0057)	0.266 ** (0.0066)	0.212 ** (0.0078)	0.293 ** (0.0053)	0.258 ** (0.0062)	0.284 ** (0.0063)
Constant	0.520 (7.620)	0.399 ** (0.0296)	-4.022 (31.46)	0.700 (13.30)	0.370 (8.029)	0.415 ** (0.0239)
Observations	310,955	201,729	176,177	336,507	259,632	253,052
Number of Students	143,427	92,490	81,323	154,594	119,758	116,159
Panel B:	Special Education	Regular Education	Below Avg. ELA Score	Above Avg. ELA Score	Grades 2-5	Grades 6-10
Prior Year (PY) OSS Days	0.0101 (0.0067)	0.007 ** (0.0017)	0.0134 **† (0.0025)	-0.0055 † (0.0034)	0.0085 (0.0050)	0.0092 ** (0.0021)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y	Y	Y
Lagged ELA Z-Score	0.296 ** (0.0118)	0.0201 ** (0.0048)	0.277 ** (0.0067)	0.139 ** (0.0058)	0.225 ** (0.0082)	0.274 ** (0.0054)
Constant	22.86 (34.22)	0.308 ** (0.0128)	0.394 (17.28)	0.400 (3.343)	0.738 ** (0.0289)	0.312 ** (0.0174)
Observations	54,294	458,390	233,186	279,498	128,568	384,112
Number of Students	28,087	211,684	107,377	128,540	127,284	199,859

Robust standard errors in parentheses

** p<0.01, * p<0.05

†Subgroup effects are statistically different

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

Note: FRL, non-FRL, White, non-White, and above or below average test scores are based on the first available observation for that student. Grade-level subgroups and special education or regular education subgroups are based on the grade level associated with each particular observation.

Table 7: Subgroup impacts of OSS days on standardized Math test scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
1-2 OSS Days in PY	0.0197 (0.0102)	0.0149 (0.0184)	0.0327 ** (0.0118)	-0.0039 (0.0133)	0.0134 (0.0110)	0.0369 * (0.0152)
3-4 OSS Days in PY	0.0154 (0.0105)	-0.00290 (0.0202)	0.0209 (0.0120)	-0.0079 (0.0144)	-0.0081 † (0.0117)	0.0538 ** (0.0157)
5-6 OSS Days in PY	0.042 ** (0.0154)	-0.0242 (0.0319)	0.0527 ** (0.0171)	-0.0192 (0.0231)	0.0196 (0.0173)	0.0602 ** (0.0232)
7-10 OSS Days in PY	0.0318 (0.0180)	-0.0472 (0.0409)	0.0227 (0.0202)	0.0180 (0.0278)	0.0073 (0.0204)	0.0527 (0.0278)
11+ OSS Days in PY	0.0785 ** (0.0257)	0.153 * (0.0635)	0.107 ** (0.0278)	0.00894 (0.0451)	0.118 ** (0.0293)	0.0209 (0.0404)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged Math Z-score	0.242 ** (0.0051)	0.152 ** (0.0065)	0.213 ** (0.0070)	0.216 ** (0.0050)	0.245 ** (0.0060)	0.164 ** (0.0059)
Constant	-4.560 (10.11)	0.361 ** (0.0263)	1.187 (11.36)	0.438 ** (0.0165)	0.420 (4.554)	0.379 ** (0.0202)
Observations	404,859	255,967	230,981	429,845	336,029	324,797
Number of Students	168,096	107,714	95,494	180,316	140,556	135,254
Panel B:	Special Education	Regular Education	Below Avg. Math Score	Above Avg. Math Score	Grades 2-5	Grades 6-10
1-2 OSS Days in PY	0.0331 (0.0324)	0.0355 ** (0.0080)	0.0374 ** (0.0113)	-0.00441 (0.0142)	0.0358 (0.0253)	0.0149 (0.0094)
3-4 OSS Days in PY	-0.0817 * (0.0348)	0.0514 ** (0.0083)	0.0272 * (0.0117)	0.00293 (0.0151)	0.0178 (0.0317)	0.0142 (-0.0096)
5-6 OSS Days in PY	-0.00135 (0.0503)	0.0569 ** (0.0124)	0.0542 ** (0.0171)	-0.00004 (0.0236)	0.134 * (0.0537)	0.0249 (0.0141)
7-10 OSS Days in PY	-0.0179 (0.0579)	0.047 ** (0.0148)	0.0444 * (0.0199)	-0.0040 (0.0292)	-0.0366 (0.0604)	0.0267 (0.0168)
11+ OSS Days in PY	-0.0795 (0.0914)	0.0752 ** (0.0208)	0.124 ** (0.0282)	-0.00591 (0.0458)	0.0780 (0.0892)	0.0886 ** (0.0241)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged Math Z-score	0.367 ** (0.0113)	-0.026 ** (0.0044)	0.283 ** (0.0057)	0.0090 (0.0055)	0.181 ** (0.0085)	0.214 ** (0.0049)
Constant	0.564 ** (0.0470)	0.314 ** (0.0114)	0.421 (7.536)	2.781 (4.788)	0.727 ** (0.0275)	5.26 ** (1.724)
Observations	72,338	588,488	333,836	326,990	129,908	530,743
Number of Students	33,897	247,024	136,869	138,941	128,565	239,462

Robust standard errors in parentheses

** p<0.01, * p<0.05

†Subgroup effects are statistically

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

Note: FRL, non-FRL, White, non-White, and above or below average test scores are based on the first available observation for that student. Grade-level subgroups and special education or regular education subgroups are based on the grade level associated with each particular observation.

Table 8: Subgroup impacts of OSS days on standardized ELA test scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
1-2 OSS Days in PY	0.0154 (0.0129)	0.0585 * (0.0232)	0.0382 ** (0.0146)	-0.00677 (0.0171)	0.0233 (0.0137)	0.00976 (0.0203)
3-4 OSS Days in PY	0.0436 ** (0.0146)	0.0189 (0.0286)	0.0579 ** (0.0161)	-0.0168 (0.0209)	0.0265 (0.0159)	0.0637 ** (0.0231)
5-6 OSS Days in PY	0.0487 * (0.0226)	0.0233 (0.0476)	0.0666 ** (0.0241)	-0.0505 (0.0365)	0.044 (0.0246)	0.0490 (0.0369)
7-10 OSS Days in PY	0.0745 ** (0.0272)	0.146 * (0.0618)	0.079 ** (0.0298)	0.0575 (0.0425)	0.0891 ** (0.0295)	0.0733 (0.0468)
11+ OSS Days in PY	0.138 ** (0.0391)	-0.114 (0.0999)	0.143 ** (0.0416)	-0.0552 (0.0709)	0.104 * (0.0427)	0.231 ** (0.0696)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged ELA Z-score	0.252 ** (0.0057)	0.266 ** (0.0066)	0.212 ** (0.0078)	0.301 ** (0.0054)	0.258 ** (0.0062)	0.284 ** (0.0063)
Constant	0.554 (9.463)	0.427 ** (0.0303)	0.31 ** (0.0227)	0.629 ** (0.0210)	0.416 (8.069)	0.42 ** (0.0238)
Observations	310,955	201,729	176,177	336,507	259,632	253,052
Number of Students	143,427	92,490	81,323	154,594	119,758	116,159
Panel B:	Special Education	Regular Education	Below Avg. ELA Score	Above Avg. ELA Score	Grades 2-5	Grades 6-10
1-2 OSS Days in PY	0.0899 * (0.0400)	0.0267 ** (0.0098)	0.0451 ** (0.0152)	-0.0098 (0.0164)	0.0464 (0.0266)	0.0134 (0.0122)
3-4 OSS Days in PY	0.00240 (0.0467)	0.0564 ** (0.0113)	0.0721 ** (0.0173)	-0.0111 (0.0194)	0.0367 (0.0336)	0.0369 ** (0.0138)
5-6 OSS Days in PY	0.126 (0.0719)	0.0436 * (0.0177)	0.0925 ** (0.0261)	-0.0694 * (0.0340)	0.14 * (0.0567)	0.0300 (0.0213)
7-10 OSS Days in PY	0.136 (0.0841)	0.0653 ** (0.0216)	0.144 ** (0.0311)	-0.122 ** (0.0446)	-0.00692 (0.0639)	0.093 ** (0.0261)
11+ OSS Days in PY	-0.0950 (0.136)	0.0887 ** (0.0308)	0.17 ** (0.0454)	-0.0454 (0.0642)	0.125 (0.0953)	0.12 ** (0.0378)
PY Infraction Counts By Category ¹	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged ELA Z-score	0.297 ** (0.0118)	0.0203 ** (0.0048)	0.277 ** (0.0067)	0.134 ** (0.0057)	0.225 ** (0.0082)	0.274 ** (0.0054)
Constant	6.732 (35.90)	0.311 ** (0.0127)	0.470 (8.481)	0.206 ** (0.0228)	0.738 ** (0.0289)	0.316 ** (0.0173)
Observations	54,294	458,390	233,186	279,498	128,568	384,112
Number of Students	28,087	211,684	107,377	128,540	127,284	199,859

Robust standard errors in parentheses

** p<0.01, * p<0.05

‡Subgroup effects are statistically

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

Note: FRL, non-FRL, White, non-White, and above or below average test scores are based on the first available observation for that student. Grade-level subgroups and special education or regular education subgroups are based on the grade level associated with each particular observation.

APPENDIX

Appendix A – Within Student Variation

Appendix Table A1: Variability of number of OSS days (lagged) between and within students, models excluding students referred to ALE or expelled

	Math Anderson-Hsiao Sample	ELA Anderson-Hsiao Sample
Overall	1.644 s.d.	1.466 s.d.
Between	1.370 s.d.	1.265 s.d.
Within	1.071 s.d.	0.922 s.d.

Appendix B – Transitions into Specific Groups of OSS days in PY

	Number of OSS Days in <i>t-1</i>					
	0	1-2	3-4	5-6	7-10	11+
Fewer OSS Days in <i>t-2</i>	0.0%	77.2%	80.1%	83.6%	85.0%	70.3%
Same Category of OSS Days in <i>t-2</i>	97.0%	8.9%	9.0%	6.2%	7.4%	29.7%
More OSS Days in <i>t-2</i>	3.0%	14.0%	10.9%	10.2%	7.6%	0.0%

Appendix C – Robustness Check to Include Students Expelled and/or Referred to ALE

Appendix Table C1: Overall impacts (comparable to Table 2)

	Dependent Variable: Math Z-Score		Dependent Variable: ELA Z-Score	
	(1)	(2)	(3)	(4)
	Pooled OLS	Student FE with Anderson-Hsiao	Pooled OLS	Student FE with Anderson-Hsiao
Prior Year (PY) Exclusion Days	-0.0038 ** (0.0004)	0.0010 (0.0006)	-0.0031 ** (0.0005)	0.0012 (0.0008)
PY Infraction Counts By Category ¹	Y	Y	Y	Y
Grade Level Indicators	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y
District Fixed Effects	Y	Y	Y	Y
Student Fixed Effects		Y		Y
Student Demographic Controls ²	Y		Y	
Lagged Z-Score	0.715 ** (0.0009)	0.208 ** (0.0042)	0.687 ** (0.0010)	0.26 ** (0.0045)
Constant		0.375 ** (0.0124)		0.403 ** (0.0142)
Observations	1,042,876	666,665	846,583	517,037
Number of Students	370,744	278,171	326,672	237,906

Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level.

** p<0.01, * p<0.05

¹PY Infraction Counts By Category are a vector of count variables representing the number of infractions of each type (alcohol/tobacco, weapons, assault, vandalism, truancy, insubordination, gangs, fighting, drugs, disorderly conduct, bullying, or other) in the prior year.

²Student demographic controls include gender, FRL-status, special education status, limited English proficiency, and a vector of race/ethnicity indicators (White, Black, Hispanic, Asian, and Other).

Exclusion Includes Out of School Suspension, Expulsion, and referrals to ALE.