12-2017

Implementation Fidelity and Outcomes of School Discipline Policy: Three Studies from Arkansas

Kaitlin Anderson
University of Arkansas, Fayetteville

Implementation Fidelity and Outcomes of School Discipline Policy:
Three Studies from Arkansas

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Education Policy

by

Kaitlin Anderson
University of Virginia
Bachelor of Science in Commerce, and Economics, 2009

December 2017
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

____
Dr. Gema Zamarro
Dissertation Director

____
Dr. Gary Ritter
Committee Member

____
Dr. Patrick Wolf
Committee Member

Abstract

In the United States, exclusionary discipline, including out-of-school suspension (OSS) and expulsion, is disproportionately administered to students of color and special education students. Exclusionary discipline is associated with lower academic achievement and higher risk of dropout, grade retention, and involvement in the juvenile justice system, but there is little causal evidence on this topic.

This dissertation reports on three analyses on school discipline, using administrative data from Arkansas public schools. The first study estimates racial disproportionalities in the use of exclusionary discipline. Controlling for reported behavior and student characteristics, my co-author and I find that Black students are 2.4 times as likely as White students to receive exclusionary discipline. Within schools, this race-based gap is insignificant, suggesting the gap is driven by differences across schools.

In the second study, my co-authors and I use student fixed effects within dynamic panel data models to attempt to estimate a causal effect of exclusionary discipline on student test scores in the following year. Counterintuitively, we find almost zero evidence of negative effects, suggesting that reductions in OSS, without additional supports or interventions, will likely not improve student achievement.

The third study examines the implementation and outcomes of a statewide policy eliminating OSS as a consequence for truancy. That study tests which school-level factors predict policy compliance and whether there were any policy-related changes in test scores, attendance, chronic absenteeism, truancy rates, or other student disciplinary outcomes. I find that compliance was low in high-minority, high-discipline schools, and there was no policy-related change in school-level test scores, attendance, and chronic absenteeism. Reports of

truancy and the use of OSS for truancy declined following the new policy, but part of this result may be due to changes in how schools report discipline.

In summary, my research indicates that real disparities in exclusionary discipline exist, but they are primarily between schools, and the negative impacts of the exclusionary discipline on its own may be minimal. What is likely more important is focusing on prevention and building positive school climate, rather than setting high-level policies and hoping schools will rise to the challenge.

Acknowledgements

I would first like to thank my committee: Gema Zamarro, Gary Ritter, and Patrick Wolf. Gema, you have been an invaluable mentor, teacher, and friend. Thanks for showing me what is possible and for pushing me in my analytical methods and thinking. Gary, this dissertation would not be remotely similar to what it is today without you. Thank you for being a great collaborator and mentor. Pat, thank you for constantly supporting my development as a researcher, writer, and teacher, and for helping me learn how to work classic rock song titles in to my academic work. Even if we are all just bricks in the wall, we clearly still need education.

I would also like to acknowledge the rest of the EDRE faculty –Jay Greene, Bob Costrell, Bob Maranto, and Julie Trivitt. Thank you for always pushing us towards greatness.

To all the other EDRE students, staff, and alumni. I believe most of my growth has come from collaborating, debating, and joking around with all of you. Nothing says "fun" like nerdy econometrics jokes.

To my cohort family – Leesa Foreman, Danish Shakeel, and Yujie Sude. I probably should have gotten Dany to write a poem for me, but this will have to suffice: thanks for all the good times and the laughter. I am so excited to see where life takes each of you.

To my family and friends, thanks for your support throughout this process. Thanks in particular to my brother, Bryan, for always pushing me intellectually and academically, and to Hannah McCollester and Amber McCrady Toeller who convinced me to take the first step in the fight for educational equity.

There are also several people who have provided helpful feedback on various parts of this research – Alexandra Boyd, Daniel Leeds, Jonathan Mills, and Matthew Steinberg. To you, and the many others that I am almost certainly forgetting, thank you.

Dedication

This dissertation is dedicated to my former students at Fordyce High School. I think about you all every day.

Table of Contents

List of Tables

Chapter 2: Disparate Use of Exclusionary Discipline: Evidence on Inequities in School Discipline from a U.S. State

Chapter 3: Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?

List of Figures

List of Abbreviations

ACTAAP = Arkansas Comprehensive Testing, Assessment, and Accountability Program

ADE = Arkansas Department of Education

ALE = Alternative learning environment

CCSS = Common Core State Standards

CITS = Comparative interrupted time series

CRDC = Civil Rights Data Collection

DD = Difference-in-differences

ELA = English Language Arts

ELL = English language learner (used interchangeably with LEP)

EOC = End of course (examinations)

FRL = Free- and reduced-price lunch

IDEA = Individuals with Disabilities Education Act

IEP = Individualized Education Program

ISS = In-school suspension

LEP = Limited English proficiency (used interchangeably with ELL)

OCR = Office for Civil Rights

OSS = Out-of-school suspension

PARCC = Partnership for Assessment of Readiness for College and Careers

SBE = (Arkansas) State Board of Education

List of Original Articles

Chapter 2:

Anderson, K. P., & Ritter, G. W. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. state. *Education Policy Analysis Archives, 25*(49). http://dx.doi.org/10.14507/epaa.25.2787

Chapter 1:

Introduction

There is much concern in the United States education community about the disproportionate use and potentially negative outcomes of exclusionary discipline such as out-of-school suspensions (OSS) and expulsions, as well as zero-tolerance policies, which remove students from school for a variety of offenses including violent misbehavior and less serious offenses such as dress code violations and truancy (Losen & Skiba, 2010; Skiba, 2014; Skiba & Peterson, 1999). Beginning in the early 1990s, many U.S. schools adopted zero-tolerance or exclusionary discipline policies in response to concerns about violence in schools, but there is now concern that this movement went too far. The arguments against exclusionary discipline are driven by three main factors.

First, there is a lack of evidence that these policies are effective disciplinary consequences that prevent future misbehavior. State zero-tolerance laws are not associated with decreases in principal perceptions of problem behaviors (Curran, 2016), and school suspension has been found to predict higher rates of misbehavior and suspensions in the future (Costenbader & Markson, 1998; Raffaele-Mendez, 2003; Tobin, Sugai, & Colvin, 1996).

A second concern is that exclusionary discipline is associated with a variety of negative student outcomes including lower academic achievement (Beck & Muschkin, 2012; Raffaele-Mendez, 2003; Skiba & Rausch, 2004), higher risk of grade retention or drop-out (American Academy of Pediatrics, 2013; Marchbanks, Blake, Smith, Seibert, & Carmichael, 2014; Swanson, Erickson, & Ritter, 2017), and involvement in the juvenile justice system (American Academy of Pediatrics, 2013; Fabelo et al., 2011; Nicholson-Crotty, Birchmeier, & Valentine, 2009). While this body of evidence paints a dreary picture about the educational outcomes of highly disciplined students, it does not necessarily follow that eliminating or reducing suspensions without additional interventions or supports would improve student outcomes. In the third chapter of this dissertation, we aim to estimate a causal impact of exclusionary discipline on student test scores (Anderson, Ritter, & Zamarro, 2017) and find almost no evidence of a negative impact. Thus, we conclude that we should not expect improvements in student achievement to follow from a reduction in OSS, on its own.

A third major concern is the disparate exposure to exclusionary discipline for students of color and students with special education needs. Black students represent 15% of U.S. students, but 35% of students suspended once, 45% of students with multiple suspensions, and 36% of expelled students (U.S. Departments of Education and Justice, 2014). Similarly, students with disabilities represent only 12% of students in the country, but they make up 20% of students suspended out-of-school once, 25% of students suspended more than once, and 19% of students expelled (U.S. Departments of Education and Justice, 2014). These findings are neither surprising nor new: many researchers have documented stark differences in suspension and expulsion rates between White and non-White students in the U.S. (Losen & Gillespie, 2012; 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; Sullivan, Klingbeil, & Van Norman, 2013).

The best estimates of the racial disparities in the use of exclusionary discipline account for the type of misbehavior a student was cited for, their behavioral history, and other student characteristics (Skiba et al., 2014; Skiba et al., 2011; Skiba et al., 2002). In addition, some studies have incorporated school-level factors to understand whether these disparities are primarily driven by differences across schools (Sartain et al., 2015; Skiba et al., 2002). Still, even the most careful of these analyses are limited in size and scope. Thus, there is still more to be learned about the sources and drivers of these disparities. This topic is the focus of Chapter 2, in which Dr. Gary Ritter and I conduct the first large scale analysis of disparities in exclusionary discipline using detailed data allowing infractions to be connected to specific consequence types, as well as to the discipline infraction history and demographic characteristics of students. In particular, we estimate the conditional disparities in the use of exclusionary discipline across the entire state as well as the average within-school disparities.

Driven by these three concerns of a lack of the intended deterrent effect, correlations to other negative outcomes, and racial disproportionalities – but without causal evidence of the actual impact of exclusionary discipline on students – many states and school districts are moving away from exclusionary discipline. According to Steinberg and Lacoe (2017), as of May 2015, 22 states and the District of Columbia had revised laws to limit exclusionary discipline and implement more supportive, non-punitive strategies. As of the 2015-16 school year, 23 of the 100 largest school districts had implemented similar reforms (Steinberg & Lacoe, 2017). These policies include reducing the length of suspensions as in Chicago (Stevens et al., 2015) and New York City (Eden, 2017), limiting suspensions for certain, minor misbehaviors as in California (Loveless, 2017; Public Counsel, 2014) and New York City (Eden, 2017), eliminating suspensions for truancy as in Arkansas (Chapter 4 of this dissertation), reducing suspensions for young students as in Seattle (Cornwell, 2015), or completely eliminating the use of suspensions as in Miami (O’Connor, 2015). Other reforms require principals to obtain written permission to suspend students for certain behaviors (Eden, 2017).

Many of these reforms are quite recent, and given the variety of policy reforms and contexts, there is little systematic, empirical evidence on their effectiveness. Yet, some researchers have expressed concerns about the ineffectiveness and unintended consequences of these reforms. For example, while OSS was officially eliminated in Miami-Dade, some students are still reportedly sent home from school, and teachers have concerns that the reduction was attempted without sufficient staff buy-in and support (Gerety, 2016). California's efforts to reduce suspensions have apparently decreased the rate of suspensions overall without actually closing the racial gaps in OSS (Loveless, 2017). Educators in California and New York City have expressed concerns about declines in safety and learning because misbehaving students remain in school (Eden, 2017; Loveless, 2017). In addition, certain schools may be more burdened by these unintended consequences: in New York City, declines in school safety in the wake of disciplinary reforms were the highest in schools with high concentrations of non-White students (Eden, 2017).

Thus, the issue is complicated, and policies may not always lead to the intended outcomes. Therefore, in Chapter 4 of this dissertation, I study a policy that intended to eliminate the use of OSS as a consequence for truancy, and test whether this policy was associated with any changes in important school outcomes including test scores, attendance, chronic absenteeism, truancy, and other disciplinary outcomes. I find no evidence that the policy reform improved school-level outcomes related to attendance and test scores, and instead, there is the possibility that schools changed how they report truancy and discipline following the reform.

For this dissertation, I dig deeper into questions related to racial disparities in disciplinary consequences, the impacts of exclusionary discipline on academic performance, and recent attempts to reform school discipline policies. Specifically, I conduct three studies on the implementation and outcomes of school discipline policies in Arkansas. Next, I provide a summary of each of the three studies included in this dissertation.

In the first study (Anderson & Ritter, 2017), which was published in *Education Policy Analysis Archives*, Dr. Gary Ritter and I use seven years of infraction-level data from all Arkansas public schools to build on the body of evidence documenting racial disparities in exclusionary discipline. Our key contributions are the ability to connect individual infractions to their corresponding consequences, cover an entire state for a period of seven years, and utilize school fixed effects analysis to distinguish between within- and across- school disproportionalities. We find that Black students are more likely than White students to receive exclusionary discipline, even after controlling for the type and frequency of disciplinary referrals, but that most of these disparities occur across rather than within schools. Across the state, Black students are about 2.4 times as likely to receive exclusionary discipline conditional on reported infraction type and frequency as well as observable student characteristics. Within school, however, this same conditional disparity is generally not statistically significant. In other words, the average within-school disparity in the conditional likelihood of exclusionary discipline between Black and White students in our fullest model is not significant. Within schools, the disproportionalities in exclusionary discipline are driven primarily by non-race factors such as free- and reduced-price lunch (FRL) eligibility and special education status. We also find that schools serving an above average share of non-White students tend to give out longer punishments, regardless of whether the school was above or below the state average FRL rate. In other words, the minority share in the school is a statistically significant and large driver of consequence severity, while FRL share is not. In combination, the results from this first study indicate that racial disparities in exclusionary discipline do exist, but that this race gap is driven primarily by differences in disciplinary practices between high-minority schools and primarily White schools.

In the second paper (Anderson et al., 2017), my co-authors and I estimate the impact of days of OSS on student achievement as measured by test scores. We use six years of de-identified demographic, achievement, and disciplinary data from all K-12 public schools in Arkansas to estimate dynamic panel data models (Anderson & Hsiao, 1981) incorporating student fixed effects to control for time-invariant unobservable student characteristics. The goal of this approach is to better control for background characteristics of students to reduce the influence of reverse causality or other confounds in our estimates. We find, counterintuitively, almost no evidence of a negative impact of OSS on student test scores. Therefore, while policymakers may have other reasons to limit exclusionary discipline, we conclude that we should not expect academic gains to follow from a reduction in OSS alone without additional interventions or supports.

Finally, the third paper (Anderson, 2017) studies the implementation and outcomes of a state-level policy reform that prohibited OSS as a consequence for truancy. While the law instituting this policy was passed in 2013, there was a lack of full compliance, even after three years. I study, using eight years of student- and infraction-level data for K-12 public school students across the state of Arkansas, which school-level factors are associated with the use of OSS as a consequence for truancy prior to the policy and which are associated with policy compliance. Further, I utilize comparative interrupted time series analysis to estimate whether school-level outcomes such as math and English Language Arts test scores, attendance, chronic absenteeism, truancy, and other disciplinary outcomes changed more in policy-affected "treatment" schools than their comparison schools. I find that schools serving more minority students, schools with higher truancy rates, and schools with higher OSS rates were less likely to comply with the policy, all else equal. This suggests that the types of schools likely targeted by this policy are the same ones not fully complying. Combined with a lack of evidence that the policy improved student achievement, attendance, or chronic absenteeism, these findings suggest that the potential impact of policy changes may be limited if policy changes are not communicated well, if there is not accountability to ensure compliance, and if there is not school capacity to handle discipline effectively.

These three papers fill significant gaps in our knowledge about how discipline policies are implemented and what impact these policies may have on students. The first and second papers also seek to highlight the fact that gaps in the literature still do exist, despite what may be interpreted by some as a vast amount of evidence. For example, many in the education community appear convinced that exclusionary discipline harms students without citing causal evidence to make that claim. In these papers, I use careful descriptive analyses to address an old problem in a new way (Chapter 2), aim to estimate – for the first time - a causal impact of exclusionary discipline on student test scores (Chapter 3), and assess the implementation fidelity and outcomes from a recent state-level policy change (Chapter 4). These types of analyses are necessary to inform effective public policy in the best interest of students, by moving the conversation beyond ideological arguments or a reliance on limited, correlational work.

References

American Academy of Pediatrics. (2013). Out-of-School Suspension and Expulsion. *Pediatrics, 131*(3).

Anderson, K. P. (2017). Just the way you are: School-level outcomes of a state-level discipline policy amidst implementation failure. Chapter 4 in *Implementation and Outcomes of School Discipline Policy: Three Studies from Arkansas* (Unpublished Dissertation).

Anderson, T. W., & 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. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. state. *Education Policy Analysis Archives, 25*(49).

Anderson, K. P., Ritter, G. W., & Zamarro, G. (2017). *Understanding a vicious cycle: Do out-of-school suspensions impact student test scores?* EDRE Working Paper 2017-09. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944346

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.

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.

Curran, F. C. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. *Educational Evaluation and Policy Analysis, 38*(4).

Eden, M. (2017). *School discipline reform and disorder: Evidence from New York City Public Schools, 2012-16.* Manhattan Institute. Retrieved from https://www.manhattan-institute.org/html/school-discipline-reform-and-disorder-evidence-nyc-schools-10103.html

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. Retrieved from https://csgjusticecenter.org/wp-content/uploads/2012/08/Breaking_Schools_Rules_Report_Final.pdf

Gerety, R. M. (October 25, 2016). School ‘suspensions’ continue despite Miami-Dade’s no suspension policy. *Miami Herald.* Retrieved from http://www.miamiherald.com/news/local/education/article110343972.html

Losen, D., & Gillespie, J. (2012). *Opportunities suspended: The disparate impact of disciplinary exclusion from school.* Civil Rights Project, UCLA. Retrieved from http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/upcoming-ccrr-research/losen-gillespie-opportunity-suspended-2012.pdf

Losen, D., Hodson, C., Keith, M., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles, CA: The Civil Rights Project, University of California – Los Angeles. Retrieved from https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/are-we-closing-the-school-discipline-gap

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. Retrieved from http://www.saferfoundation.org/files/documents/Suspended_Education.pdf

Loveless, T. (2017). *The 2017 Brown Center Report on American Education. How well are American Students Learning?* Washington, DC: Brookings Institution. Retrieved from https://www.brookings.edu/multi-chapter-report/the-2017-brown-center-report-on-american-education/

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).

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.

O’Connor, J. (July 29, 2015). Miami-Dade Schools eliminating out-of-school suspensions. *State Impact, National Public Radio.* Retrieved from https://stateimpact.npr.org/florida/2015/07/29/miami-dade-schools-eliminating-out-of-school-suspensions/

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.

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. Retrieved from https://consortium.uchicago.edu/publications/suspending-chicagos-students-differences-discipline-practices-across-schools

Skiba, R. (2014). The Failure of Zero Tolerance. *Reclaiming Children and Youth, 22*(4), 27-33.

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., 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., 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., & Peterson, R. (January 1999). ZAP zero tolerance. *Phi Delta Kappan, 80,* 372-76, 381-82.

Skiba, R., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP Scores.* Bloomington, IN: Center for Evaluation and Education Policy, Indiana University.

Steinberg, M. P., & Lacoe, J. (2017). What do we know about school discipline reform? *Education Next, 17*(1), 44-52.

Stevens, W. D., Sartain, L., Allensworth, E. M., and Levenstein, R. *with* Guiltinan, S., Mader, N., Huynh, M. H., & Porter, S. (March 2015). *Discipline practices in Chicago schools: trends in the use of suspensions and arrests.* Chicago, IL: University of Chicago Consortium of Chicago School Research. Retrieved from https://consortium.uchicago.edu/sites/default/files/publications/Discipline%20Report.pdf

Sullivan, A. L., Klingbeil, D. A., & Van Norman, E. R. (2013). Beyond behavior: Multilevel analysis of the influence of sociodemographics and school characteristics on students' risk of suspension. *Children, Research, and Public Policy, 42*(1), 99-114.

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 and U.S. Department of Justice. (2014). Dear colleague letter on the nondiscriminatory administration of school discipline. Retrieved from: http://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html

Chapter 2:

Disparate Use of Exclusionary Discipline:

Evidence on Inequities in School Discipline from a U.S. State

Kaitlin P. Anderson
University of Arkansas

Gary W. Ritter, Ph.D.
University of Arkansas

Abstract

There is much discussion in the United States about exclusionary discipline (suspensions and expulsions) in schools. According to a 2014 report from the U.S. Department of Education's Office for Civil Rights, Black students represent 15% of students, but 44% of students who were suspended more than once and 36% of expelled students. This analysis uses seven years of individual infraction-level data from public schools in Arkansas. We find that Black students are more likely to receive exclusionary discipline, even after controlling for the nature and number of disciplinary referrals, but that most of the differences occur across rather than within schools. Across the state, Black students are about 2.4 times more likely to receive exclusionary discipline (conditional on reported infractions and other student characteristics) than White students whereas, within school, this same conditional disparity is not statistically different from zero. Within schools, the disproportionalities in exclusionary discipline are driven primarily by non-race factors such as free- and reduced-price lunch (FRL) eligibility and special education status. We find, not surprisingly, that schools with larger proportions of non-White students tend to give out longer punishments, regardless of school income levels, measured by FRL rates.

Background on Issues in School Discipline

Since the early 1990s, many schools across the United States have adopted zero-tolerance and other harsh disciplinary policies in response to fears of violence in schools. The zero-tolerance philosophy is an approach that removes students from school for a variety of violations, ranging from actual serious offenses like violent behavior to dress code violations or truancy (Losen & Skiba, 2010; Skiba, 2014; Skiba & Peterson, 1999). While it is necessary for school leaders to do what is reasonable to maintain a positive learning environment and ensure the safety of the school community, these so-called zero-tolerance policies have been opposed by a growing number of researchers and observers who fear that this movement has gone too far.

Opponents of harsh disciplinary practices have voiced numerous concerns. First, there is some evidence that these policies do not have the hoped-for deterrent effect. For example, Curran (2016) recently found that state zero-tolerance laws are not associated with decreases in problem behaviors as perceived by principals. There is evidence that school suspension predicts higher rates of misbehavior and suspensions in the future, rather than reducing misbehaviors (Costenbader & Markson, 1998; Raffaele-Mendez, 2003; Tobin, Sugai, & Colvin, 1996).

Moreover, critics fear that zero tolerance might have other unintended negative consequences (Skiba, 2014). Zero-tolerance policies and exclusionary discipline practices, such as expulsions and suspensions, have been associated with lower academic achievement (Beck & Muschkin, 2012; Raffaele-Mendez, 2003; Raffaele-Mendez, Knoff, & Ferron, 2002; Skiba & Rausch, 2004), school dropout (American Academy of Pediatrics, 2013; American Psychological Association, 2008; Ekstrom, Goertz, Pollack, & Rock, 1986), and involvement in the juvenile justice system (American Academy of Pediatrics, 2013; Balfanz, Spiridakis, Neild, & Legters, 2003; Fabelo et al., 2011; Nicholson-Crotty, Birchmeier, & Valentine, 2009).

This active opposition to exclusionary discipline has made an impact and influenced some high-profile changes in school disciplinary practices. Chicago public schools enacted a policy in 2012 to reduce the length of student suspensions, and researchers from the University of Chicago's Consortium on Chicago School Research have been analyzing the impacts (Sartain et al., 2015). In September 2014, California became the first state in the nation to enact limits of student suspension for minor misbehaviors (Public Counsel, 2014). One of the nation's largest school districts, Miami-Dade, also eliminated out-of-school suspensions (OSS) ahead of the 2015-16 school year (O'Connor, 2015). In Seattle, the School Board declared a one-year moratorium on suspensions for elementary students in September 2015 (Cornwell, 2015).

Perhaps a key reason that disciplinary policies have been revised is the concern that zero-tolerance policies and exclusionary practices have been applied disproportionately to minority students. A 2014 national report from the U.S. Department of Education's Office for Civil Rights focused on the racial disparity in exclusionary disciplinary policies. The authors reported that although Black students represent only 15% of students across the nation, 35% of students suspended once are Black, 44% of students suspended more than once are Black, and 36% of expelled students are Black (U.S. Department of Education and U.S. Department of Justice, 2014). Indeed, over the past decade (and beyond), numerous researchers have documented differences in suspension rates between White students and non-White students across the nation (Children's Defense Fund, 1975; Costenbader & Markson, 1998; Losen & Gillespie, 2012; Losen, Hodson, Keith, Morrison, & Belway, 2015; Losen & Skiba, 2010; Raffaele-Mendez, 2003; Skiba et al., 2014; Skiba et al., 2011; Skiba, Michael, Nardo, & Peterson, 2002; Sullivan, Klingbeil, & Van Norman, 2013). In addition, non-White students were more likely to receive suspensions for relatively subjective offenses such as disrespect; the result is that non-White students were disproportionately missing school time, often for non-violent or trivial reasons (Skiba, Michael, Nardo, & Peterson, 2002).

In reaction to these circumstances, there is a growing - but limited - research base examining the racial disparities in the incidence of exclusionary discipline in schools across the country. Some studies rely on aggregate school- or district-level data and do not connect the actual student infractions to the disciplinary consequences (Children's Defense Fund, 1975; Losen et al., 2015; Losen & Skiba, 2010); such studies are informative but do not shed light on whether students are being treated unfairly. Others have utilized student-level data, but focus on disproportionalities in outcomes, without connecting them to the type or severity of infraction reported (Raffaele-Mendez, 2003; Sullivan et al., 2013). Some more recent studies utilize student-level or infraction-level datasets to address a more important issue: whether particular groups of students are treated differently for committing the same type of infraction (Skiba et al., 2014; Skiba et al., 2011; Skiba et al., 2002). While these analyses advance our knowledge on this issue, these studies are hampered by a variety of issues such as limited samples of students - one study (Skiba et al., 2002) involved only middle schools in a single district. In addition, certain studies do not incorporate school-level information or school fixed effects to assess whether disparities exist within certain types of schools (Skiba et al., 2011; Skiba et al., 2002).

Thus, in this article, we examine all disciplinary infractions and the resulting consequences for all K-12 students in a single U.S. state over a seven-year time period. We are able to connect individual student characteristics to specific infractions and to the resulting consequences. Using this rich dataset, we can carefully examine disparities in disciplinary outcomes by race and other student characteristics, while controlling for the infraction committed and for the school attended. By identifying the extent to which students of different racial groups are punished more or less severely for the same offenses, even within the same schools, we hope to make a meaningful contribution to the growing evidence base on this important and timely issue. Specifically, we ask three key research questions:

1) *Across schools in Arkansas, what, if any, disproportionalities exist in the use of exclusionary discipline for non-White students, low-income students, special education students, or English language learners?*
2) *Within schools, what, if any, disproportionalities exist in the use of exclusionary discipline for non-White students, low-income students, special education students, or English language learners?*
3) *What school characteristics are associated with harsher (longer) disciplinary consequences?*

The rest of this article proceeds as follows. In the next section, we present the literature on the topic of disparities in school discipline. Then, we describe our data and sample and outline our analytic methods. Following that, we present the results, and finally, we conclude with some discussion of our results.

Evidence from the Literature

We describe the relevant research in two sections. First, we present the evidence on the racial disparities in student discipline on a national level. Studies addressing this broad question generally rely on school-level data and provide only an overview of the consequences levied on groups of students. While these analyses are important, they leave many questions unanswered because they do not examine the drivers of these differences. For example, if particular groups of students are punished more severely than others for serious but similar infractions, this is likely an indication of implicit or explicit bias in disciplinary practice at the school. Thus, the second set of studies we present are particularly informative as they investigate the student- and school-level characteristics that are associated with the racial disparities in discipline.

National Overviews of Disciplinary Disproportionalities

In 2015, Dan Losen and colleagues from the Civil Rights Project at UCLA published a comprehensive report asking “Are We Closing the School Discipline Gap?” The authors focused on out-of-school suspension rates in every school district in the nation through the 2011-12 school year. The data revealed an overall increase in suspensions over the past 40 years, as well as an increasing gap in the suspension rates for White students and students of color. In 1972-73, only 6% of Black students were suspended during the year, as compared to 3% of White students and 3% of Hispanic students. By 2011-12, 16% of Black students were suspended; this rate was more than twice as great as for Hispanic students (7%) and more than three times as great as for White students (5%). Moreover, the authors also examined rates within states and districts and found much variability, indicating that district and school policies could strongly influence exclusionary discipline outcomes (Losen et al., 2015). While the authors did not mention “family fragmentation” (Pearlstein, 2011) or the decline of the two-parent family structure as factors contributing to this rise in the disciplinary gap, some scholars have argued that the rise in divorce rates and illegitimacy rates (particularly among African-Americans) during this time likely contributed to gaps in educational opportunities and outcomes including student discipline (Pearlstein, 2011). This trend was noted relatively early on (Moynihan, 1965).

Several years earlier, Losen teamed with noted discipline researcher Russell Skiba on a national study of suspension rates in middle school, using an earlier (2006) version of the Civil Rights Data Collection (CRDC). In this study, the authors analyzed suspension rates for students in more than 9,200 middle schools across the nation, as well as a sub-sample from 18 large urban districts, from the years 2002 to 2006 (Losen & Skiba, 2010). This analysis also revealed stark racial gaps in suspensions; for example, while only 10% of White male students in middle school were suspended in 2006, 28% of Black male students were suspended in that same year. In the urban sub-sample analysis, the authors found many schools in which more than one out of every three students in a particular racial group had been suspended during the year.

Overall, these and other analyses confirm that there are indeed systemic racial disparities in out-of-school suspensions. But, what factors drive these disparities? And do these differences persist even after controlling for infractions and referrals? In the next section, we summarize the emerging research literature addressing these questions. While we have not conducted a full systematic review of the literature, we searched thoroughly for literature on racial disparities in school discipline, with a focus on the use of exclusionary discipline, and used a snowball search to identify additional studies to include. We do not include theoretical or philosophical arguments for or against exclusionary discipline, but rather focus on studies that quantitatively assess the number of infractions or incidences of disciplinary consequences and the demographic characteristics of the students receiving these consequences. In general, we focus on articles since the year 2000.

Studies Examining the Drivers of Racial Discipline Gaps

In Chicago, where there has been a great deal of focus on exclusionary discipline in recent years, researchers from the Consortium on Chicago School Research scanned discipline data from roughly 85,000 high school students in the district in 2013-14 (Sartain et al., 2015). Using descriptive analyses, the authors have shown that Black students were three times as likely as Hispanic students to be suspended, and four times as likely as White and Asian students. While there was some evidence of students of different racial backgrounds systematically receiving more suspensions within the same schools, the primary driver of the differences was the school. That is, Black students attended schools, on average, that reported larger numbers of suspensions. While this investigation did consider some factors that play a role in the disparities, the authors were unable to account for the infractions allegedly committed by the students. Moreover, suspensions were the only consequence analyzed here. Nevertheless, this study moved the field forward by putting forth the idea that differing school environments or practices may be one driver of the racial discipline gap.

Welch and Payne (2010) further examined what drives the discipline gap by considering the "racial threat hypothesis" from criminal justice research. The authors posited that school leaders in buildings serving more Black students would be more likely to use punitive discipline and less likely to use restorative approaches. Analyzing data from a 1998 nationally representative survey of students and school personnel in 294 public middle schools and high schools, the authors used multivariate regression techniques and found that principals in schools with higher proportions of Black students were more likely to report using punitive disciplinary styles. Next, the authors considered the influence of differential student behavior by controlling for student reports of delinquency and teacher reports of school safety, and their findings suggest that students in schools serving high concentrations of Black students are subject to stricter discipline measures despite similarly safe and orderly environments. The weakness here, of course, is that the study is based on self-reports of disciplinary strategies rather than on actual disciplinary outcomes. Moreover, the data are all school-level and do not indicate whether Black students themselves are punished more severely or more frequently.

The studies discussed up to this point do not provide much information related to the causes of the observed disproportionalities. The disproportionalities may be due to more frequent misbehavior by Black students or a greater willingness of school staff to refer these students to the office for subjective offenses. While many of the studies described in the previous section utilized student-level data, other researchers have advanced the field by using infraction- and referral-level data to further analyze the disciplinary outcomes for certain infraction types.

Russell Skiba and a variety of colleagues have published studies that assess the drivers of actual racial disparities in discipline. First of all, Skiba et al. (2002) used student-level data on more than 11,000 students from 19 middle schools in one of the largest U.S. school districts in 1994-95 to explore what factors drive discipline disproportionalities. While this analysis did not consider the variation in disciplinary strictness between schools, the authors did pay attention to infraction type and assessed whether differential bad behavior might play a role by analyzing the reasons for the disciplinary referrals. Specifically, the authors found that White students were more likely to be referred to the office for objective infractions such as smoking or vandalism while Black students were more likely to be referred for more subjective offenses such as disrespect and noise. Thus, the authors concluded that the observed disproportionalities are not due to more "serious" or "disruptive" behavior by Black students, but rather to the higher rate of referrals for these subjective infractions (Skiba, 2002, p. 335).

Skiba et al. (2011) investigated the issue more deeply using student-infraction-level data from 364 elementary and middle schools across the United States using School-wide Positive Behavior Supports in 2005-06. Using logistic regression and multinomial logistic regression, the authors found that (1) Black students were more likely than White students to be referred to the office for a large variety of disciplinary infractions, and that (2) for the same referred infractions, Black students in all grades were significantly more likely to be given out-of-school suspension or even expulsion. Thus, even after accounting for the reported infraction type, Black students were more likely to be given exclusionary discipline. The main limitation in this analysis is that there is no control for school fixed effects; so, we do not know if the disparate strictness is occurring within or between schools.

Next, Skiba et al. (2014) used Hierarchical Linear Modeling to predict punishment as a function of infraction type and incorporated a third level to the model by incorporating school characteristics. Using information from all students in the disciplinary database in a single Midwestern state in 2007-08, the authors found that the odds of being suspended or expelled were predictably influenced by the severity of the infraction. Importantly, even after controlling for the infraction, Black students remained more likely to be given out-of-school suspensions, but were no more likely to be expelled. This analysis extends beyond the prior work due to the inclusion of level three, in which school-level characteristics, such as student race and poverty and the principal's attitude toward discipline, are incorporated into the model. When these school level variables are included, the race of the individual student was no longer significant; school-level variables, including the concentration of Black students in the school, drove the severity of the punishments allocated. Thus, these results are consistent with the "racial threat hypothesis" in schools suggested by Welch and Payne (2010). One potential weakness of the Skiba et al. (2014) study is the setting and context – the data represent a single year in a single U.S. state that serves relatively few FRL-eligible students (fewer than 40%) and very few Black students (8%). Our current study expands on this work by incorporating seven years of student-level panel data in a state that contains a more diverse population (21% Black, 12% Hispanic, and 61% FRL as of 2014-15).

Overall, evidence indicates there are racial disparities with respect to exclusionary discipline outcomes. Indeed, the Office for Civil Rights has recently demonstrated nationwide racial disparities in rates of suspensions and expulsions, and moreover, a couple of recent studies have concluded that Black students have been given disproportionate consequences for the infraction committed. However, it is still not clear whether in most cases, this disparity is due to students being treated differently within the same school or to the fact that Black students attend systematically different schools where the disciplinary practices are abnormally strict. To date, the most thorough assessment of the extent to which non-White students are more severely punished for similar disciplinary referrals, considering also whether these disparities occur within certain types of schools, has been published by Skiba et al. (2014).

Thus, although questions surrounding the sources of disciplinary disparities are critically important, the best evidence to date comes from a single school year in a single state serving relatively few FRL-eligible and Black students. Therefore, we believe it is valuable to conduct such analyses in additional settings, ideally with greater levels of student diversity and a longer study period. Our current study expands on previous work by accounting for specific infraction information (type, frequency, etc.) and school-level fixed effects whenever possible, using multiple years of data within a single U.S. state serving a student population that is approximately 61% low-income (FRL eligible), 21% Black, and 12% Hispanic.

Data and Sample

Arkansas Student Sample

First, it is important to show whether the patterns in the Arkansas data utilized in this study mirror the OCR data mentioned previously. In Table 1, we calculate the percentage of students in various subgroups, the percentage of students who received OSS at least once who were in these subgroups, and the percentage of students who were expelled in these subgroups.[1] The odds of a student in a given subgroup being in a consequence category (e.g. expelled) is the percentage of expelled students in that group divided by the percentage of total students in that group. For example, White students represent 65% of students in the state of Arkansas, and 38% of students receiving OSS, so the odds ratio is equal to (0.38/0.65) or approximately 0.58. Odds of less than one indicate that a certain group is underrepresented in a certain category, relative to their prevalence in the state, and odds of greater than one indicate that a certain group is overrepresented in a certain category.

Then, we calculate disparities (relative odds) between groups, which can be compared across different subgroups. In terms of the disparities for Black students, relative to White students, the Arkansas disparities are larger than the nationwide disparities for OSS, but smaller for expulsions. We can also see that, overall, the Black-White disparities are much larger than any other disparities, including those for special education students relative to non-special education students. In both the OCR data (nationally) and the Arkansas data, the odds indicate that Hispanic students and English language learners are somewhat underrepresented in these types of exclusionary discipline practices. Comparing the relative odds of Hispanic students to White students, however, Hispanics students are still over-represented relative to White students, at least in the OCR data. Arkansas Hispanic students are actually under-represented relative to White students in terms of expulsions (disparity = 0.76).

The Arkansas and OCR disparities are compared by simply subtracting the disparities

[1] We report differences between Black and White students, Hispanic and White students, ELL and non-ELL students, special education and regular education students. The Office for Civil Rights does not report disciplinary rates for FRL and non-FRL students separately, and the Arkansas dataset we used did not include gender, so those differences are not reported here.

(like a simple difference-in-difference). These indicate that the ELL/non-ELL disparities are quite similar, and that the Hispanic-White disparities are smaller in Arkansas than the nation. The Black-White disparities in OSS are larger in Arkansas than the nation, and Black-White disparities in expulsion are smaller in Arkansas than the nation. While Arkansas is only one of 50 states in the United States, these data indicate that, with the exception of the Hispanic-White disparities in expulsion, the pattern of disparities in Arkansas are generally similar to those in the nation as a whole.

Data and Descriptive Statistics

The study uses seven years of de-identified demographic and disciplinary data from all K-12 schools in Arkansas provided by the Arkansas Department of Education (2008-09 through 2014-15). Student demographic data include race, grade, special education status, limited English proficiency (LEP) status, and free-and-reduced-lunch (FRL) eligibility. Discipline data include indicators for 19 infraction types and 13 consequence types, the date of the infraction, and the length of the consequence. To simplify the analysis, we collapse infractions involving handguns, rifles, and shotguns into a single category, resulting in 17 distinct categories. The 13 consequence categories are collapsed into 7 (in school suspension (ISS), OSS, expulsion, referral to an alternative learning environment (ALE), corporal punishment, no action, and "other").[2]

---

[2] Our measure of out-of-school suspension includes two separately reported OSS types: Out-of-School Suspension (when the incident did not result in physical injury) and Out-of-School Suspension (when the incident did result in physical injury). Our measure of expulsion includes five separately reported expulsion types: Expelled, Expelled for Weapons (as defined by Federal, State, and Student Discipline Policy), Expelled for Drugs (does not include alcohol or tobacco), Expelled for dangerousness (the incident did not result in physical injury), and Expelled for dangerousness (the incident resulted in physical injury). Our measure of ALE referrals includes two separately reported consequence types: Alternative Learning Environment (full year) and Alternative Learning Environment (less than one year).

Table 1

Comparison of Arkansas and National Office for Civil Rights (OCR) Data

		Non-ELL		ELL			Regular Ed.		Special Ed.[a]			Non-FRL		FRL		
		% of Group	Odds	% of Group	Odds	ELL - Non-ELL Disparity	% of Group	Odds	% of Group	Odds	SpEd - Non-Sped Disparity	% of Group	Odds	% of Group	Odds	FRL - Non-FRL Disparity
Arkansas (2008-09 to 2014-15)	% Enrollment	93%		7%			89%		11%			40%		60%		
	% Stud. Receiving OSS	96%	1.03	4%	0.57	0.55	81%	0.91	19%	1.69	1.85	21%	0.52	79%	1.32	2.52
	% Stud. Expelled	96%	1.03	4%	0.55	0.53	81%	0.91	19%	1.76	1.94	21%	0.53	79%	1.31	2.48
OCR (2011-12)[b]	% Enrollment	90%		10%			88%		12%			N/A		N/A		
	% Stud. Receiving OSS	94%	1.04	6%	0.60	0.57	78%	0.89	22%	1.83	2.07	N/A	N/A	N/A	N/A	N/A
	% Stud. Expelled	95%	1.06	5%	0.50	0.47	81%	0.92	19%	1.58	1.72	N/A	N/A	N/A	N/A	N/A
Arkansas-OCR Diff-in-Diff	% Stud. Receiving OSS					-0.03					-0.22					N/A
	% Stud. Expelled					0.05					0.22					N/A

		White		Black			White		Hispanic		
		% of Group	Odds	% of Group	Odds	Black-White Disparity	% of Group	Odds	% of Group	Odds	Hispanic-White Disparity
Arkansas (2008-09 to 2014-15)	% Enrollment	65%		21%			65%		10%		
	% Stud. Receiving OSS	38%	0.58	54%	2.53	4.32	38%	0.58	6%	0.61	1.05
	% Stud. Expelled	48%	0.75	44%	2.06	2.75	48%	0.75	6%	0.57	0.76
OCR (2011-12)	% Enrollment	52%		16%			52%		24%		
	% Stud. Receiving OSS	35%	0.67	38%	2.38	3.56	35%	0.67	22%	0.91	1.37
	% Stud. Expelled	36%	0.70	36%	2.25	3.20	36%	0.70	22%	0.90	1.28
Arkansas-OCR Diff-in-Diff	% Stud. Receiving OSS					0.76					-0.32
	% Stud. Expelled					-0.46					-0.53

Note: ELL = English language learner. FRL = Free-and reduced price lunch eligible. Percentages reflect the number of students receiving OSS or expulsion at least once in a school year that are within a certain subgroup. Odds are the percent of students in a subgroup suspended or expelled divided by the subgroup's proportion of enrollment. Odds indicate over- (> 1) or under-representation (< 1). Disparities (relative odds) are the odds for one group divided by the odds for another. These indicate over- (disparities > 1) or under-representation (disparities < 1) relative to another group. The Arkansas-OCR Diff-in-Diff is an AR disparity less the corresponding OCR disparity and represents whether the AR disparities are higher or lower than the OCR disparities.

[a] Special-education students include those with an IEP, under the IDEA. Does not include handicapped students under Section 504.

[b] OCR race data collected for students without disabilities and students served under IDEA (not for those served only under Sec. 504).

A couple of these consequence types are worth elaborating on. An alternative learning environment is an "an alternate class or program within a public school or school district that affords all students an environment that seeks to eliminate barriers to learning for any student whose academic and social progress is negatively affected by the student's personal characteristics or situation" (AR Code § 6-48-104).

While the use of corporal punishment in schools is on the decline in the United States, as of 2017, it was still legal in 19 U.S. states, primarily in the South, Southwest, and Midwest (Langille, 2017). In Arkansas, school districts may use corporal punishment as long as the discipline policy includes "provisions for administration of the punishment, including that it be administered only for cause, be reasonable, follow warnings that the misbehavior will not be tolerated, and be administered by a teacher or a school administrator and only in the presence of a school administrator or his or her designee, who shall be a teacher or an administrator employed by the school district" (AR Code § 6-18-503). As long as corporal punishment is administered in "substantial compliance" with the district's written policy, teachers and administrators who administer corporal punishment are immune from civil liability (AR Code § 6-17-112). Each school district is responsible for collecting documentation that the discipline policy was received by all parents and students (AR Code § 6-18-502), and many school districts allow parents to opt-out for their child. Many schools in the state no longer use corporal punishment, with small, rural school districts most likely to continue this practice. The commissioner of the Department of Education has referred to it as a "local control issue" (Caputo, 2017).

The unit of analysis is the student-infraction level, so students can and often do have multiple observations within the same year. After removing duplicate entries (same student, discipline date, infraction type, consequence type, etc.), 1,243,555 total observations remain over the seven-year period. These observations were recorded for 240,999 individual students, which would represent about 35% of the individual students expected to attend Arkansas schools during this time period (thus, the other 65% of students in the state's public schools received no disciplinary referrals or consequences during this time period.) The breakdown by infraction and consequence, by year, can be seen in Tables 2 and 3. The vast majority of infractions (79.4%) are relatively subjective consequences such as disorderly conduct (29.7%), "other" infractions not specified in these categories (24.9%), and insubordination (24.7%). While there is not a clear definition of these terms set by the state, disorderly conduct and insubordination are subjective, catch-all type categories. Disorderly conduct could be used for any kind of disruption such as acting out in class, excessive noise, or running through the hallways, and insubordination is typically thought of as refusal to follow directions from a teacher or administrator.

Table 3 shows the trends in the reported types of disciplinary consequences. We consider OSS, expulsion, and referrals to an Alternative Learning Environment as exclusionary, given that they remove a student from the traditional learning environment, and in the case of expulsions and ALE, for long periods of time. ISS is considered non-exclusionary as the student remains in the school building, continues to receive assignments from their regularly assigned teacher, and then returns to the same classroom after a relatively short period (generally one to two days). Over the past seven years, there has been a decrease in exclusionary discipline as a proportion of total infractions (about 25% in 2008-09 compared to only about 19% in 2014-15), but much of this drop is due to large increases in the use of "other" non-specified consequences. While we have concerns about the uncertainty within this "other" non-specified category, the vast majority of these "other" non-specified outcomes are non-exclusionary.[3] Expulsions and no actions are consistently rare, and ISS was the largest category in each year, until 2014-15, in which the "other" (non-specified) category was the most common. The number of incidences of the "other" (non-specified) consequence category grew by over 300% between 2008-09 and 2014-15.

To simplify interpretation of the infraction categories, we create categories based on the type and length of consequences typically received for each infraction type. Table 4 indicates the percentage of incidences of each infraction type that result in exclusionary discipline (expulsion, OSS, or referral to an ALE), as well as the number of days of exclusionary discipline that typically result (the mean number of days of exclusionary discipline administered for each infraction type, if a non-zero number of days was reported). To group these infractions, as in Table 4, we consider, simultaneously, the percentage of incidences of that infraction type that result in exclusionary discipline, as well as the typical number of days of exclusionary discipline that results. At the same time, we consider infractions that are similar in nature (for example, substances that are not only illegal to have at school, but also illegal for even 18 year olds, such as drugs and alcohol, are somewhat different than tobacco and grouped as such).[4] The distinction between "major" and "minor" non-violent offenses, for example, is primarily based on the likelihood of each offense resulting in exclusionary discipline; there is a break in the pattern where the three "major" non-violent offenses result in the student being excluded approximately 30% of the time while "minor" offenses lead to exclusionary discipline around 20% of the time.

[3] 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.
[4] It is also possible that the typical punishments for drugs and alcohol are more likely to be similar in a state like Arkansas, where about 33% of the population report being Baptist, relative to a U.S. average of about 15% (Pew Research Center, 2015). It is possible that in heavily Baptist areas, alcohol is viewed as similarly immoral as and akin to illegal drug use.

Table 2

Infraction Types, By Year from 2008-09 to 2014-15 (Arkansas)

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	Total	% of Total
Disorderly Conduct	54,641	51,027	48,765	51,539	42,575	57,750	63,533	369,830	29.7%
Other	31,871	28,639	26,481	31,858	35,024	60,600	95,733	310,206	24.9%
Insubordination	47,273	46,151	45,765	38,798	34,759	43,068	51,200	307,014	24.7%
Fighting	12,378	12,456	12,471	12,136	12,434	13,128	14,576	89,579	7.2%
Truancy	9,968	11,834	11,734	10,465	9,407	12,914	14,987	81,309	6.5%
Bullying	3,455	4,099	4,363	4,483	4,515	5,496	5,856	32,267	2.6%
Tobacco	2,218	2,253	1,973	1,920	1,977	2,482	2,837	15,660	1.3%
Student Assault	1,856	1,820	1,615	1,645	2,007	2,153	2,232	13,328	1.1%
Drugs	944	996	954	1,146	1,259	1,295	1,511	8,105	0.7%
Vandalism	962	833	909	689	736	1,084	1,087	6,300	0.5%
Knife	401	419	384	396	443	532	497	3,072	0.2%
Staff Assault	292	312	277	314	354	350	487	2,386	0.2%
Alcohol	294	299	325	289	309	353	416	2,285	0.2%
Gangs	361	339	177	107	131	103	113	1,331	0.1%
Explosives	49	57	60	50	42	53	40	351	0.0%
Club	21	21	49	45	42	53	57	288	0.0%
Guns	38	18	32	26	35	33	62	244	0.0%
Total	167,022	161,573	156,334	155,906	146,049	201,447	255,224	1,243,555	100.0%
% of Total	13.4%	13.0%	12.6%	12.5%	11.7%	16.2%	20.5%	100.0%	

Table 3

Consequence Types, By Year from 2008-09 to 2014-15 (Arkansas)

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	Total	%
Exlusionary Discipline									
Out-of-School Suspension	41,348	39,613	36,780	37,791	40,233	42,290	47,853	285,908	23.0%
ALE	918	794	621	253	317	586	538	4,027	0.3%
Expulsion	135	322	193	95	200	249	165	1,359	0.1%
Total Exclusionary	42,401	40,729	37,594	38,139	40,750	43,125	48,556	291,294	23.4%
% of Annual Total	25.4%	25.2%	24.0%	24.5%	27.9%	21.4%	19.0%	23.4%	23.4%
Non-Exclusionary Discipline									
In-School Suspension	63,018	64,760	60,052	62,532	63,019	74,169	92,084	479,634	38.6%
Other	23,120	23,858	27,600	26,482	21,850	62,972	92,865	278,747	22.4%
Corporal Punishment	36,484	30,732	29,311	27,760	19,142	19,746	19,571	182,746	14.7%
No Action	1,999	1,494	1,777	993	1,288	1,435	2,148	11,134	0.9%
Total Non-Exclusionary	124,621	120,844	118,740	117,767	105,299	158,322	206,668	952,261	76.6%
% of Annual Total	74.6%	74.8%	76.0%	75.5%	72.1%	78.6%	81.0%	76.6%	76.6%
Total	167,022	161,573	156,334	155,906	146,049	201,447	255,224	1,243,555	100.0%
% of Seven Year Total	13.4%	13.0%	12.6%	12.5%	11.7%	16.2%	20.5%	100.0%	

Table 4

Category Groups (Based on Percent Exclusionary and Typical Exclusion Days

	% Resulting in Exclusionary Discipline	Typical Number of Days of Exclusion
Guns	77.5	11.8
Drugs and Alcohol	87.8	8.8
Drugs	88.2	9.0
Alcohol	86.4	8.0
Major Violence/Weapons	75.1	5.2
Club	83.0	4.0
Knife	74.9	5.8
Staff Assault	74.4	4.7
Minor Violence/Weapons	59.3	3.6
Gangs	63.6	5.4
Fighting	60.8	3.5
Student Assault	49.2	3.9
Explosives	47.6	4.5
Major Non-Violent	30.3	3.1
Tobacco	35.4	3.3
Vandalism	32.1	4.1
Bullying	27.5	2.8
Minor Non-Violent	19.2	3.2
Disorderly Conduct	20.4	3.6
Insubordination	18.7	2.7
Other	18.2	3.2
Truancy	12.0	2.9

Note. The typical number of days is the mean number of days of exclusionary discipline administered for each infraction type, if a non-zero number of days was reported.

Interestingly, in Table 4, we see that exclusionary discipline is not even used in all gun infractions. While expulsion is allowed for any student who brings a firearm or other weapon to school, the superintendent also has discretion to modify this requirement on a case-by-case basis, which appears to be happening in over 20% of gun-related incidents (AR Code § 6-18-507, 2015). The ability for school district leaders to adjust consequences on a case-by-case basis is perhaps further evidence that there are opportunities for disproportionalities in discipline outcomes to occur, even for infraction types in which we expect near universal exclusion. Table 4 simply presents the typical punishment (the average number of days, if a non-zero number of days was reported), but within each category, there may be variation in the type and length of punishment. For example, the average "explosive" incident may be considered "minor" if these are primarily made up of infractions such as bringing fireworks to school, rather than actually attempting an attack with a more dangerous explosive such as a bomb.

The state only codes certain types of infractions and consequences, so some categories used at a local level are coded as "other" at the state level. As a result, a large number of cases can be coded as "other" in either the infraction committed, the consequence received, or both. In the next section, we describe the analytic methods we employ to analyze these data and examine any possible disparities in disciplinary practices.

Analytic Methods

In our straightforward descriptive analyses presented in the previous section, we have described how frequently students of various subgroups are cited for various types of infractions, as well as how frequently students in these subgroups receive various types of consequences. Next, we use logistic regression and aggregated residual techniques to address our three primary research questions.

Research Question 1: Across the state, what, if any, disproportionalities exist in the use of exclusionary discipline for non-White students, low-income students, special education students, or English language learners?

We begin by testing whether students of various subgroups are more or less likely to receive exclusionary discipline, controlling for the type of infraction committed. We first analyze these disparities at a state level. Any disparities we find at this level could be due to differences across districts or schools, within district, or within school. We utilize logistic regression to predict whether certain types of students are more likely to receive exclusionary discipline (expulsion, OSS, or referral to an ALE), rather than another consequence (ISS, corporal punishment, no action, or other). The unit of analysis is the individual infraction level (there may be multiple observations, per student, per year). Whether or not a student receives exclusionary discipline (E_i) for a particular disciplinary incident, i, is defined as:

$$E_i = \begin{cases} 1 \; if \; E_i^* > 0 \\ 0 \; if \; E_i^* \leq 0 \end{cases}$$

$$E_i^* = \beta_0 + \beta_1 V_i + \beta_2 InfCat_i + \beta_3 InfOrder_i + \beta_4 Grade_i + \beta_5 Year_i + \varepsilon_i$$

Where V_i is a vector of the student-level demographic indicators (some combination of race, FRL-eligibility, special education status, and LEP-status)[5] for the student associated with the incident, i, $InfCat_i$ is a vector of 7 infraction categories, grouped by severity as in Table 4, $InfOrder_i$ is a vector of indicators for whether the infraction was the first, second, third, etc., for that student that year (a total of 10 indicators for 1-9 and 10 or more), $Year_i$ is a vector of school-year indicators, and ε_i is the infraction-level idiosyncratic error (clustered at the student level).

In this first analysis, no school-level indicators or covariates are included, so it is considered a model of state-wide racial or other disparities in disciplinary outcomes, conditional on similar infraction types, infraction history, grade level, and in some cases, other student demographic characteristics.

[5] Unfortunately, we do not have data on family structure (single-parent, two-parent, or other). FRL-eligibility serves as a rough proxy for socio-economic status and family background.

Research Question 2: Within schools, what, if any, disproportionalities exist in the use of exclusionary discipline for non-White students, low-income students, special education students, or English language learners?

Next, we seek to understand the disparities within schools, rather than across schools. We utilize similar logistic regression as in Research Question 1, but with the addition of school fixed effects. This within-school analytic strategy is motivated Anderson and Ritter's (2017) work finding that most of the disparities in the length of punishments (e.g. number of days of suspensions) at the state level diminish when school fixed effects are included, indicating that most of the disparities are across schools rather than within schools. If, in the current study, the disparities diminish when school fixed effects are included in our models, this would indicate that a great deal of the variation exists between schools. Thus, we also ask question three which seeks to disentangle the particular school characteristics driving these differences.

Research Question 3: What school characteristics are associated with harsher (longer) disciplinary consequences?

To address whether certain types of schools are more likely to assign disproportionately long punishments for similar types of infractions, we use a two-stage residuals analysis approach.[6] In the first stage, we predict the number of days of exclusionary discipline as a function of information related to the reported infraction that could reasonably predict the type or length of consequence received, as well as the cumulative number of reported infractions

[6] Alternatively, we could have used hierarchical linear modelling (Raudenbush & Bryk, 2002), but this two-stage residuals approach was chosen for ease of interpretation. This analysis is similar in concept to the two-step aggregated residuals approach used to calculate teacher value-added (Ehlert, Koedel, Parsons, & Podgursky, 2014). It is intended to partial out the effect of factors that would justifiably predict the severity of a consequence, prior to estimating how school-level factors influence severity.

associated with that student during the same school year. In this first stage, we do not include any student demographic information other than grade level, which could justifiably be associated with the type or severity of consequence used. Our first stage model utilizes ordinary least squares regression, with heteroskedastic-robust standard errors clustered at the student level (Angrist & Pischke, 2009; Huber, 1967; Rogers, 1993; White, 1980). The first stage model is:

$$DaysPunished_i = \beta_0 + \beta_1 InfCat_i + \beta_2 InfOrder_i + \beta_3 Year_i + \beta_4 GradeLevel_i + \varepsilon_i$$

Where i indexes at the incident level, $DaysPunished_i$ is the total number of days of punishment, $InfCat_i$ is a vector of infraction categories, which can be defined two ways (using all 17 categories, or our 7 infraction types, grouped generally by severity), $InfOrder_i$ is a vector of indicators for whether the infraction was the first, second, third, etc., for that student that year (a total of 10 indicators for 1-9 and 10 or more), $Year_i$ is a vector of school-year indicators, $GradeLevel_i$ is a vector of grade-level indicators, and ε_i is the infraction-level idiosyncratic error (clustered at the student level). In our primary model, we focus on days of exclusionary discipline (expulsion, OSS, or referral to an ALE) associated with a given infraction, with all other consequence types coded as zero days[7].

These residuals generated by the OLS model are then averaged at a school-by-year level to produce a measure of whether a school, on average, meted out longer punishments (residuals greater than 0) or shorter punishments (residuals less than 0), relative to the state average, for a similar type of infraction and for a student in the same grade with a similar number of past disciplinary infractions. We refer to this average school-level residual as the School Severity Index (SSI). The school-by-year SSI values are estimated using a school-level random effects

[7] Days of exclusion were at most 365 days. The average expulsion was 18.4 days, the average OSS was 3.3 days, and the average ALE was 10.6 days.

model, which shrinks the estimates towards zero for schools with relatively few observations. Schools with positive SSI values tend to give out longer punishments, and schools with negative SSI values tend to give out shorter punishments, relative to the state average for similar infraction observations.

In the second stage, we predict the SSI as a function of school-level demographic characteristics to assess which school characteristics are associated with disciplinary practices:

$$SSI_{st} = \beta_0 + \beta_1 X_{st} + \beta_2 Year_t + \varepsilon_{st}$$

Where s indexes at the school level, and t represents years of our panel. X_{st} is a vector of school-by-year level characteristics (log of enrollment, an indicator for region, an indicator for open-enrollment charter schools, indicators for elementary, middle, high school, or other school grade-level types, and the percent of the student population that is FRL-eligible, of a certain race, receiving special education services, LEP, or gifted and talented), $Year_t$ is a vector of school-year indicators, which accounts for time trends across schools, and ε_{st} is the school-by-year idiosyncratic error.

Next, we present our findings, beginning with some brief descriptive statistics and ultimately walking through the results of each of three research questions.

Results

Initial descriptive analyses focused on the frequency of consequence and infraction types for different subgroups of students. In Figure 1, it is easy to see that non-White students are disproportionately receiving all types of consequences. On average, each year, there are 29.6 in-school suspensions for every 100 Black students, but only 9.9 in-school suspensions for every 100 White students. Each year, there are 24.6 out-of-school suspensions for every 100 Black students, but only 4.3 for every 100 White students. Thus, a ratio-based measure of the Black-

White disparity in ISS indicates that Black students are about three times as likely to receive OSS as White students (29.6 divided by 9.9). For other consequence types such as referrals to ALE, this ratio is about 9.5 times, or for OSS, 5.7 times.

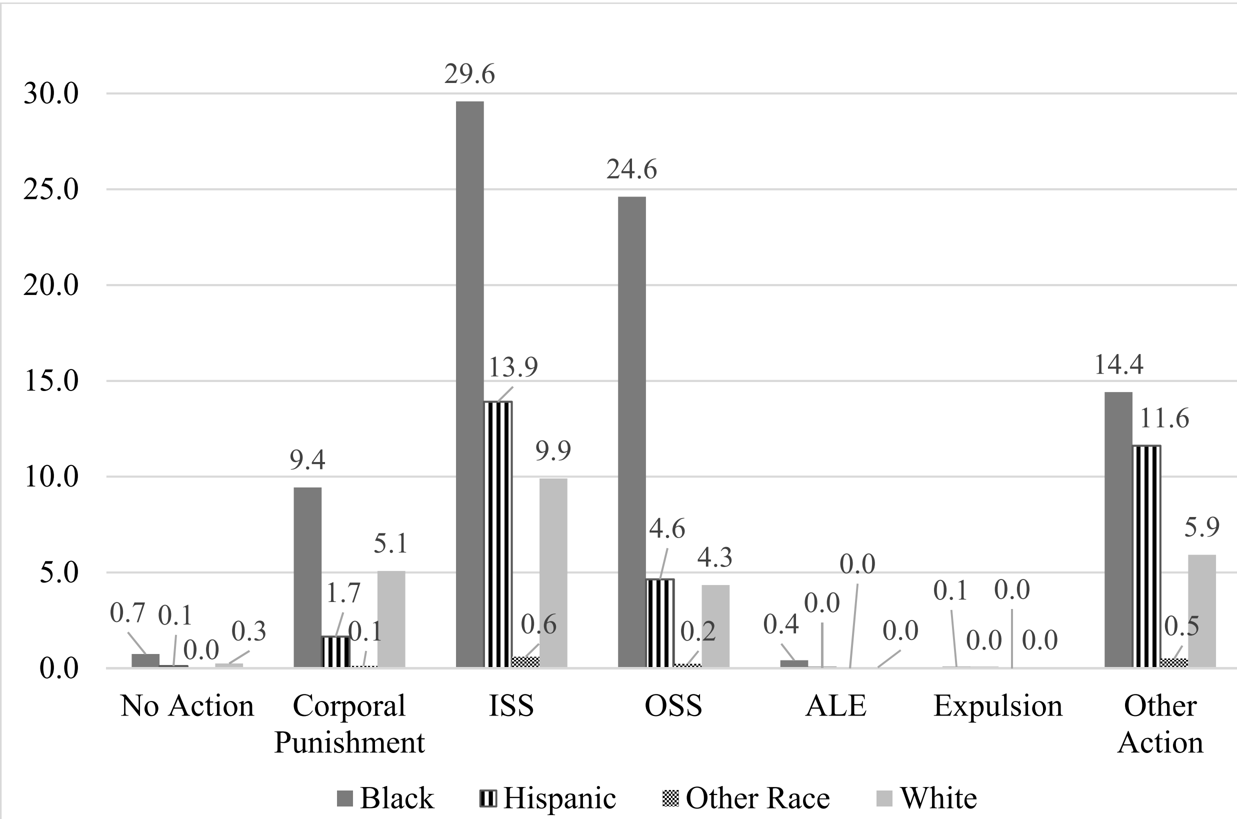

Figure 1. Disciplinary Consequences by Racial Subgroup (Annual Incidences per 100 Students, 2008-09 to 2014-15)

Looking just at the disparities in Figure 1, one might come to the quick conclusion that students are being treated unfairly, but it is also important to connect consequences to the infractions for which the students were referred. Thus, we next consider whether there are disparate rates of referrals for certain types of infractions, and indeed, we see that there are disproportionalities at this level. This does not, however, rule out the possibility that disparities may still exist conditional on infraction type, which we address with Research Questions 1 and 2.

First, a key take-away point from Figure 2 is that the vast majority (almost 80%) of incidences are minor, non-violent offenses (disorderly conduct, insubordination, and "other"). A second point is that Black students are three times more likely than White students to be referred for misbehavior but are nearly six times more likely to be given out-of-school suspensions (24.6 versus 4.3 incidences per 100 students, in Figure 1). These data indicate that Black students are being referred for discipline more often, but this only accounts for about half the difference in the rate of out-of-school suspensions. Our analyses in the next section, using logistic regression to examine incident-level data, helps us to identify more clearly whether there are disparities that still exist conditional on students' reported behavior.

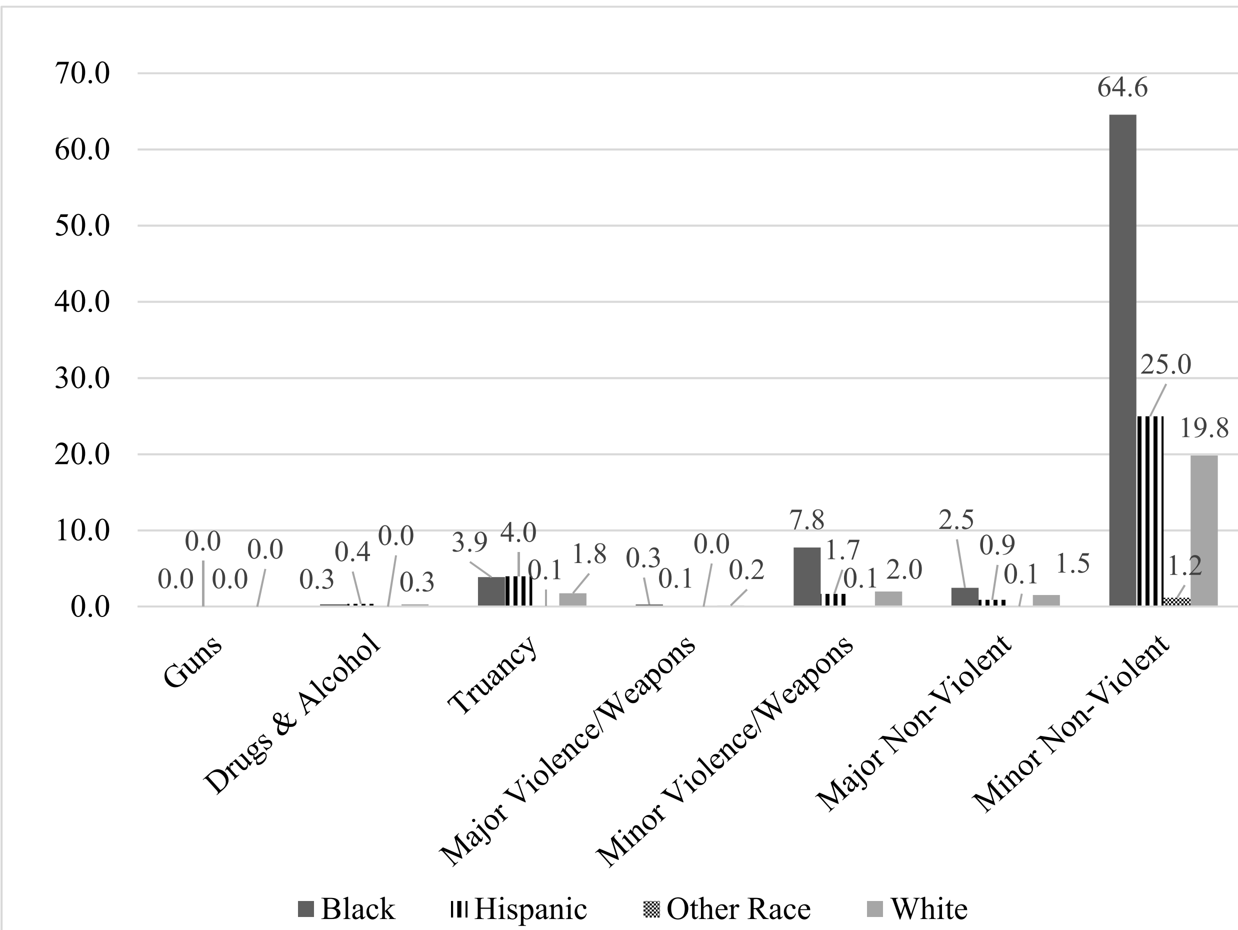

Figure 2. Referrals by Infraction Type and Race (Annual Incidences per 100 Students, 2008-09 to 2014-15)

Research Question 1: Across schools in Arkansas, what, if any, disproportionalities exist in the use of exclusionary discipline for non-White students, low-income students, special education students, or English language learners?

Logistic regression is used to determine the disparities in the likelihood of exclusionary discipline, controlling for the type of infraction committed, and the infraction history and grade level of the student. No school-level factors are taken into account, so this model indicates the extent to which different subgroups of students across the state are disproportionately exposed to exclusionary practices. Any differences by subgroup we find at this level could be due to differences at a variety of levels (across districts or schools, within district, or within school).

Relative risk ratios from several logistic regressions are indicated in Table 5. It is important to note that all models in Table 5, using infraction-level disciplinary data, are conditional on a student being referred for some infraction, so even without controlling for infraction, we can see that, holding constant that a student was referred for any misbehavior, we get a better picture of disciplinary disparities than by just comparing raw numbers of suspensions and expulsions as in Figures 1 and 2. In columns 1-3 of Table 5, we present the results of relatively naïve models that are contingent only upon the student being referred for some disciplinary infraction, but without accounting for infraction type. The primary results, based on models in which we control for the type of infraction committed and for the number of infractions committed by the student during the school year, are presented in columns 4-8.

If disciplinary consequences were handed out evenly across various subgroups of students, we would expect to see relative risk ratios for each indicator (e.g. Black) equal to one. The results in column 4 of Table 5 indicate that Black students are almost 2.5 times as likely to receive exclusionary discipline as their White peers in the same grade for similar types of infractions, with a similar number of previous infractions that year. Hispanic students as well as other non-White students are somewhat less likely than their White peers to receive exclusionary discipline. These figures may represent a lower bound of the actual racial disparities in punishments, as they focus only on the instances of students being assigned to exclusionary discipline, but ignore possibly larger disparities in the number of days punished per suspension.

Looking at columns 5, 6, and 7 of Table 5, instead of testing disproportionalities in exclusionary discipline using race indicators, we use other indicators of a student's disadvantaged status (FRL-eligibility, Special Education status, or English proficiency). FRL-elgible students are about 1.5 times as likely as their non FRL-eligible peers in the state to receive exclusionary discipline. Special education students are slightly more likely to receive exclusionary discipline, and LEP students are about half as likely to receive exclusionary discipline. The model in Column 8 includes the full combination of control variables.

The racial disparities, indicated by the relative risk ratios on Black, Hispanic, and other non-White groups, are similar between columns 1 and 4 and columns 3 and 8. Also, the disparities based on FRL-status are similar between columns 2 and 5. This result indicates the racial disparities in exclusionary discipline are not driven solely by the types of infractions reported. While infraction type, controlled for in columns 4-8, helps explain whether a student receives exclusionary disipline (higher pseudo R-squared), the relative risk ratios for various racial groups do not decline much with the inclusion of infraction-type controls. The stability of these results gives us some confidence in our first key finding, that Black students are more than twice as likely to receive exclusionary discipline after referral for the same infraction types.[8]

[8] While the pseudo R-squared values are somewhat low, they are not equivalent to the R-squared found in OLS regression, so while higher values represent better model fit, they cannot be interpreted exactly the same as the R-squared found in OLS regression. Stata's default pseudo R-

Table 5

Logistic Regression of Exclusionary Discipline (2008-09 to 2014-15)

	(1)	(2)	(3)
Black	2.215 ***		2.132 ***
	(0.021)		(0.020)
Hispanic	0.795 ***		0.838 ***
	(0.016)		(0.022)
Other Minority	0.854 ***		0.878 ***
	(0.033)		(0.035)
FRL-Eligible		1.475 ***	1.224 ***
		(0.014)	(0.012)
Special Education			1.106 ***
			(0.013)
LEP			0.860 ***
			(0.029)
Guns			
Drugs & Alcohol			
Truancy			
Major Violence/Weapons			
Minor Violence/Weapons			
Major Non-Violent			
School Year and Grade Level Indicators	Y	Y	Y
Infraction Order Indicators			
Constant	0.351 ***	0.363 ***	0.297 ***
	(0.072)	(0.073)	(0.061)
Observations	1,243,555	1,243,555	1,243,555
Wald Chi-Squared	11,571	5,645	12,146
Num. of Clusters (Students)	240,999	240,999	240,999
Pseudo R-Squared	0.036	0.012	0.037

Note: Heteroskedastic-robust standard errors in parentheses, clustered at the student level. Baseline infraction category is Minor Non-Violent Infractions.
*** p<0.01, ** p<0.05, * p<0.1.

---

squared value is McFadden's R-squared, which is equivalent to 1 minus the ratio of the log likelihood of the full model to the log likelihood of a simple intercept model (Institution for Digital Research and Education, 2011).

Table 5, Cont'd

Logistic Regression of Exclusionary Discipline (2008-09 to 2014-15)

	(4)	(5)	(6)	(7)	(8)
Black	2.471 ***				2.378 ***
	(0.024)				(0.023)
Hispanic	0.888 ***				0.897 ***
	(0.017)				(0.023)
Other Minority	0.912 **				0.920 **
	(0.035)				(0.036)
FRL-Eligible		1.518 ***			1.232 ***
		(0.015)			(0.012)
Special Education			1.068 ***		1.090 ***
			(0.013)		(0.013)
LEP				0.534 ***	0.922 **
				(0.013)	(0.031)
Guns	16.990 ***	16.270 ***	15.670 ***	15.760 ***	17.220 ***
	(2.854)	(2.621)	(2.485)	(2.502)	(2.908)
Drugs & Alcohol	38.230 ***	29.200 ***	27.880 ***	28.610 ***	38.730 ***
	(1.164)	(0.885)	(0.838)	(0.866)	(1.183)
Truancy	0.570 ***	0.517 ***	0.512 ***	0.524 ***	0.572 ***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
Major Violence/Weapons	17.500 ***	14.810 ***	14.440 ***	14.510 ***	17.330 ***
	(0.587)	(0.487)	(0.472)	(0.477)	(0.584)
Minor Violence/Weapons	6.545 ***	6.489 ***	6.500 ***	6.490 ***	6.521 ***
	(0.058)	(0.056)	(0.056)	(0.056)	(0.058)
Major Non-Violent	2.175 ***	1.898 ***	1.870 ***	1.859 ***	2.167 ***
	(0.025)	(0.021)	(0.021)	(0.021)	(0.025)
School Year and Grade Level Indicators	Y	Y	Y	Y	Y
Infraction Order Indicators	Y	Y	Y	Y	Y
Constant	0.226 ***	0.240 ***	0.347 ***	0.348 ***	0.191 ***
	(0.050)	(0.053)	(0.075)	(0.075)	(0.042)
Observations	1,243,555	1,243,555	1,243,555	1,243,555	1,243,555
Wald Chi-Squared	76,215	73,372	72,321	71,778	76,398
Num. of Clusters (Students)	240,999	240,999	240,999	240,999	240,999
Pseudo R-Squared	0.118	0.092	0.089	0.091	0.119

Note: Heteroskedastic-robust standard errors in parentheses, clustered at the student level. Baseline infraction category is Minor Non-Violent Infractions.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

The results for Research Question 1, discussed previously, are only representative of disparities in disciplinary outcomes across the state. It could be that most of these disparities only occur across schools, or it could be, instead, that disparities also exist within schools. In the next section, we utilize school fixed effects to assess what disproportionalties exist, if any, in disciplinary outcomes for students within the same schools.

Research Question 2: Within schools, what, if any, disproportionalities exist in the use of exclusionary discipline for non-White students, low-income students, special education students, or English language learners?

In this section, logistic regression is again used to assess whether student demographic factors are associated with higher rates of exclusionary discipline, this time for students within the same schools. Relative risk ratios from several logistic regressions, all including school fixed effects, are indicated in Table 6. The results in column 1 indicate that Black students are only slightly more likely to receive exclusionary discipline, relative to their White peers within the same schools. Larger disparities can be seen based on whether the student is FRL-eligible (column 2) or receiving Special Education services (column 3). This result indicates, perhaps, the multiple tiers of privilege or disadvantage – that Black students are disproportionately exposed to exclusionary discipline as a function of the school that they attend, but that, within schools, other factors such as poverty or special education status influence the likelihood of a student receiving exclusionary consequences.

The coefficients on the indicator for Black students are smaller in the school fixed effects models (indeed, in the fullest model, the odds ratio is not statistically different from one), relative to those in the models without school fixed effects, indicating that the racial disparities in exclusionary discipline are driven almost entirely by differences across schools rather than within schools. Because these analyses revealed that the between-school differences are so important, in the following section, we test which characteristics of schools drive these differences.

Table 6

Logistic Regression of Exclusionary Discipline within Schools (2008-09 to 2014-15)

	(1)	(2)	(3)	(4)	(5)
Black	1.035 ***				1.007
	(0.011)				(0.011)
Hispanic	0.935 ***				0.949 **
	(0.016)				(0.021)
Other Minority	1.011				1.023
	(0.033)				(0.033)
FRL-Eligibile		1.165 ***			1.157 ***
		(0.010)			(0.011)
Special Education			1.191 ***		1.180 ***
			(0.012)		(0.012)
Limited English Proficient				0.910 ***	0.935 ***
				(0.018)	(0.024)
Grade Level Indicators	Y	Y	Y	Y	Y
School Year Indicators	Y	Y	Y	Y	Y
Infraction Types	Y	Y	Y	Y	Y
Infraction Order Indicators	Y	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y	Y
Constant	0.081 ***	0.071 ***	0.082 ***	0.082 ***	0.070 ***
	(0.023)	(0.021)	(0.024)	(0.024)	(0.021)
Observations	1,236,401	1,236,401	1,236,401	1,236,401	1,236,401
Number of Students	239,202	239,202	239,202	239,202	239,202
Model Chi-Squared	132,531	132,473	131,941	132,507	132,333
Pseudo R^2	0.324	0.325	0.325	0.324	0.325

Note: Heteroskedastic-robust standard errors in parentheses, clustered at the student level. Baseline infraction category is Minor Non-Violent Infractions.
*** p<0.01, ** p<0.05, * p<0.1.

Research Question 3: What school characteristics are associated with harsher (longer) disciplinary consequences?

Since there are larger racial disparities across the state than within schools, it could be that there are differences in disciplinary policies and practices at the types of schools that serve large proportions of non-White students. We test this by creating a School Severity Index (SSI) for each school using the residuals from an infraction-level model predicting the length of exclusionary punishments. In this model, consequences other than exclusionary discipline (e.g. in-school suspension, corporal punishment, no action, or "other" actions) are coded as zero days of punishment, but are not removed from the model. The residuals are averaged at a school level to generate the school SSI: a positive SSI indicates that a school tends to give out longer (more exclusionary) punishments for similar types of infractions. A negative SSI indicates shorter (less exclusionary) punishments. These SSIs were created using school random effects to account for the noisy measures within schools with fewer disciplinary incidences by allocating greater weight to the schools with larger sample sizes and more precise measures.

The SSIs are then regressed on a variety of school-level characteristics. The results in Table 7 are based on SSIs created in the first-stage using the days of exclusionary punishment (OSS, expulsion, or referral to an ALE). All other (non-exclusionary) types of consequences are included as zero days. Importantly, the R-squared values in the models with the race percentage variables (columns 2, 3, 5, and 6) have about 2.5 times the predictive power of those without the race percentage variables (columns 1 and 4). Therefore, the racial breakdown of schools appears to be an important factor in explaining disciplinary outcomes within schools. The results here are stable across models and consistent with our earlier findings: schools serving greater proportions of Black students had higher scores in the severity index and thus longer punishments; schools serving greater percentages of Hispanic students had lower scores.

Other variables in Table 7 also appear to have significant relationships to the severity of punishment. Open-enrollment charter schools, all else equal, give out somewhat harsher punishments (an extra 0.4 to 1.3 days of punishment, per infraction, depending on the model). Importantly, the charter school coefficient is much lower in the models including controls for the school racial demographics than in the models without these variables. Open-enrollment charter schools in the state are primarily clustered in urban areas and serve a larger proportion of Black students (41%) than the state average (21%), and fewer White students (46%) than the state average (63%). Therefore, without controlling for these racial demographics, the charter school variable is confounded with racial demographics as well. It is possible that charter schools, conditional on student demographics, may use harsher punishments, if, for example, they focus on a so-called "no excuses" model as in the highly successful KIPP Charter Network (Arkansas had two locations in 2014-15, the last year in our dataset). In addition, evidence from Michigan indicates that students may seek out charter schools if they are already having disciplinary issues or other problems in the traditional public school district (Horn & Miron, 2000).

The coefficients on the middle and high school indicators may be surprising. Based on these coefficients alone, middle schools and high schools appear to be administering relatively less severe consequences than elementary schools in the state, for the same types of infractions. While these coefficients are statistically significant, the magnitude is somewhat small (about 0.1 days, per infraction). To understand this counterintuitive result, it is also important to note that the SSI is the average residual from a model predicting the severity (number of days) of a punishment, as a function of a variety of things, including grade level. Therefore, the SSI, in some ways, is already accounting for the increasing severity by grade-level, so we just treat these school-type variables as control variables.

Table 7

School Characteristics Associated with Harsher Punishments (Dep Var = School Severity Index Based on Days of Exclusionary Discipline, Units= Number of Days)

	(1)	(2)	(3)	(4)	(5)	(6)
Log (School Enrollment)	0.144 ***	-0.020	-0.079 ***	0.144 ***	-0.021	-0.073 ***
	(0.019)	(0.017)	(0.019)	(0.019)	(0.017)	(0.019)
School % Black		1.392 ***	1.586 ***		1.377 ***	1.552 ***
		(0.036)	(0.045)		(0.036)	(0.045)
School % Hispanic		-0.800 ***	-0.658 **		-0.887 ***	-0.761 ***
		(0.266)	(0.265)		(0.266)	(0.266)
School % Other Minority		1.716 ***	1.444 ***		1.762 ***	1.518 ***
		(0.274)	(0.275)		(0.274)	(0.276)
School % FRL	1.106 ***		-0.554 ***	1.118 ***		-0.499 ***
	(0.064)		(0.078)	(0.064)		(0.078)
School % Gifted and Talented	2.934 ***	2.093 ***	1.908 ***	2.925 ***	2.083 ***	1.919 ***
	(0.185)	(0.170)	(0.172)	(0.185)	(0.170)	(0.173)
School % Special Education	-0.144	-0.112	-0.027	-0.178	-0.143	-0.064
	(0.168)	(0.155)	(0.154)	(0.168)	(0.155)	(0.155)
School % LEP	0.410 ***	1.901 ***	2.116 ***	0.433 ***	2.012 ***	2.207 ***
	(0.117)	(0.308)	(0.308)	(0.116)	(0.308)	(0.309)
Open-Enrollment Charter	1.294 ***	0.541 ***	0.389 ***	1.292 ***	0.542 ***	0.408 ***
	(0.088)	(0.083)	(0.086)	(0.088)	(0.083)	(0.086)
Middle School	-0.148 ***	-0.085 ***	-0.098 ***	-0.145 ***	-0.082 ***	-0.094 ***
	(0.031)	(0.028)	(0.028)	(0.031)	(0.028)	(0.028)
High School	-0.115 ***	-0.057 **	-0.080 ***	-0.115 ***	-0.058 **	-0.079 ***
	(0.027)	(0.025)	(0.025)	(0.027)	(0.025)	(0.025)
Other School Type	0.680 ***	0.487 ***	0.431 ***	0.690 ***	0.500 ***	0.449 ***
	(0.098)	(0.091)	(0.090)	(0.098)	(0.091)	(0.091)
School Year Indicators	Y	Y	Y	Y	Y	Y
Constant	-1.771 ***	-0.446 ***	0.195	-1.777 ***	-0.434 ***	0.141
	(0.140)	(0.106)	(0.141)	(0.140)	(0.106)	(0.141)
Observations	6,871	6,891	6,871	6,871	6,891	6,871
R-squared	0.096	0.236	0.241	0.098	0.233	0.238

Note. Standard errors in parentheses. Models 1-3 use SSI created with all 17 infraction categories in the first stage. Models 4-6 use SSI created with the 7 infraction groups in the first stage. In the first stage, SSI were created using school random effects. Baseline school type is Elementary. *** p<0.01, ** p<0.05, * p<0.1.

Robustness Checks

As a robustness check, we estimated these SSI models (as in Table 7) excluding the log of school enrollment. The results are in Tables 8. In all cases, the coefficients on the racial groups in Table 8 are slightly smaller (or more negative) than those in Table 8, but generally, the estimated relationships in Tables 7 and 8are quite similar.

In addition, we also conducted robustness checks using days of *any* type of punishment (not just exclusionary). The results for the primary variable of interest (school percent Black) are generally similar. There are some differences, however, in terms of the coefficients on the school percent Hispanic, which has a significantly negative relationship with SSI when created using only exclusionary discipline, but a non-significant relationship in terms of days of any type of consequence. This indicates that, all else equal, while there is no relationship between the proportion of Hispanic students and the length of any type of disciplinary consequence, it is the case that schools with a greater proportion of Hispanic students generally give out shorter exclusionary type punishments.

There is a surprising result from the models that include a measure of the percent of students who are FRL-eligible as well as percent Black (columns 3 and 6 in Table 7). The coefficients on the school percent FRL are negative, despite the fact that the coefficients on FRL are positive in the models that do not control for percent Black (columns 1 and 4 in Table 7). This finding is likely due to significant correlation between the percentage of students who are FRL-eligible and the percentage of students who are Black in each school (r=0.499).

Table 8

School Characteristics Associated with Harsher Punishments (Dep Var = School Severity Index Based on Days of Exclusionary Discipline, Units= Number of Days)

	(1)	(2)	(3)	(4)	(5)	(6)
School % Black		1.388 ***	1.522 ***		1.373 ***	1.492 ***
		(0.035)	(0.042)		(0.035)	(0.042)
School % Hispanic		-0.816 ***	-0.745 ***		-0.904 ***	-0.841 ***
		(0.265)	(0.265)		(0.266)	(0.265)
School % Other Minority		1.677 ***	1.401 ***		1.722 ***	1.478 ***
		(0.272)	(0.275)		(0.272)	(0.276)
School % FRL	0.937 ***		-0.403 ***	0.948 ***		-0.359 ***
	(0.060)		(0.069)	(0.060)		(0.069)
School % Gifted and Talented	2.892 ***	2.093 ***	1.966 ***	2.883 ***	2.083 ***	1.973 ***
	(0.186)	(0.170)	(0.172)	(0.185)	(0.170)	(0.172)
School % Special Education	-0.436 ***	-0.061	0.110	-0.471 ***	-0.090	0.063
	(0.165)	(0.149)	(0.151)	(0.164)	(0.149)	(0.151)
School % LEP	0.619 ***	1.903 ***	2.064 ***	0.643 ***	2.014 ***	2.159 ***
	(0.114)	(0.308)	(0.308)	(0.114)	(0.308)	(0.309)
Open-Enrollment Charter	1.169 ***	0.559 ***	0.489 ***	1.166 ***	0.561 ***	0.500 ***
	(0.087)	(0.081)	(0.083)	(0.087)	(0.081)	(0.083)
Middle School	-0.127 ***	-0.089 ***	-0.109 ***	-0.124 ***	-0.086 ***	-0.105 ***
	(0.031)	(0.028)	(0.028)	(0.031)	(0.028)	(0.028)
High School	-0.119 ***	-0.058 **	-0.079 ***	-0.119 ***	-0.059 **	-0.077 ***
	(0.027)	(0.025)	(0.025)	(0.027)	(0.025)	(0.025)
Other School Type	0.673 ***	0.487 ***	0.447 ***	0.683 ***	0.500 ***	0.464 ***
	(0.098)	(0.091)	(0.090)	(0.098)	(0.091)	(0.091)
School Year Indicators	Y	Y	Y	Y	Y	Y
Constant	-0.793 ***	-0.568 ***	-0.356 ***	-0.796 ***	-0.559 ***	-0.371 ***
	(0.053)	(0.037)	(0.052)	(0.052)	(0.037)	(0.052)
Observations	6,871	6,891	6,871	6,871	6,891	6,871
R-squared	0.089	0.236	0.239	0.091	0.233	0.236

Note. Table corresponds to Table 7 but with exclusion of log of school enrollment as a covariate. Standard errors in parentheses. Models 1-3 use SSI created with all 17 infraction categories in the first stage. Models 4-6 use SSI created with the 7 infraction groups in the first stage. In the first stage, SSI were created using school random effects. Baseline school type is Elementary. *** p<0.01, ** p<0.05, * p<0.1.

To further understand what is happening within schools in terms of both racial and economic demographics, we created indicators for four types of schools (Low-Income Mostly White, Low-Income Mostly non-White, Higher-Income Mostly White, and Higher-Income Mostly non-White). These four categories are based on whether a school is above or below the state average on two separate indicators (percent White and percent FRL). The state averages during the study period were about 65% White and about 60% FRL. The uneven distribution of observations across these groups, as in Table 9, reflects the relative presence of these types of schools in the state, in the sense that there are relatively few schools that are mostly-non-White and higher-income (8%), relative to the other three types.

Table 9

Distribution of Four School Types

		Higher-Income <60% FRL	Low-Income ≥60%FRL
Mostly-Non-White	<65% White	585 School Year Combinations (8%)	2,185 School-Year Combinations (32%)
Mostly-White	≥65% White	2,237 School-Year Combinations (32%)	1,886 School-Year Combinations (27%)

According to the results in Table 10, it seems that the schools with more non-White students (regardless of whether those schools tend to be higher income or lower income), tend to administer harsher (longer) punishments than the baseline schools serving more White students and higher-income students. The first set of coefficients of interest indicates that schools serving more non-White students, who are also higher income, still receive an additional half a day (roughly) of exclusionary discipline, per infraction, relative to their peers in schools serving more White students. Similarly, the third set of coefficients of interest indicates that students in these relatively poor, relatively non-White schools, receive about 0.6 days of extra punishment, relative to students in the relatively wealthy, relatively White schools. Therefore, these two findings indicate that schools with more non-White students tend to give out longer punishments, regardless of the percentage of students receiving FRL in a school.

Table 10

School Characteristics Associated with Harsher Punishments (Dep Var = School Severity Index Based on Days of Exclusionary Discipline, Units= Number of Days)

	(1)	(2)	(3)	(4)
Log (School Enrollment)	0.002	-0.004	0.004	0.000
	(0.019)	(0.018)	(0.019)	(0.018)
Schools serving <60%FRL, <65% White	0.479 ***	0.544 ***	0.471 ***	0.537 ***
	(0.041)	(0.040)	(0.041)	(0.040)
Schools serving ≥60%FRL, ≥65% White	0.033	0.009	0.042	0.019
	(0.028)	(0.028)	(0.028)	(0.029)
Schools serving ≥60%FRL, <65% White	0.624 ***	0.611 ***	0.619 ***	0.609 ***
	(0.028)	(0.026)	(0.028)	(0.026)
School % Gifted and Talented	2.235 ***		2.235 ***	
	(0.182)		(0.182)	
School % Special Education	-0.270		-0.301 *	
	(0.165)		(0.165)	
School % LEP	-0.062		-0.018	
	(0.116)		(0.116)	
Open-Enrollment Charter	0.782 ***	0.599 ***	0.785 ***	0.604 ***
	(0.088)	(0.085)	(0.088)	(0.085)
Middle School	-0.125 ***	0.015	-0.122 ***	0.017
	(0.030)	(0.028)	(0.030)	(0.028)
High School	-0.111 ***	0.013	-0.112 ***	0.011
	(0.026)	(0.025)	(0.027)	(0.025)
Other School Type	0.563 ***	0.526 ***	0.575 ***	0.534 ***
	(0.096)	(0.095)	(0.096)	(0.095)
School Year Indicators	Y	Y	Y	Y
Constant	-0.396 ***	-0.252 **	-0.404 ***	-0.274 **
	(0.123)	(0.113)	(0.123)	(0.113)
Observations	6,891	6,892	6,891	6,892
R-squared	0.142	0.122	0.140	0.120

Note. Standard errors in parentheses. Models 1-2 use SSI created with all 17 infraction categories in the first stage. Models 3-4 use SSI created with the 7 infraction groups in the first stage. In the first stage, SSI were created using school random effects. Baseline school type is Elementary. Baseline school type is schools serving <60%FRL and ≥65% White students.
*** p<0.01, ** p<0.05, * p<0.1.

One of the most interesting results, however, is that there were no significant differences between the length of punishments in wealthier and less wealthy schools, conditional on serving ≥65% White students. This finding suggests that racial factors appear more important than income factors for predicting the severity of disciplinary consequences. This result is consistent with our earlier models (Table 7); the magnitude and sign on the race variable is mostly unchanged by the inclusion of the poverty variable in the model. On the other hand, the poverty result is very sensitive to the inclusion of the race variable. In Table 10, open-enrollment charter schools still appear to use more severe consequences (about an extra 0.6 to 0.8 days of punishment, per infraction).

As an additional robustness check, we also created the SSI using days of any kind of punishment, rather than only the days of exclusionary discipline. The coefficients for each of the four school types in this additional model are nearly identical to the coefficients from the primary model reported in Table 10.

Discussion and Conclusions

There have been numerous studies over the past twenty years documenting the existence of racial disparities in disciplinary consequences. From this research base, we know, with some confidence, based on multiple studies across many years, that Black students are both referred for discipline more often and receive exclusionary disciplinary more often than other students (e.g. Skiba et al., 2002). Moreover, some more recent studies have suggested that Black students receive more severe and longer consequences than their peers, who have committed identical infractions (e.g. Skiba et al., 2002). Finally, based on a single recent study in a single state for a single year (Skiba et al., 2014), along with inferences we can draw from a 1998 survey and consequence-data from Chicago (Sartain et al., 2015), we believe these disparities are more likely driven by differences between schools than by differential treatment of students within a given school.

In this study, we aimed to build on this growing research base by analyzing all infraction-level disciplinary data for every public school in Arkansas over a seven year time span. Consistent with earlier evidence, we found disproportionate use of exclusionary discipline for Black students, and we also found that these disparities are primarily due to differences in discipline practices across schools, rather than within schools. This result supports the important work of Skiba et al. (2014), and builds upon that work by providing analysis of an entire state over seven school years, rather than just one school year.

When school fixed effects are not included, Black students are about 2.4 times as likely as their White peers in the state (in the same grade and with similar numbers of previous infractions) to receive exclusionary discipline for similar infraction types. Hispanic students are slightly less likely than their similar White peers in the state to receive exclusionary discipline. Importantly, the disparities are not only based on race. Depending on whether or not race was also controlled for, our results indicate that FRL students in the state are about 1.2 to 1.5 times as likely to receive exclusionary discipline as their non-FRL peers.

However, we conclude that most of the racial differences in rates of exclusionary discipline are across schools, because these racial disparities diminished greatly when school fixed effects were included. Within schools, Black students are only slightly more likely than White students to receive exclusionary discipline (relative risk ratio of 1.04, significant at the 99% confidence level). Within schools, there still appear to be persistent gaps in the use of exclusionary discipline for FRL students and special education students (relative risk ratios of about 1.2). The implications of this for state or district-level policy are not exactly clear.

These results indicate that the large racial disparities tend to be across schools, and therefore a function of the types of schools that non-White students are likely to attend, whereas within schools, there may be larger concerns about disparities based on socio-economic status and special education status. Since the results indicate that the state-level racial disparities are likely a function of the school attended, we also tested which school-level factors were associated with a measure of school disciplinary severity (SSI), and found that the percent of the school that is Black or the percent of the school that is of another non-White, non-Black, non-Hispanic group are both significant predictors of harsher (longer) consequences, which supports the idea that most of the racial disparities occur due to different disciplinary practices being used in districts/schools serving different racial compositions of students.

When schools were split into four categories based on the proportions of FRL students and White students in the school, we found that schools serving more non-White students (regardless of the proportion of FRL students) administered longer punishments than schools serving mostly White and mostly non-FRL students. However, lower income, mostly White schools were actually quite similar to the higher income, mostly White schools, again indicating that differences in exclusionary practices across schools appear to be more driven by racial demographics than by income or poverty.

Overall, then, there seem to be two broad conclusions from this work. First, non-White students are far more likely to receive exclusionary discipline for a given infraction than their White peers, and this disparity is driven by disciplinary practices employed at the schools non-White students attend. Second, the differences by race are far more impactful than the differences by poverty status (measured by FRL). So, what are the implications for policy?

Based on the analyses presented here, and on our interactions with state level policymakers, we believe there are two broad lessons for policymakers and school leaders, the first related to data transparency and the second to targeted reforms.

First, we have uncovered patterns in disciplinary consequences that were previously not well known in the education community. Thus, we believe that a critical first step in creating positive change with regard to student discipline is broadly sharing discipline data with education stakeholders including staff, administration, families, and communities. For example, school leaders and state policymakers would benefit from reports that allow for comparisons of disciplinary practices and statistics across schools. Given that the disparities are primarily across schools and not within, school leaders may not be aware of a problem until they can compare their school to others in the state. When awareness of potential disparities is raised, school leaders may seek out more concrete programs or strategies to address such issues. It is possible that simply sharing data on school-level rates of exclusion may create awareness that serves as a catalyst for action within communities as well. Moreover, as Tatto et al. (2001) have shown, parental perceptions of unequal or overly strict disciplinary practices can undermine school culture. When policymakers and/or school leaders actively share discipline data, parents can be empowered to advocate for their children and work with school leaders to devise solutions.

Second, the primary conclusion policymakers should draw from our analyses is that the clearest evidence of racial disparities in discipline occur across schools. That is, schools serving predominantly Black students impose more severe (longer) exclusionary consequences on students, even after controlling for the type of infraction. Thus, to address these disparities statewide, policymakers can focus on these particular schools which serve mostly Black students and are engaging in particularly severe disciplinary practices.

One strategy that policymakers might adopt would be to mandate reductions in suspensions – at least for minor non-violent infractions – in targeted schools. This sort of change could be impactful as nearly half (46%) of the infractions that lead to exclusionary discipline are minor, non-violent, and subjective. These infractions include disorderly conduct (~26%) and insubordination (~20%). In addition, it has been argued using OSS for truancy (another non-violent offense which represents 3% of total infractions in our dataset) is hard to justify, as it further removes truant students from the learning environment (Smink & Heilbrunn, 2005; U.S. Departments of Education and Justice, 2014). Perhaps states might aim to eliminate these types of counterintuitive consequences. However, as I find in Chapter 4 of this dissertation, a state-level policy eliminating the use of OSS for truancy was not implemented well and had very little impact on important school outcomes. Again, perhaps, this suggests the need for a more targeted approach, focusing on particular schools with the highest rates of OSS, and supplementing policies with programmatic interventions.

It seems possible to address minor non-violent infractions with preventative or restorative alternatives to exclusionary discipline, if local school districts choose to do so. For example, there is some evidence that simply revising codes of conduct (or setting policies) to reduce the use of suspensions for minor offenses and limit the length of suspensions may be effective (Lacoe & Steinberg, 2016; Mader, Sartain, & Steinberg, 2016) and at little cost to school climate (Mader et al., 2016).

Further, there are school-based interventions, some of which have been rigorously evaluated, designed to improve school climate and disciplinary outcomes. For example, there are non-experimental studies that find reductions in referrals or suspensions and expulsions with programs such as Response to Intervention (RTI), which attempts to prevent recidivism by responding to behavioral issues as they arise (Fairbanks, Sugai, Guardino, & Lathrop, 2007). Another strategy, commonly known as restorative justice, is viewed as a movement to "institutionalize peaceful and non-punitive approaches for addressing harm" that in a school setting can serve as an alternative to exclusionary discipline (Fronius, Persson, Guckenberg, Hurley, & Petrosino, 2016). Essentially, restorative justice is a non-punitive approach to handling conflicts, but these programs can take the form of whole-school interventions or "add-ons" to respond to specific situations (Fronius et al., 2016). Finally, School-Wide Positive Behavioral Interventions and Supports (SWPBIS a.k.a. PBIS) may be the most well-known behavioral intervention and, fortunately, has been subject to some rigorous evaluation. Indeed, there is some experimental evidence that indicates implementation of the PBIS framework improves student perceptions of school safety and test scores (Horner et al., 2009). Other non-experimental studies have linked PBIS to fewer disciplinary incidents (Flannery, Fenning, Kato, & McIntosh, 2014; Freeman et al., 2015) and increased attendance (Freeman et al., 2015) in high schools implementing the framework with fidelity. As of June 2016, there are at least 49 Arkansas schools implementing PBIS (Saarnio & Merten, 2016).

While mandated reductions in exclusionary discipline may be appealing and are certainly simple, we have two reservations about this approach. First, it is possible that school leaders may respond to mandates superficially, by changing reporting patterns without substantially improving their disciplinary practices. For example, I find, in Chapter 4 of this dissertation, that when Arkansas prohibited OSS as a consequence for truancy, this policy was associated with a rise in "other" non-specified consequences for truancy, as well as "other" infractions. Unfortunately, we know little about what is actually included in these categories.

Our second reservation is that mandated reductions without any other supports are unlikely to be effective. In fact, evidence from Arkansas suggests that a reduction in OSS will likely not improve student achievement as measured by test scores (Chapter 3 of this dissertation), and that when the state eliminated OSS as a legal consequence for truancy, there was no policy-related change in school level outcomes such as test scores, attendance, and chronic absenteeism (Chapter 4 of this dissertation). I conclude that this second result was due, at least in part, to a lack of communication from the state to school leaders about the policy, a lack of accountability for compliance with the policy, and a lack of resources and capacity for schools with the greatest need to adequately respond. Front line educators in the schools need effective alternatives to exclusionary discipline if such mandates are put in place, and school culture is unlikely to improve if educators do not have the necessary capacity to respond to behavioral infractions. Thus, if state policymakers are to mandate reductions, they should also consider providing schools with access to more positive alternative disciplinary strategies.

While we advocate for data transparency, we understand that there are limitations to the conclusions that should be drawn and there is a real potential for unintended consequences. First of all, it is not obvious that high numbers of disciplinary referrals and consequences are bad – or good. For example, a school with very few reported infractions may either be one with a great school climate, or one where administrators fail to address real problems related to student discipline. Thus, school context matters. Finally, according to Campbell's (1979) Law, which states that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor," the greater public attention paid to disciplinary data may have the unintended effect of encouraging school personnel to simply under-report or code disciplinary infractions and consequences in vague categories (such as “other”).

We are not suggesting that policymakers ignore disciplinary data, but that these numbers be interpreted with appropriate caution and in context. While the discussion of disciplinary disparities has been ongoing for several decades, the practice of public reporting of school level discipline data is relatively new. Thus, while policymakers should pay attention to these data, we would argue that it is premature to attach high-stakes consequences to disciplinary outcomes.

Ultimately, while the results presented here do not provide step-by-step solutions, they do provide further confidence in the early findings from the research literature that Black students face disciplinary disparities, even conditional on the type of infraction reported. We have also provided some information about which schools in Arkansas are more likely to impose relatively severe consequences that remove students from classrooms. The first step in addressing a potential problem is identifying it. It is our hope that policymakers and researchers and school leaders collaborate on the next step: to implement potentially effective strategies and rigorously evaluate the results to improve the schooling experience for students in the future.

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.

Anderson, K. P. & Ritter, G. W. (2017). *Do school discipline policies treat students fairly? A second look at school discipline rate disparities.* EDRE Working Paper No. 2015-11. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2700707

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arkansas Code § 6-17-112 (2016). Corporal punishment – immunity from liability. Retrieved from http://law.justia.com/codes/arkansas/2016/title-6/subtitle-2/chapter-17/subchapter-1/section-6-17-112/

Arkansas Code § 6-18-502 (2016). Guidelines for development of school district student discipline policies. Retrieved from http://law.justia.com/codes/arkansas/2016/title-6/subtitle-2/chapter-18/subchapter-5/section-6-18-502/

Arkansas Code § 6-18-503 (2016). Written student discipline policies required. Retrieved from http://law.justia.com/codes/arkansas/2016/title-6/subtitle-2/chapter-18/subchapter-5/section-6-18-503/

Arkansas Code § 6-18-507 (2016). Suspension – Expulsion. Retrieved from http://law.justia.com/codes/arkansas/2016/title-6/subtitle-2/chapter-18/subchapter-5/section-6-18-507/

Arkansas Code § 6-48-104 (2016). Department of Education responsibilities. Retrieved from http://law.justia.com/codes/arkansas/2016/title-6/subtitle-3/chapter-48/section-6-48-104/

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. Retrieved from https://doi.org/10.1525/sop.2012.55.4.637

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67-90.

Caputo, I. (March 15, 2017). Bid to eliminate corporal punishment in schools fails. *Arkansas Blog.* [Web log post]. Retrieved from https://www.arktimes.com/ArkansasBlog/archives/2017/03/15/bid-to-eliminate-corporal-punishment-in-schools-fails

Children's Defense Fund. (1975). School Suspensions: Are they Helping Children? Washington Research Project, Cambridge, MA. Retrieved from http://files.eric.ed.gov/fulltext/ED113797.pdf

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.

Curran, F. C. (2016). Estimating the Effect of State Zero Tolerance Laws on Exclusionary Discipline, Racial Discipline Gaps, and Student Behavior. *Educational Evaluation and Policy Analysis, 38*(4). Retrieved from https://doi.org/10.3102/0162373716652728

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. J. (2014). The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Statistics and Public Policy, 1*(1), 19-27.

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. Retrieved from https://doi.org/10.1177/001440290707300302

Flannery, K. B., Fenning, P., Kato, M. 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.

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2015). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions, 18*(1).

Fronius, T., Persson, H., Guckenberg, 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

Horn, J., & Miron, G. (2000). *An evaluation of the Michigan charter school initiative: Performance, accountability, and impact.* The Evaluation Center, Western Michigan University. Retrieved from http://homepages.wmich.edu/~miron/publics/michigan/Executive_Summary.pdf

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.

Institute for Digital Research and Education. (2011). UCLA. Retrieved from http://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/

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.

Langille, S. (July 24, 2017). Texas school district approves corporal punishment 'by way of paddling.' *Washington Examiner.* Retrieved from http://www.washingtonexaminer.com/texas-school-district-approves-corporal-punishment-by-way-of-paddling/article/2629258

Losen, D., & Gillespie, J. (2012). *Opportunities Suspended: The Disparate Impact of Disciplinary Exclusion from School.* Civil Rights Project, UCLA. Retrieved from http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/upcoming-ccrr-research/losen-gillespie-opportunity-suspended-2012.pdf

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.

Mader, N., Sartain, L., & Steinberg, M. (2016). *When suspensions are shorter: The effects on school climate and student learning.* Working paper.

Moynihan, D. P. (1965). *The Negro family: The case for national action.* Washington, D.C.: U.S. Government Printing Office, 1965

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.

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/

Pearlstein, M. (2011). *From family collapse to America's decline: The educational, economic, and social costs of family fragmentation.* Lanham, MD: Rowman & Littlefield Education.

Pew Research Center. (2015). *2014 Religious Landscape Study.* Interactive Tool. Retrieved from http://www.pewforum.org/religious-landscape-study/

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.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin 13: 19–23.* Reprinted in Stata Technical Bulletin Reprints, Vol. 3, 88–94. College Station, TX: Stata Press.

Saarnio, D. & Merten, A. (June 28, 2016). PBIS State Contacts for Arkansas. Retrieved from http://cce.astate.edu/pbis/wp-content/uploads/2016/07/PBIS_Schools_in_Arkansas_-for-Natl-PBIS-Center_6-28-163.pdf

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.

Skiba, R. (2014). The Failure of Zero Tolerance. *Reclaiming Children and Youth, 22*(4), 27-33.

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., 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., 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. Retrieved from https://doi.org/10.1023/A:1021320817372

Skiba, R., & Peterson, R. (January 1999). ZAP Zero Tolerance. *Phi Delta Kappan, 80,* 372-76, 381-82.

Skiba, R., & 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.

Smink, J., & Heilbrunn, J. Z. (2005). *The legal and economic implications of truancy*. Clemson, SC: National Dropout Prevention Center/Network. Retrieved from http://files.eric.ed.gov/fulltext/ED491290.pdf

Sullivan, A. L., Klingbeil, D. A., & Van Norman, E. R. (2013). Beyond behavior: Multilevel analysis of the influence of sociodemographics and school characteristics on students' risk of suspension. *Children, Research, and Public Policy, 42*(1), 99-114.

Tatto, M. T., Rodriguez, A., Gonzalez-Lantz, D., Miller, C., Busscher, M., Trumble, D., Centeno, R., & Woo, A. The challegnes and tensions in reconstructing teacher-parent relations in the context of school reform: a case study. *Teachers and Teaching: Theory and Practice, 7*(3), 315-333.

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 and U.S. Department of Justice. (2014). Dear Colleague Letter on the Nondiscriminatory Administration of School Discipline. Retrieved from http://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html

Welch, K., & Payne, A. N. (2010). Racial threat and punitive school discipline. *Social Problems, 57*(1), 25-48.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48,* 817–838.

Chapter 3

Understanding a Vicious Cycle:

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

Kaitlin P. Anderson
University of Arkansas

Gary W. Ritter, Ph.D.
University of Arkansas

Gema Zamarro, Ph.D.
University of Arkansas

Abstract

A vast body of research has found a 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. However, 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 aim at estimating the causal relationship. We estimate dynamic panel data models incorporating student fixed effects using Anderson-Hsiao methods. We find, counterintuitively, a generally null (and sometimes slightly positive) impact of out-of-school suspensions on test scores. Therefore, while policymakers may have other reasons to limit exclusionary discipline, we conclude we should not expect academic gains to follow from a reduction in OSS without additional school interventions or supports.

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, 2017; 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).

We know little about the causal effect of exclusionary discipline on student outcomes, yet a vast amount of prior work indicates correlational evidence. 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, 2003; Raffaele-Mendez, Knoff, & Ferron, 2002; Rausch & Skiba, 2005; Skiba & Rausch, 2004), school drop-out and grade retention (American Academy of Pediatrics, 2013; American Psychological Association, 2008; Balfanz, Byrnes, & Fox, 2014; Cobb-Clark et al., 2015; Ekstrom, Goertz, Pollack, & Rock, 1986; Fabelo et al., 2011; Gregory & Weinstein, 2008; Krezmien, Leone, & Achilles, 2006; Marchbanks et al., 2014; Raffaele-Mendez, 2003; Raffaele-Mendez & Sanders, 1981; Rodney, Crafter, Rodney, & Mupier, 1999; Stearns & Glennie, 2006; Wald & Kurlaender, 2003), and involvement in the juvenile justice system (American Academy of Pediatrics, 2013; Balfanz, Spiridakis, Neild, & Legters, 2003; Fabelo et al., 2011; Nicholson-Crotty, Birchmeier, & Valentine, 2009).

If these relationships are causal rather than simply correlational, the economic impact of reducing out-of-school suspensions 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, & Catalano, 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; Stallings, Cory, Fairweather, & Needels, 1978; Wang, Haertel, & Walberg, 1997). This argument is consistent with studies that find suspensions precede lower performance although these findings are not necessarily causal (Balfanz et al., 2014; Cobb-Clark et al., 2015, McIntosh, Flannery, Sugai, Braun, & Cochrane, 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 future suspensions, 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 the opposite - that low academic performance predicts a variety of undesirable behaviors in the future (Arcia, 2006; Choi, 2007; McIntosh et al., 2008; Miles & Stipek, 2006). 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 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.

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. It could also be that the effects work both ways, or that other unobservable factors are causing both suspensions and poor achievement. Therefore, the ability to direct public policy based on correlational studies is limited. However, despite the ambiguity, 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, nonpunitive) discipline strategies; and provide support services such as counseling, dropout prevention, and guidance services for at-risk students" (Steinberg & Lacoe, 2017, 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 & Lacoe, 2017).

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. For example, it may help identify the mechanism through which students are affected or the types of students most affected. 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, the declining achievement of a struggling student and the associated disengagement from school leads to disciplinary problems, or causality works in both directions. 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. Policy discussion around reducing suspensions requires causal evidence of the impacts on students, but sorting out the causal effect is a 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 suspension (OSS) 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 et al., 1986; Fabelo et al., 2011; Gregory & Weinstein, 2008; Krezmien, Leone, & Achilles, 2006; Marchbanks et al., 2014; Raffaele-Mendez, 2003; Raffaele-Mendez & Sanders, 1981; Rodney et al., 1999; Stearns & Glennie, 2006; Wald & Kurlaender, 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 English Language Arts (ELA) test scores.

In this study, we focus on the academic impacts on suspended students, but many have hypothesized that disciplinary practices, which influence school culture and perceived or actual school safety, might affect the academic achievement of other students in the school. One study found that high levels of suspensions are 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, causal inference from these studies is limited by the potential for reverse causality and confounding effects of factors that influence 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 OSS on student test scores. Next, we describe the data utilized for this study and the analytic sample.

Data and Sample

This study uses six years of de-identified student demographic, achievement (test score), and disciplinary data from all K-12 schools in Arkansas provided by the Arkansas Department of Education (ADE) for 2008-09 through 2013-14. Demographic data include race/ethnicity, gender, grade, special education status, limited English proficiency (LEP) 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). Math and English Language Arts (ELA) exams were administered in grades 3-8, and End of Course (EOC) examinations were administered in Algebra I (typically 9th grade), Geometry (typically 10th grade), Algebra II (typically 11th grade), and 11th Grade Literacy.[9] 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 and scaling methods.

Discipline data are provided at the individual infraction level and indicate which type of infraction (out of 19) was recorded, the corresponding consequence (out of 13 types), the infraction date, and the length of the consequence. These infraction-level data are linked to students and schools. To simplify the analysis, we grouped similar infraction types, resulting in

[9] 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.

12 groups.[10] See Table 1 for the frequency of each infraction type. The three most common types of infractions, disorderly conduct (31.0%), insubordination (25.9%), and “other” non-specified infractions (21.7%),[11] represent almost 80% of all infractions during the study period.

Furthermore, 13 consequence categories were collapsed into 7 (in school suspension (ISS), OSS, expulsion, referral to an alternative learning environment (ALE), corporal punishment, no action, and “other”). [12] See Table 2 for the frequency of each disciplinary consequence type. The most common consequence types during the study period were ISS (39.2%), OSS (24.1%), “other” non-specified consequences (18.8%), 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 specify the number of times in a given school year the student was cited for some particular type of infraction and received some particular type of consequence. In addition, days of each type of punishment, when applicable, were aggregated to a student-by-school year level. These disciplinary data are merged with the student level demographic and achievement data using unique student identifiers.

---

[10] 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.

[11] “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. This category was provided by the ADE and not researcher-created.

[12] “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. This category was provided by the ADE and not researcher-created.

Table 1

Infraction Types, By Year (Arkansas) from 2008-09 to 2013-14

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	Total	% of Total
Disorderly Conduct	54,641	51,027	48,765	51,539	42,575	57,750	306,297	31.0%
Insubordination	47,273	46,151	45,765	38,798	34,759	43,068	255,814	25.9%
Other	31,871	28,639	26,481	31,858	35,024	60,600	214,473	21.7%
Fighting	12,378	12,456	12,471	12,136	12,434	13,128	75,003	7.6%
Truancy	9,968	11,834	11,734	10,465	9,407	12,914	66,322	6.7%
Bullying	3,455	4,099	4,363	4,483	4,515	5,496	26,411	2.7%
Alcohol and Tobacco	2,512	2,552	2,298	2,209	2,286	2,835	14,692	1.5%
Student or Staff Assault	2,148	2,132	1,892	1,959	2,361	2,503	12,995	1.3%
Drugs	944	996	954	1,146	1,259	1,295	6,594	0.7%
Vandalism	962	833	909	689	736	1,084	5,213	0.5%
Weapons	509	515	525	517	562	671	3,299	0.3%
Gangs	361	339	177	107	131	103	1,218	0.1%
Total	167,022	161,573	156,334	155,906	146,049	201,447	988,331	100.0%
% of Total	16.9%	16.3%	15.8%	15.8%	14.8%	20.4%	100.0%	

Table 2

Consequence Types, By Year (Arkansas) from 2008-09 to 2013-14

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	Total	%
In-School Suspension	63,018	64,760	60,052	62,532	63,019	74,169	387,550	39.2%
Out-of-School Suspension	41,348	39,613	36,780	37,791	40,233	42,290	238,055	24.1%
Other	23,120	23,858	27,600	26,482	21,850	62,972	185,882	18.8%
Corporal Punishment	36,484	30,732	29,311	27,760	19,142	19,746	163,175	16.5%
No Action	1,999	1,494	1,777	993	1,288	1,435	8,986	0.9%
ALE	918	794	621	253	317	586	3,489	0.4%
Expulsion	135	322	193	95	200	249	1,194	0.1%
Total	167,022	161,573	156,334	155,906	146,049	201,447	988,331	100.0%
% of Total	16.9%	16.3%	15.8%	15.8%	14.8%	20.4%	100.0%	

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. For these students, data for years prior to an expulsion or ALE referral were also excluded from the sample. We seek to estimate the impact on students who are suspended out-of-school, relative to receiving some other, non-exclusionary consequence, so excluding expelled students and students referred to an ALE ensures the reference category of consequence (ISS, corporal punishment, no action, or “other”) is less exclusionary than OSS. Thus, 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. 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 combined impact of all three kinds of exclusionary discipline on student test scores.

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 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 estimates are meant to be understood as correlation coefficients as the model most probably suffers from endogeneity as the error terms ε_{it} are likely correlated to other explanatory variables in the model due to potential reverse causation and other confounding factors.

Our pooled OLS model (the correlational benchmark) is:

$$y_{it} = \beta_1 OSSdays_{it-1} + \beta_2 infractioncount_{it-1} + \beta_3 gradelevel_{it} + \beta_4 schoolyear_{it} + \beta_5 y_{it-1} + \beta_6 X_{it} + d_{it} + \varepsilon_{it} \tag{1}$$

The variable of interest, $OSSdays_{t-1}$, is defined as the number of days of out-of-school suspension student *i* receives in year $t-1$. Note, this variable has been included in lagged form to limit the potential for reverse causation in the model. Although this specification assumes that each day of OSS has the same effect on student outcomes, we relax this assumption in some model specifications. 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). For example, a student may have been referred for two insubordination infractions and one fighting infraction in the same year. We control for disciplinary infraction history using these infraction counts in order to determine the impact of OSS in the past separately from the impact of having conducted a particular infraction in the same year as the suspension. For example, we estimate the impact of receiving OSS for fighting, relative to being written up for fighting, but receiving some other, less exclusionary consequence. This is important for isolating the effect of the OSS itself, not the disengagement from or misbehavior in school.

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 the 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, LEP status, 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 relatively simple model might not fully account for other unobservable student characteristics. Unobservable characteristics such as family backgrounds and community characteristics may relate to both the students' risk of OSS and their academic outcomes, so omitted variable bias remains 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 and allow for a certain type of unobserved heterogeneity.

The fixed effects model allows 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 over time. Even if there is "enough" variation within students, biases may remain if endogeneity is driven by time-varying shocks with persistent effects 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, as there may be time-persistent shocks to a student's life such as divorce or death in the family, we believe it is more plausible than that of no omitted variable bias 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 estimates of coefficients of other variables included in the model.[13] 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 are 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, by relaxing the assumption of strict exogeneity, is allowed to be correlated to our other regressors. With the inclusion of a_i, we exclude the vector of student characteristics, X_{it}, in Equation (1). Other variables in Equation (2) are the same as those included in Equation (1).

To estimate this model, one could transform Equation (2) using first differencing as:

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

[13] 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% of students in the state contribute to estimation of the impact of OSS.

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 these 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:

<u>*First Stage:*</u>

$$\Delta OSSdays_{it-1} = \hat{\pi}_0 + \hat{\pi}_1 OSSdays_{it-2} + \sum_{j=1}^{k} \hat{\pi}_{2j} infractioncount_{ijt-2} +$$

$$\hat{\pi}_3 \Delta gradelevel_{it} + \hat{\pi}_4 \Delta schoolyear_{it} + \hat{\pi}_5 y_{it-2} + \hat{\pi}_6 \Delta d_{it} + \eta_{it}^{OSS} \qquad (4)$$

$$\Delta y_{it-1} = \hat{\pi}_0 + \hat{\pi}_1 OSSdays_{it-2} + \sum_{j=1}^{k} \hat{\pi}_{2j} infractioncount_{ijt-2} + \hat{\pi}_3 \Delta gradelevel_{it} +$$

$$\hat{\pi}_4 \Delta schoolyear_{it} + \hat{\pi}_5 y_{it-2} + \hat{\pi}_6 \Delta d_{it} + \eta_{it}^{test\_scores} \qquad (5)$$

And k=12 equations for each infraction type:

$$\Delta infractioncount_{ijt-1} = \hat{\pi}_0 + \hat{\pi}_1 OSSdays_{it-2} + \sum_{j=1}^{k} \hat{\pi}_{2j} infractioncount_{ijt-2} +$$

$$\hat{\pi}_3 \Delta gradelevel_{it} + \hat{\pi}_4 \Delta schoolyear_{it} + \hat{\pi}_5 y_{it-2} + \hat{\pi}_6 \Delta d_{it} + \eta_{it}^{infrac}$$

for each j = 1, …, 12 (6)

<u>*Second Stage:*</u>

$$\Delta y_{it} = \beta_0 + \beta_1 \Delta \widehat{OSSdays}_{it-1} + \beta_2 \Delta \widehat{infractioncount}_{it-1} + \beta_3 \Delta gradelevel_{it} +$$

$$\beta_4 \Delta schoolyear_{it} + \beta_5 \widehat{\Delta y_{it-1}} + \beta_6 \Delta d_{it} + \Delta\varepsilon_{it} \qquad (7)$$

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 more reasonable than the assumptions accompanying pooled OLS or other descriptive methods.

We describe our analytic samples in Table 3. The analytic samples are similar, regardless of method reflective of the state population. The ELA samples are smaller due to fewer grades tested. While students in grades 3-8 are tested in both subjects, students typically take three math End of Course exams in roughly grades 9-11, but only one Literacy test in 11th grade.

Table 3

Descriptive Statistics for State and Analytic Samples

	Entire State	Math POLS Sample	Math Student FE (Anderson-Hsiao) Sample	ELA POLS Sample	ELA Student FE (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

Results

The math results are in Columns 1 and 2 of Table 4. 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 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,[14] 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 4, 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 causal impact, the effect of an additional day of exclusionary discipline on math test scores, among those who experience OSS at least once, if anything, is positive but very small.

The results of the ELA analysis are shown in the final two columns of Table 4. Based on the pooled OLS model, as displayed in Column 3, 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.

[14] For example, Arcia (2006) estimated that students with high suspension rates were roughly three grade levels behind their peers with no suspensions after one year and almost five grade levels behind after two years).

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
Prior Year (PY) OSS Days	-0.0060 **	0.0039 **	-0.0056 **	0.0095 **
	(0.0006)	(0.0013)	(0.0008)	(0.0019)
PY Infraction Counts By Category[a]	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[b]	Y		Y	
Lagged Z-Score	0.714 **	0.208 **	0.686 **	0.261 **
	(0.001)	(0.004)	(0.001)	(0.005)
Constant		0.330		0.403 **
		(13.50)		(0.015)
Observations	1,033,936	660,826	839,542	512,684
Number of Students	367,759	275,810	324,033	235,917

Note. Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level.
[a] PY 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.
[b] Student demographic controls include gender, FRL-status, special education status, LEP-status and a vector of race/ethnicity indicators (White, Black, Hispanic, Asian, and Other).
** p<0.01, * p<0.05.

Within-student variation over time

One diagnostic test of whether fixed effects are appropriate in these data is to investigate the amount of variation in the counts of OSS days between and within students over time. Student fixed effects models use only within-student variation to identify the impact of OSS. Table 5 indicates that there is almost as much variation within students as across students, so a fixed effects model allows us to control for unobserved heterogeneity, without sacrificing much in terms of identifying variation.

Table 5

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.

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 4 but drop students who never received OSS. The estimated relationships in Appendix A are similar (-0.004 s.d. in math and -0.003 s.d. in ELA) to those seen in columns 1 and 3 of Table 4 (approximately -0.006 s.d. in both math and ELA). Thus, 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 the 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 6 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 7, 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 4, but we focus on the student fixed effects models. The sign of the coefficients for the math model in Column 2 are generally consistent with the results in Table 4 (which included only a linear count of OSS days). However, not all are statistically significant. We find that 1-2 days of OSS, relative to none, leads to slight 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 6, the lack of significance on the 3-4 OSS days and 7-10 OSS days could be due to low statistical power. Still, the results at least suggest a lack of a clear linear relationship between OSS day and impacts on math test scores.

Table 6

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.

The ELA impacts in Column 4 of Table 7 are similar to the results in Table 4 (positive in magnitude), with all but the impact of 1-2 OSS days statistically significant at the 95% confidence level or higher. These results are not necessarily generalizable, however, as the impact of 1-2 days, for example, is estimated using the students with some years of 1-2 days, and some years of zero, while the impact of 7-10 days, for example, is estimated only using the students with similar variation (some years with zero days, and some years with 7-10 days). Still, these results are suggestive of whether there could be differential impacts at different thresholds.

Table 7

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	-0.0470	**	0.0190	*	-0.0505	**	0.0215	
	(0.0057)		(0.0088)		(0.0071)		(0.0111)	
3-4 OSS Days in PY	-0.0668	**	0.0126		-0.0551	**	0.0380	**
	(0.0058)		(0.0092)		(0.0075)		(0.0127)	
5-6 OSS Days in PY	-0.0521	**	0.0319	*	-0.0552	**	0.0461	*
	(0.0084)		(0.0137)		(0.0113)		(0.0199)	
7-10 OSS Days in PY	-0.0695	**	0.0228		-0.0775	**	0.0828	**
	(0.0100)		(0.0162)		(0.0137)		(0.0242)	
11+ OSS Days in PY	-0.0842	**	0.0852	**	-0.0696	**	0.1280	**
	(0.0128)		(0.0233)		(0.0169)		(0.0352)	
PY Infraction Counts By Category[a]	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[b]	Y				Y			
Lagged Z-Score	0.714	**	0.208	**	0.685	**	0.261	**
	(0.001)		(0.004)		(0.001)		(0.005)	
Constant			0.393				0.406	**
			(12.00)				(0.014)	
Observations	1,033,936		660,826		839,542		512,684	
Number of Students	367,759		275,810		324,033		235,917	

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

[a] PY 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.

[b] Student 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).

** p<0.01, * p<0.05.

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 observed test score was above (or below) average for their grade and school year and results for observations recorded in grades 2-5 and grades 6-10.

Table 8 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 8 for FRL students, non-White students, male students, regular education students, below average students, and students in grades 6-10. There are no math impacts on the remaining groups. None of the analyses in Table 8 indicate heterogeneous impacts in math. For example, even though there is a slight 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 on ELA scores are in Table 9. 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 slight positive ELA impacts on female students. 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 students simply have idiosyncratically low or idiosyncratically high scores the first time we observe them.

Table 8

Subgroup Impacts of OSS Days on Standardized Math Scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
Prior Year (PY) OSS Days	0.0040 **	0.0029	0.0051 **	-0.0006	0.0044 **	0.0027
	(0.0014)	(0.0032)	(0.0016)	(0.0023)	(0.0016)	(0.0022)
PY Infraction Counts By Category[a]	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 **	0.152 **	0.213 **	0.216 **	0.245 **	0.165 **
	(0.005)	(0.007)	(0.007)	(0.005)	(0.006)	(0.006)
Constant	-0.666	0.468	-0.441	0.436 **	0.390	0.378 **
	(22.24)	(3.445)	(10.50)	(0.017)	(8.370)	(0.020)
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	0.0050 **	0.0064 **	-0.0009	0.0052	0.0039 **
	(0.0046)	(0.0012)	(0.0016)	(0.0023)	(0.0048)	(0.0013)
PY Infraction Counts By Category	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 **	-0.026 **	0.283 **	0.009	0.181 **	0.214 **
	(0.011)	(0.004)	(0.006)	(0.006)	(0.009)	(0.005)
Constant	0.563 **	0.312 **	0.394 **	0.646	0.726 **	22.1 **
	(0.047)	(0.012)	(0.015)	(12.09)	(0.028)	(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

Note. Robust standard errors in parentheses. 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.
ƚ Subgroup effects are statistically different.
[a] PY 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.
** p<0.01, * p<0.05.

Table 9

Subgroup Impacts of OSS Days on Standardized ELA Scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
Prior Year (PY) OSS Days	0.0093 **	0.0069	0.0099 **	0.0021	0.0083 **	0.0145 **
	(0.0021)	(0.0049)	(0.0024)	(0.0033)	(0.0024)	(0.0037)
PY Infraction Counts By Category[a]	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 **	0.266 **	0.212 **	0.293 **	0.258 **	0.284 **
	(0.006)	(0.007)	(0.008)	(0.005)	(0.006)	(0.006)
Constant	0.520	0.399 **	-4.022	0.700	0.370	0.415 **
	(7.620)	(0.030)	(31.46)	(13.30)	(8.029)	(0.024)
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	0.0070 **	0.0134 **ł	-0.0055 ł	0.0085	0.0092 **
	(0.0067)	(0.0017)	(0.0025)	(0.0034)	(0.0050)	(0.0021)
PY Infraction Counts By Category	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 **	0.020 **	0.277 **	0.139 **	0.225 **	0.274 **
	(0.012)	(0.005)	(0.007)	(0.006)	(0.008)	(0.005)
Constant	22.860	0.308 **	0.394	0.400	0.738 **	0.312 **
	(34.220)	(0.013)	(17.280)	(3.343)	(0.029)	(0.017)
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

Note. Robust standard errors in parentheses. 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.

łSubgroup effects are statistically different.

[a] PY 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.

** *p*<0.01, * *p*<0.05.

We conducted 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. These results are not necessarily generalizable, as the 7-10 days, for example, is estimated only identifying off the variation within students who had zero days in some years and 7-10 in others. The results are suggestive of differential impacts at different thresholds. As in Table 8, regular education students' math scores appear to be slightly positively impacted, if anything by OSS, although these estimated effects are still quite small. Further, students whose first observed test score was below average generally have positive test score growth the year following OSS days, 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, except that there is only one coefficient (out of 60) in Table 10 that is statistically significant and negative (the impact of 3-4 days of OSS, relative to zero, for special education students). 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 11 shows the same subgroup analyses, but predicting ELA test scores. As in Table 9, there are generally slightly positive or null impacts of OSS on ELA test scores, with consistently positive (but very small) 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 driven by reversion to the mean as described previously. 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. No other estimated effects in Table 11 are negative and significant.

Table 10

Subgroup Impacts of OSS Days on Standardized Math Scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
1-2 OSS Days in PY	0.0197	0.0149	0.0327 **	-0.0039	0.0134	0.0369 *
	(0.0102)	(0.0184)	(0.0118)	(0.0133)	(0.0110)	(0.0152)
3-4 OSS Days in PY	0.0154	-0.0029	0.0209	-0.0079	-0.0081 ł	0.0538 **ł
	(0.0105)	(0.0202)	(0.0120)	(0.0144)	(0.0117)	(0.0157)
5-6 OSS Days in PY	0.0420 **	-0.0242	0.0527 **	-0.0192	0.0196	0.0602 **
	(0.0154)	(0.0319)	(0.0171)	(0.0231)	(0.0173)	(0.0232)
7-10 OSS Days in PY	0.0318	-0.0472	0.0227	0.0180	0.0073	0.0527
	(0.0180)	(0.0409)	(0.0202)	(0.0278)	(0.0204)	(0.0278)
11+ OSS Days in PY	0.0785 **	0.1530 *	0.1070 **	0.0089	0.1180 **	0.0209
	(0.0257)	(0.0635)	(0.0278)	(0.0451)	(0.0293)	(0.0404)
PY Infraction Counts By Category[a]	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged Math Z-score	0.242 **	0.152 **	0.213 **	0.216 **	0.245 **	0.164 **
	(0.005)	(0.007)	(0.007)	(0.005)	(0.006)	(0.006)
Constant	-4.560	0.361 **	1.187	0.438 **	0.420	0.379 **
	(10.11)	(0.026)	(11.36)	(0.017)	(4.554)	(0.020)
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	0.0355 **	0.0374 **	-0.0044	0.0358	0.0149
	(0.0324)	(0.0080)	(0.0113)	(0.0142)	(0.0253)	(0.0094)
3-4 OSS Days in PY	-0.0817 *ł	0.0514 **ł	0.0272 *	0.0029	0.0178	0.0142
	(0.0348)	(0.0083)	(0.0117)	(0.0151)	(0.0317)	(0.0096)
5-6 OSS Days in PY	-0.0014	0.0569 **	0.0542 **	-0.0000	0.1340 *	0.0249
	(0.0503)	(0.0124)	(0.0171)	(0.0236)	(0.0537)	(0.0141)
7-10 OSS Days in PY	-0.0179	0.0470 **	0.0444 *	-0.0040	-0.0366	0.0267
	(0.0579)	(0.0148)	(0.0199)	(0.0292)	(0.0604)	(0.0168)
11+ OSS Days in PY	-0.0795	0.0752 **	0.1240 **	-0.0059	0.0780	0.0886 **
	(0.0914)	(0.0208)	(0.0282)	(0.0458)	(0.0892)	(0.0241)
PY Infraction Counts By Category	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged Math Z-score	0.367 **	-0.026 **	0.283 **	0.009	0.181 **	0.214 **
	(0.011)	(0.004)	(0.006)	(0.006)	(0.009)	(0.005)
Constant	0.564 **	0.314 **	0.421	2.781	0.727 **	5.260 **
	(0.047)	(0.011)	(7.536)	(4.788)	(0.028)	(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

Note. Robust standard errors in parentheses. 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.

łSubgroup effects are statistically different.

[a] PY 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.

** $p<0.01$, * $p<0.05$.

Table 11

Subgroup Impacts of OSS Days on Standardized ELA Scores (Anderson-Hsiao)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
1-2 OSS Days in PY	0.0154	0.0585 *	0.0382 **	-0.0068	0.0233	0.0098
	(0.0129)	(0.0232)	(0.0146)	(0.0171)	(0.0137)	(0.0203)
3-4 OSS Days in PY	0.0436 **	0.0189	0.0579 **ɫ	-0.0168 ɫ	0.0265	0.0637 **
	(0.0146)	(0.0286)	(0.0161)	(0.0209)	(0.0159)	(0.0231)
5-6 OSS Days in PY	0.0487 *	0.0233	0.0666 **	-0.0505	0.0440	0.0490
	(0.0226)	(0.0476)	(0.0241)	(0.0365)	(0.0246)	(0.0369)
7-10 OSS Days in PY	0.0745 **	0.1460 *	0.0790 **	0.0575	0.0891 **	0.0733
	(0.0272)	(0.0618)	(0.0298)	(0.0425)	(0.0295)	(0.0468)
11+ OSS Days in PY	0.1380 **	-0.1140	0.1430 **	-0.0552	0.1040 *	0.2310 **
	(0.0391)	(0.0999)	(0.0416)	(0.0709)	(0.0427)	(0.0696)
PY Infraction Counts By Category[a]	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged ELA Z-score	0.252 **	0.266 **	0.212 **	0.301 **	0.258 **	0.284 **
	(0.006)	(0.007)	(0.008)	(0.005)	(0.006)	(0.006)
Constant	0.554	0.427 **	0.310 **	0.629 **	0.416	0.420 **
	(9.463)	(0.030)	(0.023)	(0.021)	(8.069)	(0.024)
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 *	0.0267 **	0.0451 **	-0.0098	0.0464	0.0134
	(0.0400)	(0.0098)	(0.0152)	(0.0164)	(0.0266)	(0.0122)
3-4 OSS Days in PY	0.0024	0.0564 **	0.0721 **ɫ	-0.0111 ɫ	0.0367	0.0369 **
	(0.0467)	(0.0113)	(0.0173)	(0.0194)	(0.0336)	(0.0138)
5-6 OSS Days in PY	0.1260	0.0436 *	0.0925 **ɫ	-0.0694 *ɫ	0.1400 *	0.0300
	(0.0719)	(0.0177)	(0.0261)	(0.0340)	(0.0567)	(0.0213)
7-10 OSS Days in PY	0.1360	0.0653 **	0.1440 **ɫ	-0.1220 **ɫ	-0.0069	0.0930 **
	(0.0841)	(0.0216)	(0.0311)	(0.0446)	(0.0639)	(0.0261)
11+ OSS Days in PY	-0.0950	0.0887 **	0.1700 **ɫ	-0.0454 ɫ	0.1250	0.1200 **
	(0.1360)	(0.0308)	(0.0454)	(0.0642)	(0.0953)	(0.0378)
PY Infraction Counts By Category	Y	Y	Y	Y	Y	Y
Grade, School, District, and Student FE	Y	Y	Y	Y	Y	Y
Lagged ELA Z-score	0.297 **	0.020 **	0.277 **	0.134 **	0.225 **	0.274 **
	(0.012)	(0.005)	(0.007)	(0.006)	(0.008)	(0.005)
Constant	6.732	0.311 **	0.470	0.206 **	0.738 **	0.316 **
	(35.90)	(0.013)	(8.481)	(0.023)	(0.029)	(0.017)
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

Note. Robust standard errors in parentheses. 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.

ɫSubgroup effects are statistically different.

[a] PY 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.

** $p<0.01$, * $p<0.05$.

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 with these students added back in, estimating the effect of the cumulative days of exclusionary discipline (including OSS, ALE, and expulsion).[15] 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 4 (0.0039 s.d. in math and 0.0095 in ELA), the overall results using all three forms of arguably exclusionary discipline are null impacts on both types of test scores (see Appendix Table B1).

In addition, while the subgroup impacts of OSS (using the linear specification of OSS days) in Tables 8 and 9 were always null or slightly 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.). The results for all other subgroups are still null or slightly positive (see Appendix Tables B2 and B3).

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, 10, and 11 but with all three types of exclusionary discipline, rather than just OSS). The results are in Appendix Tables B4 (for the math and ELA estimates for the overall sample), B5 (for the subgroup effects on math test

[15] In some cases, the number of days of suspension or expulsion was recorded as something higher than the typical number of school days in an academic year (180). For example, in some cases, the number of days was listed as 365, which seems to indicate the intention of expelling a student for a full year or 180 school days, not 365. In all cases in which the number of days of suspension, expulsion, or ALE was greater than 180, they were recoded to equal 180 days. If no days were listed, it was coded as zero days. In less than 0.07% of OSS infractions, there were not days of OSS reported, in 2.07% of expulsion cases, no days were reported, and in 10.76% of ALE Referrals no days were reported.

scores), and B6 (for the subgroup effects on ELA test scores). No estimated math test score impacts were negative. In Table B6, 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 of all types (OSS, expulsion, and ALE) are generally null or very small positive impacts, so we have confidence that the null to slightly positive impacts of OSS are not driven by our sample restrictions.

Discussion and Conclusions

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 helps us avoid the likely impact of contemporary shocks on OSS and test scores in the same year. A remaining concern for a causal interpretation of these results is that there may be time-varying shocks to students that affect discipline and student test scores over time. However, the remaining assumptions needed for causal interpretation in this case are certainly less strict and more realistic than those imposed by OLS, so these findings are still an important contribution to the field.

In general, we find that OSS days have a very small 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 slightly 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 be due to mere chance. In general, any positive impacts we do find are quite small, but still significant due to high analytic power, 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 some of the most extreme disciplinary offenders (the students expelled or referred to an ALE for disciplinary reasons during our study period). In addition, 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 the most highly disciplined students (those who were expelled or referred to an ALE for disciplinary reasons), the general finding is of null to small positive impacts, with very few (and inconsistently) statistically significant negative impacts.

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 or from the community 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). While holding back a student is certainly not an educational intervention to be used to improve test scores, perhaps we observe a slight test scores benefit in some students if they are suspended out of school, retained a grade, and then receive an extra year of math or ELA instruction in a course in which they were previously struggling. However, this may also come with higher risk of drop-out (Marchbanks et al., 2014).

Implications

These results are important given the trend toward reining in the use of OSS in schools. According to Steinberg and Lacoe (2017), 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 nonpunitive discipline strategies and/or limit suspension use (Steinberg & Lacoe, 2017). Based on our results, if policymakers continue to push for reductions in exclusionary discipline, they should not expect improvements in academic impacts to follow.

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. California middle schools and schools serving high proportions of poor or black students tended to have elevated suspension rates for Black students, and some educators have expressed concerns about declines in safety and learning because more trouble-makers remain in school (Loveless, 2017).

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. 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 suspensions in kindergarten through grade three for mid-level offenses such as "disruptive behavior," shoving, using racial slurs, or inappropriate physical contact. Eden argues that school climate measures based on survey responses stayed relatively constant during Bloomberg's reforms, but that school climate deteriorated following a different type of reform during de Blasio's time as mayor. Beginning under de Blasio, principals were required 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 reform, teachers reported less order and discipline, and students reported more violence, drug use, alcohol use, and gang activity, as well as 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 the issues highlighted in Loveless (2017) and Eden (2017) focus on systemic effects on schools, it is clear from the current study that even the expected impacts on the actual students suspended 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. In fact, Chapter 4 of this dissertation reports on a study utilizing eight years of data from Arkansas in which I find that a policy prohibiting the use of OSS as a consequence for truancy had low implementation fidelity, particularly in high-minority schools with high truancy and OSS rates, and that the policy may not have had any impact on school level outcomes such as math and ELA test scores, student attendance, and chronic absenteeism. (Anderson, 2017). Given the lack of a clear benefit from reducing exclusionary discipline, 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, 2017; Anyon et al., 2014; Losen et al., 2015; Losen & Skiba, 2010; Sartain et al., 2015; Skiba et al., 2014; Skiba et al., 2011; Skiba et al., 2002), and these documented disparities suggest that the use of exclusionary discipline should at least be evaluated and researched further. Perhaps, regardless of the lack of a negative impact on student test scores, the use of exclusionary discipline, if perceived as overly harsh or unfair, 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 can have a positive impact on perceptions of school safety and test scores (Horner et al., 2009).

In this study, we find a counterintuitive result that OSS does not harm student test scores, which suggests that educators and policy makers should think carefully about policies designed solely with the goal of reducing suspensions for the sake of increasing test scores. Therefore, 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. In addition, while large-scale policy changes may achieve the narrow goal of reducing suspensions, future research is needed to determine systemic impacts on all students within a school or school system.

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.

Anderson, K. P. (2017). Just the way you are: School-level outcomes of a state-level discipline policy amidst implementation failure. Chapter 4. *Implementation and Outcomes of School Discipline Policy: Three Studies from Arkansas* (Unpublished Dissertation).

Anderson, T. W., & 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. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. state. *Education Policy Analysis Archives, 25*(49). Retrieved from http://dx.doi.org/10.14507/epaa.25.2787

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

Angrist, J. D., & 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 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 Bulletin, 80*(579), 49.

Carter, L. (1984). The sustaining effects study of compensatory education. *Educational Researcher, 13*(7), 4-13.

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*(4), 741-785.

Cooley, W., & Leinhardt, G. (1980). The instructional dimensions study. *Educational Evaluation and Policy Analysis, 2*(1), 7-25.

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.* New York, NY: Council of State Governments Justice Center.

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*(1), 2-15.

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., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology, 46*(4), 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*(5), 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*(5), 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, 29*(1).

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*(4), 217 – 226.

Losen, D., Hodson, C., Keith, M., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles, CA: The Civil Rights Project, University of California – Los Angeles.

Losen, D., & Skiba, R. (2010). *Suspended education: Urban middle schools in crisis.* Los Angeles, CA: The Civil Rights Project, University of California – Los Angeles.

Loveless, T. (2017). *The 2017 Brown Center Report on American Education. How well are American Students Learning?* Washington, DC: Brookings Institution.

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. *Econometrics, 49*(6), 1417-1426.

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).

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

Raffaele-Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New Directions for Youth Development,* 99, 17-33.

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., & 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. Retrieved from http://www.indiana.edu/~atlantic/wp-content/uploads/2014/03/Academic-Cost-of-School-Discipline.pdf

Reynolds, A. J., & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology, 83*(1), 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*, *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.* Chicago, IL: 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*(1), 21-27.

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 dolor of discipline: Sources 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*(3), 295-315.

Skiba, R. J., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores.* Bloomington, IN: Center for Evaluation and Education Policy, Indiana University.

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 school. (2006). *Youth & Society, 38*(1), 29-57.

Steinberg, M. P., & Lacoe, J. (2017). What do we know about school discipline reform? *Education Next, 17*(1), 44-52. 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.

Wald, J., & Kurlaender, M. (2003). Connected in Seattle? An exploratory study of student perceptions of discipline and attachments to teachers. *New Directions for Youth Development, 2003*(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.

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*(4), 817–838.

Appendices

Appendix A– Comparing OLS Results Excluding Students who Never Received OSS

Appendix Table A1

OLS Models Comparing Full Sample to Sample Excluding Students Who Never Received OSS

	Dependent Variable: Math Z-Score		Dependent Variable: ELA Z-Score	
	(1)	(2)	(3)	(4)
	Original OLS	Robustness Check	Original OLS	Robustness Check
Prior Year (PY) OSS Days	-0.0060 **	-0.0039 ***	-0.0056 **	-0.0033 ***
	(0.0006)	(0.0007)	(0.0008)	(0.0009)
PY Infraction Counts By Category[a]	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				
Student Demographic Controls[b]	Y	Y	Y	Y
Lagged Z-Score	0.714 **	0.672 ***	0.686 **	0.681 ***
	(0.001)	(0.002)	(0.001)	(0.003)
Constant				
Observations	1,033,936	183,848	839,542	137,795
Number of Students	367,759	61,425	324,033	52,053

Note. Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level. Robustness check sample excludes students who never received OSS.
[a] PY 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.
[b] Student 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).
** $p<0.01$, * $p<0.05$.

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

Appendix Table B1

Overall Impacts Including Students Expelled and/or Referred to ALE (Comparable to Table 4)

	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 **	0.0010	-0.0031 **	0.0012
	(0.0004)	(0.0006)	(0.0005)	(0.0008)
PY Infraction Counts By Category[a]	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[b]	Y		Y	
Lagged Z-Score	0.715 **	0.208 **	0.687 **	0.260 **
	(0.001)	(0.004)	(0.001)	(0.005)
Constant		0.375 **		0.403 **
		(0.012)		(0.014)
Observations	1,042,876	666,665	846,583	517,037
Number of Students	370,744	278,171	326,672	237,906

Note. Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level. Exclusion days includes Out of School Suspension, expulsion, and referrals to ALE.
[a] PY 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.
[b] Student 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).
** p<0.01, * p<0.05.

Appendix Table B2

Subgroup Impacts in Math Including Students Expelled and/or Referred to ALE (Comparable to Table 8)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
Prior Year (PY) Exclusion Days	0.0015	-0.0012	0.0029 **ł	-0.0021 *ł	0.0011	0.0005
	(0.0007)	(0.0014)	(0.0009)	(0.0009)	(0.0008)	(0.0012)
PY Infraction Counts By Category[a]	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	0.242 **	0.152 **	0.212 **	0.217 **	0.245 **	0.164 **
	(0.005)	(0.007)	(0.007)	(0.005)	(0.006)	(0.006)
Constant	-0.718	0.360 **	0.297 **	0.435 **	0.375 **	0.383 **
	(13.04)	(0.026)	(0.019)	(0.016)	(0.016)	(0.020)
Observations	409,955	256,710	235,038	431,627	340,237	326,428
Number of Students	170,147	108,024	97,115	181,056	142,284	135,887

Panel B:	Special Education	Regular Education	Below Avg. Math Score	Above Avg. Math Score	Grades 2-5	Grades 6-10
Prior Year (PY) Exclusion Days	-0.0023	0.0012 *	0.0022 **ł	-0.0022 ł	-0.0019	0.0014 *
	(0.0023)	(0.0006)	(0.0008)	(0.0012)	(0.0017)	(0.0007)
PY Infraction Counts By Category	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	0.365 **	-0.027 **	0.283 **	0.009	0.183 **	0.214 **
	(0.011)	(0.004)	(0.006)	(0.006)	(0.009)	(0.005)
Constant	0.399	0.308 **	0.405 **	0.304 **	0.726 **	0.299 **
	(24.16)	(0.011)	(0.015)	(0.026)	(0.027)	(0.014)
Observations	73,284	593,381	338,404	328,261	130,841	535,646
Number of Students	34,327	249,027	138,715	139,456	129,481	241,596

Note: Robust standard errors in parentheses. 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.
łSubgroup effects are statistically different.
[a] PY 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.
** $p<0.01$, * $p<0.05$.

Appendix Table B3

Subgroup Impacts in ELA Including Students Expelled and/or Referred to ALE (Comparable to Table 9)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
Prior Year (PY) Exclusion Days	0.0016	-0.0006	0.0036 **ł	-0.0027 *ł	0.0013	0.0008
	(0.0010)	(0.0016)	(0.0012)	(0.0011)	(0.0010)	(0.0016)
PY Infraction Counts By Category[a]	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	0.250 **	0.264 **	0.210 **	0.293 **	0.257 **	0.283 **
	(0.006)	(0.007)	(0.008)	(0.005)	(0.006)	(0.006)
Constant	0.399 **	0.396 **	0.321 **	0.467 **	0.388 **	0.436 **
	(0.017)	(0.029)	(0.023)	(0.018)	(0.018)	(0.024)
Observations	314,794	202,243	179,271	337,766	262,765	254,272
Number of Students	145,167	92,739	82,712	155,194	121,201	116,705

Panel B:	Special Education	Regular Education	Below Avg. ELA Score	Above Avg. ELA Score	Grades 2-5	Grades 6-10
Prior Year (PY) Exclusion Days	-0.0031	0.0013	0.0021 *	-0.0008	-0.0004	0.0013
	(0.0026)	(0.0007)	(0.0011)	(0.0013)	(0.0018)	(0.0009)
PY Infraction Counts By Category	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	0.293 **	0.017 **	0.276 **	0.135 **	0.225 **	0.273 **
	(0.012)	(0.005)	(0.007)	(0.006)	(0.008)	(0.005)
Constant	0.646 **	0.306 **	1.466	0.307 **	0.738 **	0.314 **
	(0.055)	(0.013)	(13.93)	(0.027)	(0.029)	(0.017)
Observations	54,985	462,052	236,546	280,491	129,469	387,562
Number of Students	28,435	213,377	108,901	129,005	128,169	201,629

Note: Robust standard errors in parentheses. 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.

łSubgroup effects are statistically different.

[a] PY 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.

** p<0.01, * p<0.05.

Appendix Table B4

Overall Impacts Including Students Expelled and/or Referred to ALE (Comparable to Table 7)

	Dependent Variable: Math Z-Score				Dependent Variable: ELA Z-Score			
	(1)		(2)		(3)		(4)	
	Pooled OLS		Student FE with Anderson-		Pooled OLS		Student FE with Anderson-	
1-2 Exclusion Days in PY	-0.0553	**	0.0191	**	-0.0642	**	0.0106	
	(0.0034)		(0.0058)		(0.0044)		(0.0075)	
3-4 Exclusion Days in PY	-0.0762	**	0.0443	**	-0.0759	**	0.0286	**
	(0.0042)		(0.0072)		(0.0055)		(0.0098)	
5-6 Exclusion Days in PY	-0.0845	**	0.0365	**	-0.0921	**	0.0507	**
	(0.0062)		(0.0104)		(0.0083)		(0.0145)	
7-10 Exclusion Days in PY	-0.1150	**	0.0417	**	-0.1060	**	0.0328	
	(0.0069)		(0.0122)		(0.0096)		(0.0173)	
11+ Exclusion Days in PY	-0.1210	**	0.0714	**	-0.1270	**	0.0395	
	(0.0088)		(0.0164)		(0.0122)		(0.0235)	
PY Infraction Counts By Category[a]	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[b]	Y				Y			
Lagged Z-Score	0.714	**	0.207	**	0.686	**	0.260	**
	(0.001)		(0.004)		(0.001)		(0.005)	
Constant			0.375	**			0.405	**
			(0.012)				(0.014)	
Observations	1,042,876		666,665		846,583		517,037	
Number of Students	370,744		278,171		326,672		237,906	

Note. Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level. Reference group is 0 days of OSS in prior year.

[a] PY 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.

[b] Student 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).

** p<0.01, * p<0.05.

Appendix Table B5

Subgroup Impacts in Math Including Students Expelled and/or Referred to ALE (Comparable to Table 10)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
1-2 Exclusion Days in PY	0.0269 **	-0.0046	0.0363 **ł	-0.0005 ł	0.0161 *	0.0273 **
	(0.0070)	(0.0109)	(0.0088)	(0.0078)	(0.0074)	(0.0097)
3-4 Exclusion Days in PY	0.0495 **	0.0237	0.0505 **	0.0276 **	0.0326 **	0.0678 **
	(0.0083)	(0.0148)	(0.0099)	(0.0104)	(0.0091)	(0.0121)
5-6 Exclusion Days in PY	0.0385 **	0.0325	0.0500 **	0.0099	0.0271 *	0.0528 **
	(0.0119)	(0.0222)	(0.0139)	(0.0156)	(0.0130)	(0.0181)
7-10 Exclusion Days in PY	0.0456 **	0.0236	0.0505 **	0.0153	0.0274	0.0684 **
	(0.0137)	(0.0279)	(0.0158)	(0.0193)	(0.0152)	(0.0212)
11+ Exclusion Days in PY	0.0843 **	0.0040	0.1070 **ł	-0.0148 ł	0.0646 **	0.0777 **
	(0.0184)	(0.0390)	(0.0209)	(0.0266)	(0.0203)	(0.0290)
Lagged Math Z-score	0.242 **	0.152 **	0.212 **	0.217 **	0.245 **	0.162 **
	(0.005)	(0.007)	(0.007)	(0.005)	(0.006)	(0.006)
Constant	1.352	5.890	0.299 **	0.434 **	0.375 **	0.384 **
	(11.25)	(14.58)	(0.019)	(0.016)	(0.016)	(0.020)
Observations	409,955	256,710	235,038	431,627	340,237	326,428
Number of Students	170,147	108,024	97,115	181,056	142,284	135,887

Panel B:	Special Education	Regular Education	Below Avg. Math Score	Above Avg. Math Score	Grades 2-5	Grades 6-10
1-2 Exclusion Days in PY	0.0195	0.0425 **	0.0274 **	0.0171 *	0.018	0.0177 **
	(0.0230)	(0.0052)	(0.0078)	(0.0087)	(0.0167)	(0.0062)
3-4 Exclusion Days in PY	-0.0045 ł	0.0795 **ł	0.0624 **	0.0305 **	0.0194	0.0467 **
	(0.0279)	(0.0065)	(0.0093)	(0.0114)	(0.0236)	(0.0075)
5-6 Exclusion Days in PY	0.0159	0.0734 **	0.0547 **	0.0189	0.0544	0.0335 **
	(0.0389)	(0.0094)	(0.0131)	(0.0173)	(0.0356)	(0.0108)
7-10 Exclusion Days in PY	-0.0078	0.0853 **	0.0569 **	0.0354	-0.0296	0.0459 **
	(0.0452)	(0.0110)	(0.0151)	(0.0213)	(0.0418)	(0.0127)
11+ Exclusion Days in PY	-0.0010	0.0948 **	0.1170 **ł	-0.0359 ł	0.0088	0.0753 **
	(0.0612)	(0.0148)	(0.0202)	(0.0293)	(0.0533)	(0.0172)
Lagged Math Z-score	0.366 **	-0.028 **	0.282 **	0.009	0.182 **	0.213 **
	(0.011)	(0.004)	(0.006)	(0.006)	(0.009)	(0.005)
Constant	0.548 **	0.309 **	20.95	0.304 **	0.725 **	0.300 **
	(0.047)	(0.011)	(12.82)	(0.026)	(0.027)	(0.014)
Observations	73,284	593,381	338,404	328,261	130,841	535,646
Number of Students	34,327	249,027	138,715	139,456	129,481	241,596

Note. Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level. Reference group is 0 days of OSS in prior year. All models include PY Infraction Counts by Category and Grade, School, District, and Student Fixed Effects.
łSubgroup effects are statistically different.
[a] PY 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.
[b] Student 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).
** p<0.01, * p<0.05.

Appendix Table B6

Subgroup Impacts in ELA Including Students Expelled and/or Referred to ALE (Comparable to Table 11)

Panel A:	FRL	Non-FRL	Non-White	White	Male	Female
1-2 Exclusion Days in PY	0.0180 *	0.0020	0.0372 **ł	-0.0106 ł	0.0113	-0.0103
	(0.0090)	(0.0139)	(0.0112)	(0.0102)	(0.0094)	(0.0132)
3-4 Exclusion Days in PY	0.0397 **	0.0107	0.0500 **	-0.0022	0.0317 **	0.0014
	(0.0114)	(0.0204)	(0.0132)	(0.0146)	(0.0122)	(0.0178)
5-6 Exclusion Days in PY	0.0644 **	0.0177	0.0691 **	0.0204	0.0467 **	0.0406
	(0.0167)	(0.0311)	(0.0191)	(0.0221)	(0.0177)	(0.0275)
7-10 Exclusion Days in PY	0.0402 *	0.0354	0.0437 *	0.0133	0.0424 *	-0.0163
	(0.0197)	(0.0405)	(0.0221)	(0.0281)	(0.0211)	(0.0331)
11+ Exclusion Days in PY	0.0570 *	-0.0517	0.0613 *	-0.0343	0.0549	-0.0215
	(0.0266)	(0.0557)	(0.0300)	(0.0375)	(0.0282)	(0.0472)
Lagged ELA Z-score	0.250 **	0.264 **	0.211 **	0.292 **	0.257 **	0.283 **
	(0.006)	(0.007)	(0.008)	(0.005)	(0.006)	(0.006)
Constant	0.402 **	0.396 **	0.323 **	0.466 **	0.389 **	0.437 **
	(0.017)	(0.029)	(0.022)	(0.018)	(0.018)	(0.024)
Observations	314,794	202,243	179,271	337,766	262,765	254,272
Number of Students	145,167	92,739	82,712	155,194	121,201	116,705
Panel B:	Special Education	Regular Education	Below Avg. ELA Score	Above Avg. ELA Score	Grades 2-5	Grades 6-10
1-2 Exclusion Days in PY	0.0725 *	0.0277 **	0.0303 **	0.0000	-0.0017	0.0096
	(0.0285)	(0.0066)	(0.0108)	-0.0101	(0.0176)	(0.0083)
3-4 Exclusion Days in PY	0.0308	0.0546 **	0.0674 **ł	-0.0115 ł	-0.0061	0.0305 **
	(0.0360)	(0.0087)	(0.0135)	(0.0143)	(0.0248)	(0.0107)
5-6 Exclusion Days in PY	0.1030 *	0.0667 **	0.1050 **ł	-0.0423 ł	0.1500 **ł	0.0278 ł
	(0.0521)	(0.0129)	(0.0195)	(0.0225)	(0.0375)	(0.0157)
7-10 Exclusion Days in PY	0.1360 *	0.0387 *	0.0937 **ł	-0.0864 **ł	-0.0155	0.0329
	(0.0620)	(0.0154)	(0.0228)	(0.0290)	(0.0442)	(0.0189)
11+ Exclusion Days in PY	0.0471	0.0316	0.0936 **ł	-0.0835 *ł	0.0330	0.0281
	(0.0813)	(0.0210)	(0.0308)	(0.0396)	(0.0563)	(0.0258)
Lagged ELA Z-score	0.292 **	0.0169 **	0.275 **	0.135 **	0.224 **	0.273 **
	(0.019)	(0.005)	(0.007)	(0.006)	(0.008)	(0.005)
Constant	0.646 **	0.309 **	0.408	0.307 **	0.737 **	0.316 **
	(0.055)	(0.013)	(9.363)	(0.027)	(0.029)	(0.017)
Observations	54,985	462,052	236,546	280,491	129,469	387,562
Number of Students	28,435	213,377	108,901	129,005	128,169	201,629

Note. Robust standard errors in parentheses. Standard errors in OLS models are clustered at the student level. Reference group is 0 days of OSS in prior year. All models include PY Infraction Counts by Category and Grade, School, District, and Student Fixed Effects.
łSubgroup effects are statistically different.
[a] PY 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.
[b] Student 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).
** $p<0.01$, * $p<0.05$.

Chapter 4

Just the Way You Are:

School-Level Outcomes of a State-Level Discipline Policy amidst Implementation Failure

Kaitlin P. Anderson
University of Arkansas

Abstract

Exclusionary discipline such as out-of-school suspension (OSS) and expulsion is associated with lower student achievement, higher risk of drop-out or grade retention, and involvement in the juvenile justice system. Further, there is concern about higher rates of exposure to exclusionary discipline for students of color and special education students. In response, many schools, districts, and states are moving toward less exclusionary consequences. In 2013, the Arkansas state legislature passed a bill prohibiting the use of OSS as a consequence for truancy. Yet, even after three school years, there has not been a full reduction in the use of OSS for truancy. If the policy did not succeed in eliminating the use of OSS for truancy, what did happen?

In this paper, I use eight years of student- and infraction-level data for all K-12 students in Arkansas public schools to assess descriptively which school-level factors are associated with the use of OSS as a consequence for truancy in the year the law was passed, and which are associated with policy compliance. In addition, I utilize a comparative interrupted time series analysis to estimate whether school-level outcomes (test scores, attendance, chronic absenteeism, truancy, and student disciplinary outcomes) may have been affected by the policy.

I find schools that served more minority students, schools that had higher rates of truancy, and schools that had higher rates of OSS were less likely to comply with the policy, all else equal. Combined, these results suggest that the types of schools likely targeted by this policy are the same ones not fully complying with it. Combined with a lack of evidence that the policy was associated with improved student achievement, attendance, or chronic absenteeism, these findings suggest that the potential impact of state-level policies, which sound beneficial *prima facie*, may be limited if policy changes are not communicated well to schools, if there is not accountability to ensure compliance, and if there is not school capacity to handle discipline effectively.

Introduction and Literature Review

There is growing concern in the education community about chronic absenteeism, which the U.S. Department of Education (2016) has called a "hidden educational crisis." Absenteeism refers to any period of not attending school, while chronic absenteeism is defined as student absenteeism representing over 10% of the school year (Gottfried, 2015). Across the country, 10-15% of students are chronically absent in a typical school year (Balfanz & Byrnes, 2012), but the problem is even greater in urban areas such as New York City, where over 23 percent of elementary school students were chronically absent in 2009 (Nauer, Mader, Robinson, & Jacobs, 2014). The picture is even worse within particularly low-income parts of the city, with almost 40 percent of students chronically absent in some areas (Nauer et al., 2014).

Chronic absenteeism is related to truancy, which has been linked to negative outcomes such as gang activity, burglary, auto theft, and vandalism (Bell, Rosen, & Dynlacht, 1994; Dryfoos, 1990; Garry, 1996; Huizinga, Loeber, & Thornberry, 1995; Rohrman, 1993). Adults who were frequently truant earlier in life are more likely to have poor physical and mental health, lower paying jobs, increased reliance on welfare, children who exhibit problem behaviors, and an increased risk of incarceration (Bell et al., 1994; Dryfoos, 1990; Ingersoll & LeBoeuf, 1997; Rohrman, 1993).

Some researchers consider absenteeism and truancy to be distinct in that absenteeism is defined as periods of not attending school, while truancy is an unexcused and unlawful absence from school without parental knowledge or consent, with the student typically spending time away from home (Bell et al., 1994; Lee & Miltenberger, 1996). Other scholars have argued that the shift to labeling absenteeism as "truancy" occurs as students become older (e.g. high school students) and when absenteeism accompanies other risky behaviors (Gottfried, 2015). It is very likely that different schools and districts have varied definitions for what constitutes truancy.

Schools recognize truancy as a problem and use a variety of methods to respond to truancy including automatically failing students who miss a designated amount of school, disciplinary consequences such as in-school suspension (Kube & Ratigan, 1992) or out-of-school suspension (OSS) (Pell, 2000), and legal consequences for parents (Smink & Heilbrunn, 2005).[16] Regardless of whether OSS is an appropriate consequence for truancy, many in the educational community are troubled by the general use of OSS and other types of exclusionary discipline for several reasons.

OSS and expulsions are the two main types of exclusionary discipline, which remove the student from the traditional learning environment. In Arkansas, out-of-school suspensions longer than 10 days are considered expulsions (Arkansas Code § 6-18-507). Exclusionary discipline has been linked to lower academic achievement (Arcia, 2006; Beck & Muschkin, 2012; Cobb-Clark, Kassenboehmer, Le, McVicar, & Zhang, 2015; Raffaele-Mendez, 2003; Raffaele-Mendez, Knoff, & Ferron, 2002; Rausch & Skiba, 2005; Skiba & Rausch, 2004), school drop-out and grade retention (American Academy of Pediatrics, 2013; American Psychological Association, 2008; Balfanz, Byrnes, & Fox, 2014; Cobb-Clark et al., 2015; Fabelo et al., 2011; Gregory & Weinstein, 2008; Krezmien, Leone, & Achilles, 2006; Marchbanks, Blake, Smith, Seibert, & Carmichael, 2014; Raffaele-Mendez, 2003; Stearns & Glennie, 2006; Swanson, Erickson, & Ritter, 2017; Wald & Kurlaender, 2003), and involvement in the juvenile justice system (American Academy of Pediatrics, 2013; Fabelo et al., 2011; Nicholson-Crotty, Birchmeier, &

[16] For example, parents can be held civilly or criminally liable for their children's misbehavior (Geis & Binder, 1991; Siegel, 2002).

Valentine, 2009). There is particular concern about the disparate use of exclusionary discipline for students of color and special education students (Anderson & Ritter, 2017; 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; U.S. Departments of Education and Justice, 2014).

To date, very little of this research is causal. One study (Chapter 3 of this dissertation) estimates a causal impact, using rigorous dynamic panel data methods within a student fixed effects framework, and concludes there is no evidence of a negative impact of OSS on student test scores (Anderson, Ritter, & Zamarro, 2017). That study suggests a reduction in suspensions - if not accompanied by other reforms - may not improve student academic achievement.

Despite a lack of causal evidence that suspensions harm students, many states and school districts have tried to reduce the use of exclusionary discipline. As of May 2015, 22 states and the District of Columbia had revised laws to limit exclusionary discipline and implement less punitive strategies, and as of the 2015-16 school year, 23 of the 100 largest school districts had implemented similar reforms (Steinberg & Lacoe, 2017). These reforms include reducing the length of suspensions as in Chicago (Stevens et al., 2015), limiting suspension for certain minor misbehaviors as in California (Public Counsel, 2014), reducing suspensions for young students as in Seattle (Cornwell, 2015), or even completely eliminating the use of suspensions as in Miami-Dade (O'Connor, 2015). However, while OSS has officially been eliminated in Miami-Dade, some students are still reportedly sent home from school, and teachers have concerns that the reduction was attempted without sufficient staff buy-in and support (Gerety, 2016).

Choosing to exclude students from school as a consequence for truancy is certainly counterintuitive and it may well be that this practice is particularly harmful or ineffective. It is hard to justify using OSS as a response to truancy given that the consequence for not attending school further removes the student from school (U.S. Departments of Education and Justice, 2014). However, because the students who are truant or chronically absent often are the same ones with disciplinary issues (Baker, Sigmon, & Nugent, 2001; Balfanz et al. 2014; Huizinga, Loeber, Thornberry, & Cothern, 2000), the causal relationship between disciplinary policies and absenteeism is difficult to untangle. Truancy is found to predict exclusionary discipline, but the reverse is true as well. Excluding students from school for disciplinary reasons is related to lower attendance, higher risk of course failure, and a path of disengagement from school (Balfanz et al., 2014). Truancy is a major risk factor for multiple suspensions, expulsions, and school dropout (Baker et al., 2001; Huizinga et al., 2000), and students may begin to disengage from school through nonattendance at very early ages (Alexander, Entwisle, & Kabanni, 2001; Epstein & Sheldon, 2002). In addition, there is potential for confounding factors to influence both truancy and discipline referrals. However, no evidence identifies a clearly causal relationship between truancy and other disciplinary outcomes.

Despite a lack of causal evidence, there is support for the idea that exclusionary discipline should not be used for truancy (Smink & Heilbrunn, 2005; U.S. Departments of Education and Justice, 2014). The U.S. Departments of Education and Justice (2014) wrote: "policies that impose out-of-school suspensions or expulsions for truancy also raise concerns because a school would likely have difficulty demonstrating that excluding a student from attending school in response to the student's efforts to avoid school was necessary to meet an important educational goal" (p. 12). Truancy and dropout prevention groups recommend that schools never assign OSS as a punishment for truancy (Smink & Heilbrunn, 2005). However, OSS is still used as a consequence for truancy in some contexts. Arkansas and Ohio have recently adopted legislation to remove OSS as a legal consequence for truancy (Arkansas § 6-18-507; O'Donnell, 2016). Arkansas is the focus of this study.

In March 2013, the Arkansas state legislature passed a bill prohibiting OSS as a consequence for truancy (Arkansas Code § 6-18-507). This policy was a single sentence in a longer Act outlining, among other things, a requirement for the state to produce an annual report on disciplinary outcomes for various subgroups of students. There was not a full reduction in the use of OSS as a consequence for truancy, as the law intended, even by the 2015-16 school year. While 14% of truancy cases resulted in OSS during the 2012-13 school year, this figure only fell to 10% in 2013-14 and about 9% in 2014-15 and 2015-16. In February 2016, the Arkansas State Board of Education was notified that some schools were still using OSS as a consequence for truancy, but additional communication reminding schools of this change was not distributed until January 2017 (after the study period). Thus, the implementation failure may indicate a lack of communication about the law, an inability on the part of schools to respond fully, or that schools had full knowledge and ability to respond but chose not to. In Appendix A, I document the timeline of events relevant to this policy change in more detail.

This paper addresses the implementation fidelity and outcomes of this state-level policy-based solution prohibiting the use of OSS as a consequence for truancy. Specifically, I ask three key research questions:

1. *What school-level factors predict whether a school was using OSS as a consequence for truancy in the baseline year, 2012-13?*
2. *What school-level factors predict whether a school, among those using OSS as a consequence for truancy in 2012-13, complies with the policy in the following years?*
3. *How did school level outcomes such as average math and ELA test scores, school*

attendance, chronic absenteeism, and disciplinary outcomes change in "policy-affected" schools (i.e., those using OSS as a consequence for truancy in 2012-13)?

Little is known currently about how these types of state-level policies are implemented at a local level, or their impact. Thus, this work fills a significant gap in our knowledge base regarding whether these types of broad-based policy solutions are feasible or effective. In the next section, I describe the data used in this study. Then, I describe the analytic methods. Finally, I present the results and discuss what they imply for future policy design.

Data and Descriptive Statistics

This study uses eight years (2008-09 through 2015-16) of de-identified student demographic, achievement, attendance, and discipline referral data from all K-12 schools in Arkansas provided by the Arkansas Department of Education (ADE). Demographic data include race, gender, grade level, special education status, limited English proficiency (LEP) status, and free-and reduced-price lunch (FRL) status. Academic achievement data include standardized scores on state tests in ELA and mathematics for eight school years from 2008-09 to 2015-16. From 2008-09 to 2013-14, state tests in ELA and math were administered in grades 3-8, and End of Course (EOC) examinations were administered in Algebra I, Geometry, Algebra II, and 11th Grade Literacy as part of the Arkansas Comprehensive Testing, Assessment, and Accountability Program (ACTAAP). During the 2014-15 school year, and again in 2015-16, there was a change in standardized testing. In 2014-15, the state administered the Partnership for Assessment of Readiness for College and Careers (PARCC) tests aligned with the Common Core State Standards (CCSS). Under PARCC, literacy exams were administered in grades 3-10, grade level math exams were administered in grades 3-8, and end of course exams were administered for Algebra and Geometry. In 2015-16, Arkansas administered the ACT Aspire summative assessments in math and ELA in grades 3-10. Student achievement test scores are standardized within grade, year, test subject, and testing group (e.g. with accommodations or without) to account for differences in test administrations and scaling methods.

Student-level attendance data are reported quarterly. In a small minority of cases, the sum of the days present and days absent for a quarter was unreasonable. For example, about 0.01% of cases had more than 65 days in total, which would be roughly equal to the maximum number of days per quarter if there were no breaks for weekends or holidays.[17] In 2.3% of cases, there were more than 50 days, and in 3.4% of cases, there were more than 45 days, which would be a typical quarter-length given Arkansas's normal 180 school-day calendar. To remove the impact of outliers, I use the Winsorization technique introduced by Charles P. Winsor (1895-1951) to replace the top five percent of values with the value at the 95th percentile.[18] In these cases, an adjusted days absent was calculated so that the student-by-quarter-level percent of days absent remained constant pre- and post-Winsorization.

Discipline data are at the individual infraction level and specify, for each incident, the infraction type (out of 19 types) and the corresponding consequence (out of 13 types). I group

[17] These outliers were not clustered in particular schools, so I assume these outliers are not due to certain schools having a long or irregular school calendar. Rather, these outliers are likely due to some students being counted in more than one school (which is possible as some students switch schools during the school year, and school attendance counts may not accurately reflect the exact entry and exit dates for each student). Unfortunately, the attendance data files did not include school or district indicators, so it was impossible to determine whether duplicate entries for the same student and school year are due to multiple schools reporting for the same student.

[18] Winsorization is different from censoring, in that it replaces extreme values with less extreme values, rather than "trimming" or dropping the observations altogether (Locker, 2001). The choice between Winsorization and trimming is data- and context-specific, but Winsorization is used in this case, because the outliers are assumed to be real observations for real students that should not be dropped, but for some reason include too many days. Particularly in the case of non-symmetrical censoring (only censoring at one end of the distribution), statistical efficiency is better maintained with Winsorization relative to trimming (Dixon, 1960).

handgun, rifle, and shotgun infractions into one category for gun-related infractions. I collapse 13 consequence categories into 7: in school suspension (ISS), OSS, expulsion, referral to an alternative learning environment (ALE), corporal punishment, no action, and "other."[19] The most common consequences (for any infraction type) during the study period are ISS (38.3%), "other" consequences (24.9%), OSS (22.2%), and corporal punishment (13.3%). The "other" consequences do not fit into a state-designated reporting category and indicate a range of consequences such as morning, lunch, or after-school detentions, Saturday school,[20] or other. Expulsions (0.10%), referrals to ALE (0.31%), and no action (0.85%) are rare. The focus of this study is on the interaction between truancy infractions, representing 6.4% of all infractions, and OSS, which was used in 11.4% of truancy cases during the study period.

Table 1 shows descriptive statistics for school-level truancy rates (incidents per 100 students) for each year from 2008-09 to 2015-16. This time period includes four school years prior to the policy change, the year of the policy change (2012-13), and three outcome years. These statistics are presented for all schools, as well as for the set of schools with at least one truancy incident in the year observed. Between 60% and 66% of schools report zero truancy.[21] Generally, over time, fewer schools report zero truancy, indicating that either truancy is a growing concern, schools are increasing their reporting, or both. Some schools report high

[19] "Other" non-specified consequences were coded as a particular type of consequence at the school level, but when combined and reported by the state, they are grouped into an "other" category. This group was labeled as "other" when we receive the data, and is not a researcher-created group of consequences. There is a similar "other" category for infractions.

[20] It is unknown whether days of Saturday school are included in reported attendance data.

[21] About 80% of the schools reporting no truancy are elementary schools, which typically represent less than 50% of all schools. Schools reporting no truancy are proportionally distributed across the state's five regions. Schools with no truancy are slightly poorer on average (the average school with no truancy is 66-67% FRL instead of 64-65% for the average school in the state). Finally, schools with no reported truancy serve a similar share of black students (20% in the average school with no truancy, compared to 21% for the average school in the state.)

truancy rates, with a maximum of 179.7 infractions per 100 students in 2011-12. It is possible that high truancy rates in some schools in 2011-12 were a factor leading to the state pursuing this policy change, although the average school's truancy rate in this year (1.6 truancy infractions per 100 students) was not particularly high relative to other years. On average, schools report 1.4 to 2.4 incidents per 100 students per year, but among the schools that report at least some truancy, the typical rate is about 4.1 to 6.2 truancy incidents per 100 students.[22]

Table 1

School-Level Truancy Rates (Infractions per 100 Students)

	All Schools							Schools with ≥ 1 Truancy Incident				
School Year	*n* of Schools	Num. of Schools with Zero Truancy	% of Schools with Zero Truancy	*Mean*	*SD*	*Min*	*Max*	*n* of Schools	*Mean*	*SD*	*Min*	*Max*
2008-09	1,090	705	65%	1.58	4.54	0.00	57.82	385	4.47	6.74	0.09	57.82
2009-10	1,085	719	66%	1.81	6.12	0.00	98.10	366	5.38	9.60	0.07	98.10
2010-11	1,073	703	66%	1.76	5.67	0.00	91.93	370	5.10	8.74	0.09	91.93
2011-12	1,071	700	65%	1.61	6.89	0.00	179.74	371	4.65	11.10	0.07	179.74
2012-13	1,063	689	65%	1.44	3.95	0.00	43.69	374	4.08	0.06	0.08	43.69
2013-14	1,065	666	63%	1.81	5.68	0.00	81.09	399	4.83	0.08	0.08	81.09
2014-15	1,055	659	62%	2.32	7.48	0.00	136.71	396	6.19	0.11	0.12	136.71
2015-16	1,044	629	60%	2.36	7.78	0.00	111.84	415	5.93	0.11	0.11	111.84

The trends in truancy and the consequences used in response are in Table 2. Following the passage of Act 1329, there was a small dip in the use of OSS as a consequence for truancy. In the school year in which Act 1329 was passed (2012-13), 13.8% of truancy cases resulted in OSS, and by 2015-16 this figure had decreased to 8.7%. At the same time, there was an increase in the use of "other" consequences which are specified at the local level, but then roll up to an "other" category at the state level. Combined, these two trends suggest that some schools are

[22] These rates include multiple incidents per student, if applicable. Thus, given that there are some students with multiple truancy incidents per year, the percent of students written up for truancy in an average school is less than 4.1 percent per year.

shifting away from OSS towards these "other" types of consequences. There is also evidence of a shift away from using ISS, and this shift accounts for more of the increase in the "other" consequences than the shift away from OSS. For example, while between 2012-13 and 2013-14, the "other" consequences share increased by 17.1 percentage points, the ISS share dropped 13.3 percentage points, and the OSS share only dropped 3.9 percentage points.

Table 2

Consequences Administered for Truancy, All Arkansas Schools

	Pre-Policy					Policy Passed	Post-Policy			
	2008-09	2009-10	2010-11	2011-12	4-Year Average	2012-13	2013-14	2014-15	2015-16	3-Year Average
Pct. Of Truancy Cases Resulting in:										
ISS	71.4%	67.3%	68.5%	67.1%	68.5%	74.9%	61.5%	56.3%	51.9%	56.3%
Other Action	8.3%	16.7%	14.5%	17.4%	14.4%	9.0%	26.1%	31.7%	36.5%	31.7%
OSS	14.9%	12.1%	12.8%	12.6%	13.0%	13.8%	10.0%	9.1%	8.7%	9.2%
Corporal Punishment	4.9%	2.7%	2.8%	2.7%	3.3%	2.1%	1.6%	1.9%	1.8%	1.8%
No Action	0.3%	1.0%	1.3%	0.1%	0.7%	0.1%	0.7%	0.9%	1.0%	0.9%
ALE	0.1%	0.2%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Expulsion	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
Total Truancy Cases	9,968	11,834	11,734	10,464	44,000	9,407	12,914	14,987	15,598	43,499
Number of Schools	385	366	370	371		374	399	396	415	

Note. Number of schools indicates the number of schools that reported at least one truancy incident. Pre-policy 4-year average represents the weighted average from 2008-09 to 2011-12, or in other words, the percent of all truancy cases over those four years that resulted in each consequence type. Post-policy 3-year average represents the weighted average from 2013-14 to 2015-16, or in other words, the percent of all truancy cases over those three years that resulted in each consequence type.

Understanding what these "other" consequences represent is important for interpreting whether these shifts represent a meaningful change for students. Very limited qualitative evidence indicates these are primarily non-exclusionary consequences.[23]

[23] I contacted an assistant superintendent/former principal, a principal, and a former teacher in three of the school districts with the largest number of instances of "other" consequences administered for truancy in the past two years (2014-15 and 2015-16). The superintendent/former principal stated that the response to truancy typically is ISS, but the

The trends in Table 2 and the apparent replacement of ISS and OSS as consequences for truancy with "other" consequences represent the trends for the entire state, including schools that never used OSS for truancy. To look more closely at the schools likely to be affected by the policy change, I assess the trend in disciplinary practices in the subset of 192 schools that were using OSS as a consequence for truancy in the baseline year, 2012-13 (Table 3). For each year from 2008-09 to 2015-16, I report the relative frequencies of each consequence type used for truancy. In these schools, 18.3% of truancy cases in 2012-13 resulted in OSS. Following the policy change, this rate decreased gradually to 10.7% in 2015-16. At the same time, there was an increase in "other" consequences from 5.7% in the baseline year to a high of 34.8% by 2015-16. There was also a continuous decline in the use of ISS for truancy from 74.0% in 2012-13 to 51.6% by 2015-16. Thus, across all schools (Table 2), as well as among the schools that used OSS as a consequence for truancy in 2012-13 (Table 3), schools were shifting away from using ISS for truancy. This decline in the use of ISS as a consequence for truancy suggests there may be trends in discipline policy other than the passage of Act 1329 that are affecting how schools handle truancy cases, since this Act did not mention the use of ISS for truancy. In the schools that did not use OSS as a consequence for truancy in 2012-13, there was a similar decline in the use of ISS for truancy from 77.5% of truancy cases in 2012-13 to 52.3% of cases in 2015-16. Thus, there was a general trend away from ISS and towards "other" consequences in both types of schools (those that were using OSS for truancy in 2012-13 and those that were not).

"other" consequences may include loss of course credit or credit recovery (making up missed instructional time). For example, seniors may be required to come in extra days after they would usually be required to be at school. The high school principal indicated that his school primarily uses after school suspension and Saturday school as a consequence for truancy, practices which do not exclude the truant student from the regular learning environment. A former teacher from another high school indicated that some of the "other" consequences are likely a mix of morning detentions and students being "sent home" for part of the day.

Histograms of school-level use of OSS as a consequence for truancy during the baseline year (2012-13) and the three outcome years are presented in Figure 1. The corresponding descriptive statistics for these four years, as well as the prior four years, are in Table 4. In each year, about half of schools with at least one truancy incident did not use OSS as a consequence for truancy. In 2012-13, 12% of schools with at least one truancy incident used OSS in 100% of truancy cases. This figure was lower previously, and decreased slightly in later years, but was still as high as 9% in 2015-16. This group of schools using OSS exclusively as a consequence for truancy is evident in the far right of the bimodal distributions in Figure 1.

Table 3

Consequences Administered for Truancy, Arkansas Schools using OSS for Truancy in 2012-13

	Pre-Policy					Policy Passed	Post-Policy			
	2008-09	2009-10	2010-11	2011-12	4-Year Average	2012-13	2013-14	2014-15	2015-16	3-Year Average
Pct. Of Truancy Cases Resulting in:										
ISS	68.1%	66.6%	67.5%	67.8%	67.4%	74.0%	58.1%	54.5%	51.6%	54.6%
Other Action	7.3%	14.7%	12.7%	13.6%	12.3%	5.7%	27.4%	31.6%	34.8%	31.4%
OSS	20.6%	15.8%	16.9%	16.4%	17.3%	18.3%	12.0%	11.1%	10.7%	11.2%
Corporal Punishment	3.6%	1.8%	1.7%	2.0%	2.2%	1.9%	1.4%	1.4%	1.4%	1.4%
No Action	0.4%	0.7%	1.1%	0.1%	0.6%	0.0%	1.0%	1.2%	1.5%	1.2%
ALE	0.1%	0.2%	0.1%	0.0%	0.1%	0.1%	0.0%	0.1%	0.1%	0.1%
Expulsion	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
Total Truancy Cases	5,976	7,532	7,658	7,110	28,276	7,134	8,843	9,696	9,814	28,353
Number of Schools	141	135	147	160		192	164	155	157	

Note. Number of schools indicates the number of schools that reported at least one truancy incident. Pre-policy 4-year average represents the weighted average from 2008-09 to 2011-12, or in other words, the percent of all truancy cases over those four years that resulted in each consequence type. Post-policy 3-year average represents the weighted average from 2013-14 to 2015-16, or in other words, the percent of all truancy cases over those three years that resulted in each consequence type.

This study focuses on school-level analyses,[24] so infraction- and student-level data are aggregated to the school-by-year level. In the next section, I outline the analytic methods used.

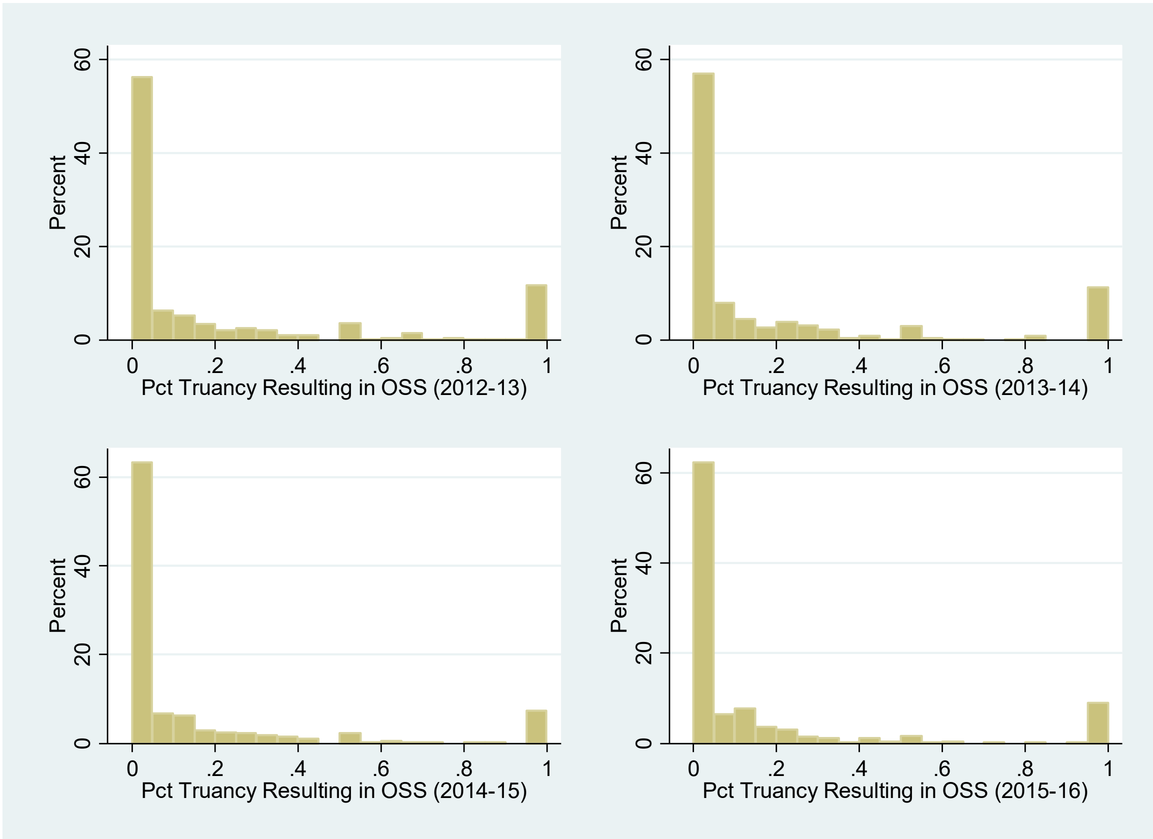

Figure 1. Distribution of School-level Use of OSS as a Consequence for Truancy (Schools with at Least One Truancy Incident)

[24] I focus on school-level analyses here, because a) this was a broad-based state-level policy that theoretically school leaders chose to either comply with or not, and thus, the level of intervention is not the student level, and b) the lack of implementation failure suggested a need for understanding what happens at a school-level. In the future, estimating student-level impacts on students who were truant or on the peers of truant students would both be topics worthy of study.

Table 4

Percentage of Truancy Infractions Resulting in OSS

				Schools Using OSS for Truancy:			
				In 0% of Incidents		In 100% of Incidents	
School Year	*n* of Schools	*Mean*	*SD*	Num. of Schools	% of Schools	Num. of Schools	% of Schools
2008-09	385	0.19	0.32	196	51%	33	9%
2009-10	366	0.17	0.30	179	49%	28	8%
2010-11	370	0.15	0.27	197	53%	21	6%
2011-12	371	0.19	0.32	197	53%	36	10%
2012-13	374	0.21	0.34	182	49%	44	12%
2013-14	399	0.19	0.33	203	51%	45	11%
2014-15	396	0.14	0.28	218	55%	29	7%
2015-16	415	0.15	0.30	220	53%	37	9%

Note. Sample only includes schools with at least one truancy incident in the year reported.

Analytic Methods

Research Question 1: What school-level factors predict whether a school was using OSS as a consequence for truancy in the baseline year, 2012-13?

For this question, I exclude schools with zero truancy incidents in 2012-13, (which, as indicated in Table 1, represent about 65% of schools). While this is a large proportion of schools, it is not reasonable to compare the rate of OSS as a response for truancy among schools with and without truancy incidents. Using school-by-year level data, I predict, using a probit model, whether a school used OSS as a consequence for truancy at least once during 2012-13, denoted by $Y_{S2012-13}$ and defined as:

$$Y_{S2012-13} = \begin{cases} 1\ if\ Y^*_{S2012-13} > 0 \\ 0\ if\ Y^*_{S2012-13} \leq 0 \end{cases}$$

$$Y^*_{S2012-13} = \beta_0 + \beta_1 TruancyRate_{S2011-12} + \beta_2 OSSRate_{S2011-12} + \beta_3 ISSRate_{S2011-12} +$$
$$\beta_4 ExpulsionRate_{S2011-12} + \beta_5 CorporalRate_{S2011-12} + \beta_6 ALERate_{S2011-12} +$$

$\beta_7 OtherConsRate_{S2011-12} + \ \beta_8 NoActionRate_{S2011-12} + \ \boldsymbol{X_{S2011-12}\beta} +$

$region_S + \varepsilon_{S2012-13}$ (1)

$Y^*_{S2012-13}$ is a latent variable representing the propensity for school s to have used OSS as a consequence for truancy in 2012-13.

Many of the included explanatory variables such as the truancy rate are likely endogenous. To limit endogeneity concerns, these explanatory variables are lagged. However, despite using lagged variables, there could still be constant unobservable school characteristics that are correlated with these variables, so the results should be interpreted as descriptive, not causal. For example, although I control for school-level demographic characteristics to capture some information about the family and community background of students in each school, truancy rates might be correlated with unobservable characteristics about families and communities that also are related to a preference to use OSS for truancy.

The explanatory variables include $TruancyRate_{S2011-12}$, the number of truancy infractions per 100 students in 2011-12, because it may be that schools struggling with high rates of truancy have a different propensity to use OSS as a consequence for truancy. The explanatory variables also include $OSSRate_{S2011-12}$, $ISSRate_{S2011-12}$, $ExpulsionRate_{S2011-12}$, $CorporalRate_{S2011-12}$, $ALERate_{S2011-12}$, $OtherConsRate_{S2011-12}$, and $NoActionRate_{S2011-12}$, which are variables representing the number of infractions per 100 students resulting in each of the seven consequence types in school s in the previous school year (2011-12). These variables are included to capture the school's typical rates of disciplinary consequences, which is important given that discipline practices are clearly different in the types of schools that do or do not use OSS as a consequence for truancy. In addition, it is probable that schools with high rates of exclusionary discipline may be more likely to use OSS as a consequence for truancy or that the types of schools that generally rely on non-exclusionary consequences such as ISS, would be less likely to use OSS as a consequence for truancy. In total, these seven consequence types also represent the total number of infractions in the school.

$\boldsymbol{X_{S2011-12}}$ is a vector of lagged school level characteristics including the log of school enrollment and the percent of students who are FRL-eligible, receiving special education services, Black, Hispanic,[25] or of another non-white race. This vector also includes indicators for open-enrollment and district conversion charter schools (with traditional public schools as the reference group), as well as indicators for high schools and middle schools (with elementary as the reference group). $\boldsymbol{X_{S2011-12}}$ also includes the percent of students disciplined and school-level academic indicators including school-average math and ELA test scores in standard deviation units. To capture remaining time-invariant unobservable regional differences, I include a set of region fixed effects for the five regions in the state, $region_S$, with the largest region, Northwest Arkansas, as the omitted group. Robust standard errors are clustered at the district level.[26]

In the primary model, a school is included if it had at least one truancy case in 2012-13, which could be an outlier or a data reporting issue. As a robustness check, I estimate models among the schools with at least five truancy cases.[27]

[25] The percent of students who are identified as limited English proficient in each school was not included as an explanatory variable due to its high correlation with Hispanic (r = 0.946).
[26] As of 2015-16, there were 259 districts in Arkansas, 192 represented in the models restricted to schools with at least one truancy incident, and 140 represented in the models restricted to schools with at least five truancy incidents.
[27] Using a cut-off of five truancy cases was admittedly arbitrary. Future iterations of this work will test the sensitivity of this choice using a variety of cut-offs, both in terms of the number of truancy cases and the percent of students reported truant.

Research Question 2: What school-level factors predict whether a school, among those using OSS as a consequence for truancy in 2012-13, complies with the policy in the following years?

I address this research question with two analyses: 1) separate probit models for each outcome year (2013-14, 2014-15, and 2015-16) and 2) an ordered probit model predicting whether a school complied in zero, one, two, or three of the outcome years.

Year-by-year Analyses

To examine what school-level factors predict compliance in each of the outcome years, I estimate three separate models and restrict the sample to schools that used OSS as a consequence for truancy at least once in the baseline year (2012-13) and schools with at least one truancy incident in the outcome year. The goal is to see which types of schools are likely to comply in a given year, regardless of whether the school complied in other years. Thus, a school that complied in 2013-14 would remain in the 2014-15 model. Some schools switch back to non-compliance after complying, so given my interest in modelling which types of schools comply in a particular year, I conduct a series of probit models rather than a hazard or duration model.

I estimate one model for each outcome year (2013-14, 2014-15, and 2015-16). Whether or not a school complies with the policy in year *t* (C_{st}) is defined as:

$$C_{st} = \begin{cases} 1 \; if \; C^*_{st} > 0 \\ 0 \; if \; C^*_{st} \leq 0 \end{cases}$$

$$\begin{aligned} C^*_{st} = {} & \beta_0 + \beta_1 TruancyRate_{st-1} + \beta_2 OSSRate_{st-1} + \beta_3 ISSRate_{st-1} + \\ & \beta_4 ExpulsionRate_{st-1} + \beta_5 CorporalRate_{st-1} + \beta_6 ALERate_{st-1} + \\ & \beta_7 OtherConsRate_{st-1} + \beta_8 NoActionRate_{st-1} + \boldsymbol{X_{st-1}\beta} + region_s + \varepsilon_{st} \qquad (2) \end{aligned}$$

Where C^*_{st} is a latent variable representing the propensity to comply with the policy in year *t*. Compliance in year *t*, C_{st}, is equal to one if school *s* had truancy incidents in year *t* and reported using OSS in zero of these cases. C_{st} equals zero if school s had truancy incidents in year t and reported using OSS in at least one of these cases. Thus, C_{st} can be interpreted as an indicator for full compliance in year *t*. The explanatory variables are similar to those in Equation 1 and are lagged one year. Robust standard errors are clustered at the district level.

As in Equation 1, many variables are potentially endogenous, even though they are lagged, so the results cannot be interpreted as causal. As a robustness check, I also estimate models where I limit the sample to schools that had at least five truancy cases.

Ordered Probit Models

In addition to the three separate year models, I estimate an ordered probit model predicting whether a school complied with the policy in zero, one, two, or three of the outcome years. For this model, I only include the schools that used OSS for truancy at least once in the baseline year, 2012-13, and schools that reported at least one truancy incident in all three outcome years. Of the 132 schools included in this model, 65 schools (49%) complied in zero years, 32 (24%) complied in one year, 21 (16%) complied in two years, and 14 schools (11%) complied in all three outcome years. These schools represent a relatively small portion of all schools in the state. For example, in 2012-13 there were 1,063 schools in the state, so the 132 schools represented in Table 9 represent only 12.4% of all schools.

Ordered probit models are used to estimate a series of outcomes with a natural order. The outcomes are assumed to arise sequentially as a latent variable, Y_s^*, crosses progressively higher thresholds, τ^j. It is assumed that the thresholds are the same for all schools in the sample. In this case, there are four possible outcomes (0, 1, 2, or 3 years of compliance), where the observed outcome, Y_s, is equal to some value, j, if and only if, the latent variable Y_s^* passes threshold τ^{j-1} but does not pass the next threshold, τ^j. Specifically:

$Y_s = j \iff \tau^{j-1} < Y_s^* \leq \tau^j$; j = 0, 1, 2, 3

Where $\tau^{-1} = -\infty, \tau^3 = \infty$ (3)

The latent variable Y_s^* is modeled as:

$$Y_s^* = \beta_0 + \beta_1 TruancyRate_{S2012-13} + \beta_2 OSSRate_{S2012-13} + \beta_3 ISSRate_{S2012-13} + \beta_4 ExpulsionRate_{S2012-13} + \beta_5 CorporalRate_{S2012-13} + \beta_6 ALERate_{S2012-13} + \beta_7 OtherConsRate_{S2012-13} + \beta_8 NoActionRate_{S2012-13} + \boldsymbol{X_{S2012-13}\beta} + region_s + \varepsilon_s \quad (4)$$

Research Question 3: How did school-level outcomes such as average math and ELA test scores, school attendance, chronic absenteeism, and disciplinary outcomes change in "policy-affected" schools (i.e., those using OSS for truancy in 2012-13)?

To estimate whether the policy was related to changes in treatment schools relative to comparison schools, I rely on a comparative interrupted time series (CITS) analysis.[28] This CITS compares changes between schools that would theoretically be affected by the policy because they were using OSS as a consequence for truancy in the baseline year and those that would theoretically not be affected because they were not using OSS as a consequence for truancy in the baseline year. CITS, an interrupted time series with a non-equivalent comparison group, has been used to estimate the impacts of school accountability policies (Dee & Jacob, 2011; Wong, Cook, & Steiner, 2011) as well as programs such as Reading First (Somers, Zhu, Jacob, & Bloom, 2013) and Jobs-Plus (Bloom & Riccio, 2005). Despite its relative rigor, CITS is used less

[28] I estimate the policy-related change in outcomes at a school-level, because a) this was a broad-based state-level policy that theoretically school leaders chose to either comply with or not, and thus, the level of intervention is not the student level, and b) the lack of implementation failure suggested a need for understanding what happens at a school-level. In the future, estimating student-level impacts on students who were truant or on the peers of truant students would both be topics worthy of study.

frequently in program or policy evaluation than difference-in-differences (DD), because it requires at least four time points before the intervention or policy change (Somers et al., 2013).

The goal of this approach is to compare deviations in trends between a set of "treatment" schools affected by the policy and a set of theoretically unaffected "comparison" schools. The CITS approach is similar in concept to a DD design, but the CITS assesses whether the treatment group deviates more than the comparison group from its *baseline trend*, whereas DD assesses whether the treatment group deviates more than the comparison group from its *baseline mean* (Somers et al., 2013). According to Somers et al. (2013), CITS is generally more rigorous than DD, because it controls for differences in the baseline mean and trends between the treatment and comparison groups, whereas DD assumes the baseline trends in the treatment groups and comparison groups to be similar. In other words, CITS accounts not only for differences in the levels between treatment and comparison groups, but also for differences in their natural growth rates. This is important because the types of schools that were or were not using OSS for truancy prior to the policy change likely had other unobservable differences affecting trends in disciplinary outcomes, test scores, or attendance in different ways. In addition, CITS was found to be better able to estimate long-term impacts than DD (Somers et al., 2013).

However, causal inference from CITS relies on the assumption that deviations from prior trends in the comparison schools provide a valid counterfactual for what would have happened in the treatment schools in the absence of the policy change. This is a strong assumption, because schools that did use OSS as a consequence for truancy in 2012-13 may be very different from the schools that did not and perhaps in ways that cannot be accounted for with observable characteristics. While it is possible to estimate a model comparing treatment schools and propensity-score matched comparison schools, even this more statistically complicated procedure cannot ensure that I have accounted for unobservable characteristics. Given the relatively limited set of covariates typically available in administrative data (e.g. FRL, special education, race/ethnicity, school size, school type, school grade span, disciplinary outcomes, etc.), this is a key limitation for causal inference from CITS in this case.

I define treatment schools as those that used OSS as a consequence for truancy at least once in 2012-13 and comparison schools as those that did not, and therefore had no changes to make with respect to this one particular policy. I exclude schools that reported no truancy at all in 2012-13 as it would not be reasonable to compare treatment schools to these schools.

The following regression illustrates the CITS design following Dee and Jacob (2011):

$$Y_{st} = \beta_0 + \beta_1 YEAR_t + \beta_2 POLICY_t + \beta_3 YR\_SINCE\_POLICY_t + \beta_4 (T_s \times YEAR_t) + \beta_5 (T_s \times POLICY_t) + \beta_6 (T_s \times YR\_SINCE\_POLICY_t) + \boldsymbol{X_{st}\beta} + \mu_s + region_{st} + \varepsilon_{st} \quad (5)$$

Where Y_{st} is a school-by-year level outcome such as average test scores in math or ELA, school overall non-attendance (student days absent as a percent of total instructional days),[29] the school chronic absenteeism rate, defined as the percent of students missing at least 10% of instructional days in a particular school year, or disciplinary outcomes hypothesized to be affected by the policy. $YEAR_t$ is a time-trend variable starting with 1 for the first year in the analytic dataset, 2008-09, $POLICY_t$ is an indicator variable equal to 1 for any time period after the policy change (i.e. in 2013-14 and later) and 0 for each earlier time period, and $YR\_SINCE\_POLICY_t$ is a time trend variable defined such that it equals zero in all periods prior to and including the year of the policy-change (through 2012-13), one in 2013-14, two in 2014-15, and three in 2015-16.

[29] Total instructional days is the sum of all student's days absent and days present, which estimates the total number of days the student theoretically should have been present, even if it was not a full school year, as some students may enter and exit schools within the year. For more details on these variables, see the Data and Descriptive Statistics section of this chapter.

$\boldsymbol{X_{st}}$ is a vector of school-level characteristics varying within schools over time such as the log of enrollment, the percentage of students that are FRL-eligible, Black, Hispanic, another non-white race, or eligible for special education services, etc.[30] In some models, I include only a few of the key variables identified as predictors of using truancy for OSS in the baseline year (based on results from Research Question 1), as these variables predict selection into treatment. These variables include the log of school enrollment, school percent Black, region indicators, and truancy frequency (incidents per 100 students). In other models, I also include seven controls for the frequency of the seven consequence types (counts per 100 students). As with my previous analyses, these variables control for the school's overall discipline practices and collectively represent the level of general misbehavior, as reported by disciplinary consequences. For my fullest specification, I add other observable school-level control variables (school percent special education, school percent FRL, school percent Hispanic, school percent other minority, and indicators for middle schools and high schools (with elementary schools as the omitted group). These variables are all included to further account for school-level characteristics that may be associated with whether a school complies with the policy in the future. To control for time-invariant unobservable characteristics of schools and regions, I include school fixed effects, μ_s, and region fixed effects, $region_s$. The idiosyncratic random error term is indicated by ε_{st}. Standard errors are clustered at the school level.

In the first specification, T_s is a time-invariant variable that identifies whether a school was using OSS as a consequence for truancy in the baseline year, 2012-13, and defines the treatment (T_s = 1) and comparison (T_s = 0) schools. Utilizing interactions with the other

[30] The percent of students who are identified as limited English proficient in each school was not included as an explanatory variable due to its high correlation with Hispanic (r = 0.946).

variables, T_s allows the effect of the policy to be reflected in both a level shift (i.e. β_5) and a shift in trend (i.e. β_6). Thus, the total estimated effect of the policy after three years (by 2015-16), would be $(\widehat{\beta_5} + 3 \times \widehat{\beta_6})$. The term β_4 indicates whether there was a statistically significant difference between the treatment and comparison schools' baseline trends. I do not report these coefficients here, but their statistical significance is not problematic, as it simply indicates that using CITS rather than a DD is important, because CITS is able to account for these baseline trend differences.

In this first specification of Equation 5, T_s simply indicates whether a school used OSS as a consequence for truancy at least once in the 2012-13 school year. It is possible, however, that the effect size may be influenced by the degree to which a school was "subject" to the policy. Schools with very few cases of truancy resulting in OSS may experience little impact relative to schools that use OSS frequently as a consequence for truancy. Alternatively, schools that exclusively or almost exclusively administer OSS for truancy may be less likely to change practices in response to the policy. Therefore, I also estimate models where treatment is defined as the school's percentage of truancy cases in 2012-13 resulting in OSS.

Outcome measures for this analysis are of two main types: end outcomes such as average ELA and math test scores and intermediate outcomes such as student attendance, chronic absenteeism, and disciplinary outcomes theoretically related to the policy.

I estimate policy-related changes in three disciplinary outcomes. One disciplinary outcome that may change following this policy reform is the percent of truancy cases that result in "other" consequences. Perhaps the policy caused treatment schools to use "other" consequences as an alternative to OSS, but given that there was also a decline in the use of ISS for truancy (Tables 2 and 3), it is unclear whether this is really an effect of the policy, or some broader trend. In addition, it is unclear whether this shift towards "other" consequences is occurring more in treatment schools than in comparison schools. Thus, this analysis address whether the policy was associated with this trend.

In addition, the policy could potentially influence how schools report truancy-type incidents: if schools seek to continue administering OSS for truancy-type infractions, but without raising red flags, they could potentially code these cases as "other" infractions instead of as truancy. Therefore, I test whether there are policy-related changes in the number of reported truancy and "other" infractions per 100 students.

In addition to the CITS model utilizing a full set of comparison schools, I also conduct a separate analysis using propensity-score matched schools. Under the assumption that potential outcomes are independent of treatment status, conditional on the included covariates, X, it can be assumed that potential outcomes are also independent of treatment, conditional on propensity score, $\pi(X)$. In other words, if you are able to account for all the covariates that predict treatment, treatment selection behaves as if it has been conditionally randomized, and if this is the case, the only covariate needed to control for selection into treatment is the propensity score, defined as the likelihood of being in the treatment group given the observable baseline characteristics (Rosenbaum & Rubin, 1984). For propensity score matching to produce consistent estimates, the distribution of covariates must be the same for treatment and comparison groups, conditional on estimated propensity score. The advantage of propensity-score matching is that it can test whether the estimates are robust to a model in which comparison and treatment schools are similar in terms of their propensity to be in the "treatment" group, based on observable characteristics. However, because units off common support are excluded, the results may not generalize to the entire population. In addition, given the limited set of observable characteristics, propensity-score matching does not ensure causal estimates.

I conduct propensity-score matching using a probit model to predict whether or not a school used OSS as a consequence for truancy in 2012-13 (which defines treatment status) as a function of the school's log of enrollment, percent FRL, percent Black, percent Hispanic, percent other minority, percent special education, average math scores, average ELA scores, percent of students chronically absent, and percent of student days absent. The results from this probit model are used to estimate a propensity score for each school, and each treatment school is matched, using nearest neighbor matching, to one comparison school, without replacement. A caliper of 0.1 is used, so that each matched school is within a 10 percentage point propensity score of its matched partner. Units without common support are excluded. The sample includes 228 schools or 61% of the full analytic sample for Research Question 3. Baseline equivalence for this matched sample is in Table 5. There are no statistically significant differences in observable characteristics between the treatment and comparison schools within this matched sample.

In the following section, I present the results for each of my three research questions.

Table 5

Baseline Equivalency for Matched Sample

	Treatment	Comparison	Diff.	*p*
Math Z-score	-0.025	-0.052	0.028	0.572
ELA Z-score	-0.048	-0.067	0.019	0.661
School Enrollment	489	509	-20	0.638
School % FRL	58.7%	59.1%	-0.4%	0.856
School % Black	15.4%	17.9%	-2.6%	0.377
School % Hispanic	8.4%	9.4%	-1.0%	0.529
School % Other Minoirty	3.7%	3.9%	-0.2%	0.736
School % Special Education	10.8%	11.3%	-0.5%	0.277
Percent of Students Chronically Absent	13.8%	14.1%	-0.3%	0.850
School Percent of Days Absent	5.2%	5.3%	0.0%	0.899
N of Schools	114	114		

Note: All variables are from baseline year (2012-13).

Results

Research Question 1: What school-level factors predict whether a school was using OSS as a consequence for truancy in the baseline year, 2012-13?

The first question descriptively assesses which schools were administering OSS as a disciplinary consequence for truancy in the baseline year (i.e. exhibiting the behavior Act 1329 sought to address). In addition, Question 1 addresses which schools would have theoretically been affected by the policy in the sense that, at the time the policy was passed, they were using OSS as a consequence for truancy. Thus, these results also provide evidence on which school-level factors are associated with treatment status. I estimate probit models, following Equation 1, among two samples of schools: schools with at least one case of truancy in 2012-13 (columns 1 and 2), and schools with at least five truancy cases in 2012-13 (columns 3 and 4).

The results, in Table 6, indicate that larger schools were more likely to use OSS as a consequence for truancy at least once in 2012-13. In columns 2 and 4, which control for the lagged frequencies of the seven consequence categories, schools with higher rates of truancy, higher rates of OSS, and lower rates of "other" actions in the prior year were more likely to use OSS as a consequence for truancy in 2012-13. In only column four, there is evidence that the rates of corporal punishment use and ISS use in the prior year negatively predict whether OSS was used as a consequence for truancy, but this is marginally significant for ISS. In only column one, there is a positive relationship between the lagged school percent black and whether that school used OSS for truancy in 2012-13. Among schools with at least five truancy infractions (columns 3 and 4), there is a negative relationship between the lagged proportion of other (non-Black, non-Hispanic) minorities and the likelihood of administering OSS for truancy in 2012-13. All other estimated relationships in Table 6 are not statistically significant.

Table 6

Probit Models: Use of OSS as a Consequence for Truancy at Least Once in 2012-13

	Schools with ≥ 1 Truancy Incident		Schools with ≥ 5 Truancy Incidents	
	(1)	(2)	(3)	(4)
Truancy Per 100 Students	0.003	0.014 ***	0.003	0.014 ***
	(0.003)	(0.004)	(0.003)	(0.004)
Percent of Students Disciplined	0.151	0.033	0.207	0.665
	(0.263)	(0.507)	(0.316)	(0.516)
ln(school enrollment)	0.162 ***	0.171 ***	0.277 ***	0.265 ***
	(0.057)	(0.057)	(0.070)	(0.067)
School % FRL	-0.021	0.121	-0.005	0.077
	(0.280)	(0.261)	(0.341)	(0.299)
School % Black	0.348 **	0.040	0.218	-0.076
	(0.161)	(0.159)	(0.217)	(0.220)
School % Hispanic	0.413	0.278	0.192	-0.020
	(0.304)	(0.334)	(0.355)	(0.414)
School % Other Minority	-1.030	-0.583	-3.177 **	-2.440 **
	(1.112)	(0.970)	(1.241)	(1.197)
School % Special Education	0.881	0.939	1.035	0.797
	(0.845)	(0.850)	(1.067)	(1.055)
District Conversion Charter School	0.295	0.059		
	(0.265)	(0.260)		
High School	0.119	0.074	0.018	-0.014
	(0.080)	(0.075)	(0.166)	(0.141)
Middle School	-0.019	-0.0599	-0.009	-0.041
	(0.072)	(0.072)	(0.159)	(0.137)
Average Math Z-score	-0.143	-0.082	-0.11	-0.042
	(0.104)	(0.094)	(0.118)	(0.108)
Average ELA Z-score	0.095	0.146	0.017	-0.005
	(0.118)	(0.116)	(0.148)	(0.142)
ALE Per 100 Students		0.158		0.176
		(0.232)		(0.313)
Corporal Punishment Per 100 Students		-0.000		-0.004 **
		(0.001)		(0.002)
Expulsion Per 100 Students		-0.121		-0.293
		(0.194)		(0.193)
ISS Per 100 Students		-0.002		-0.003 *
		(0.002)		(0.002)
No Action Per 100 Students		-0.005		-0.038
		(0.012)		(0.033)
Other Action Per 100 Students		-0.002 **		-0.002 **
		(0.001)		(0.001)
OSS Per 100 Students		0.014 ***		0.013 ***
		(0.003)		(0.004)
Region Indicators	Y	Y	Y	Y
Observations	361	361	219	219

Note. Estimates are marginal effects. Robust standard errors, clustered at the district level, are in parentheses. Dependent variable is a binary indicator for whether OSS was used as a consequence for truancy at least once in 2012-13. Explanatory variables are lagged. Variables omitted due to perfect collinearity: open enrollment charter school and, in some models, district conversion charter school. Five region indicators with largest region (NW) as omitted region. Omitted school type is elementary.
*** p<0.01, ** p<0.05, * p<0.1.

Research Question 2: What school-level factors predict whether a school, among those using OSS as a consequence for truancy in 2012-13, complies with the policy in the following years?

First, I study the schools that, as of 2012-13, were using OSS as a consequence for truancy, and assess what school-level characteristics are associated with whether those schools fully comply (use OSS for zero truancy cases) in each of the three following years. Results for three probit models, one for each outcome year, are in Table 7. The sample in Table 7 only includes schools with at least one truancy case in the outcome year.

In the first year following the passage of Act 1329 (2013-14), schools with higher truancy rates in the prior year were less likely to comply. In addition, schools with higher proportions of minority students were generally less likely to comply in 2013-14. For example, a 10 percentage point increase in the proportion of students in the school who are Black is associated with a 4.1 to 6.1 percentage point decrease in the likelihood of full compliance in 2013-14. Schools with higher proportions of Hispanic and other minority students also have a lower likelihood of compliance, although the coefficient on percent Hispanic is not significant in column 2. High schools were 22.8 to 27.1 percentage points less likely to comply in 2013-14, relative to elementary schools. In column 2, which includes the rates of each consequence type, schools that used ISS more frequently in the prior year were slightly more likely to comply, and schools that used OSS more frequently in the prior year were less likely to comply.

The results from the models predicting full compliance in 2014-15 are somewhat different. The percent of students who are Black or Hispanic are still generally negatively associated with compliance, however the percent of students who are other minorities (Asian, Native American, Hawaiian/Pacific Islander, or two or more races) is positively associated with compliance in 2014-15. These other minority groups represent only about 5% of Arkansas's public school students, so I do not focus much on this result. In column 3 only, there is suggestive evidence that schools with more FRL students were more likely to comply, all else equal, although it is possible this estimated positive relationship is due in part to the correlation between school percent Black and school percent FRL (r = 0.4867). In column 4 only, there is evidence that larger schools were less likely to comply in 2014-15, schools with a higher proportion of students in special education were more likely to comply in 2014-15, and schools with higher truancy rates, higher OSS rates, lower "other" consequence rates, or lower corporal punishment rates in 2013-14, all else equal, were less likely to comply in 2014-15.

In columns 5 and 6, schools with higher 2014-15 truancy rates were less likely to comply in 2015-16. In one model (column 6), high schools and middle schools were less likely to comply in 2015-16 than elementary schools, and schools with higher prior-year ALE rates, higher prior-year "no action" rates, or lower prior-year OSS rates were more likely to comply in 2015-16. The relationship between the prior year frequency of ALE referrals and policy compliance in column 6 is quite large, but the direction of this relationship does not appear consistent across years. In addition, this result is not consistent with the results from the ordered probit models that I discuss later in this section. Furthermore, only 0.31% of all the disciplinary consequences in Arkansas during this time period were ALE referrals, so I do not emphasize this result and treat the ALE rate primarily as a control variable.

In Table 7, the primary consistent drivers of compliance are truancy rates and OSS rates. Schools with higher prior year truancy rates were slightly less likely to comply in 2013-14 and 2015-16, and perhaps in 2014-15 as well, although this last result was not significant in column 3. Schools with higher prior year OSS rates were less likely to comply in each of the three outcome years. Schools that use consequences other than OSS more frequently generally appear to be as likely or more likely to comply with the policy, all else equal. These results are not consistent across all years, but may indicate that schools that do not rely as much on OSS are better able to respond to a policy change that seeks to reduce the use of OSS. Another takeaway from Table 7 is that racial demographics predicted compliance in the first two years, but not the final outcome year. Instead, in the final year, compliance is predicted by school type, truancy rates, and OSS rates, suggesting that the schools struggling to comply even after three years are likely the ones targeted by the policy (high schools, middle schools, schools with more truancy, and schools with greater reliance on OSS).

To reduce the possibility that these results are driven by schools with very few truancy cases, I estimate the same models but limit the sample to schools with at least five truancy cases in the outcome years. The results, in Table 8, are generally robust to this restriction, although many relationships lose or gain statistical significance, compared to Table 7. Given that the results from Table 7 include some schools with very small numbers of truancy (1-4), the results in Table 8 may be more representative of schools that struggle with truancy at a significant level. I do not discuss all the differences between Tables 7 and 8 here, but a few are worth highlighting.

First, the inconsistent coefficient on school percent other minority from Table 7 (negative in 2013-14 but positive in 2014-15) tells a more consistent story in Table 8 (negative or null across all models). Further, while in Table 7, the only significant coefficients for high schools and middle schools are negative, in Table 8, these coefficients are estimated to be positive in 2014-15. Thus, there is an inconsistent story about what types of schools (in terms of grade levels served) are likely to comply.

Table 7

Probit Models of Compliance (Schools with at Least One Truancy Incident)

	2013-14		2014-15		2015-16	
	(1)	(2)	(3)	(4)	(5)	(6)
Truancy Per 100 Students	-0.024 **	-0.029 **	-0.002	-0.012 **	-0.024 ***	-0.026 ***
	(0.012)	(0.013)	(0.005)	(0.006)	(0.009)	(0.008)
Percent of Students Disciplined	-0.092	-0.808	-0.094	-0.497	0.126	1.073
	(0.303)	(0.709)	(0.341)	(0.460)	(0.333)	(0.683)
ln(school enrollment)	-0.032	-0.106	-0.089	-0.142 *	-0.066	-0.113
	(0.062)	(0.066)	(0.078)	(0.074)	(0.058)	(0.074)
School % FRL	0.285	0.033	0.849 *	0.484	0.503	0.260
	(0.414)	(0.464)	(0.449)	(0.452)	(0.411)	(0.453)
School % Black	-0.607 ***	-0.412 *	-0.823 ***	-0.208	-0.264	-0.063
	(0.197)	(0.220)	(0.278)	(0.308)	(0.247)	(0.262)
School % Hispanic	-0.784 **	-0.529	-1.348 ***	-1.271 ***	-0.027	0.223
	(0.363)	(0.395)	(0.365)	(0.389)	(0.379)	(0.367)
School % Other Minority	-3.546 **	-3.082 **	2.601 **	3.299 ***	-1.667	-1.547
	(1.396)	(1.417)	(1.104)	(1.163)	(1.107)	(1.049)
School % Sped	-0.464	-0.210	0.664	2.560 **	-0.369	-1.124
	(1.026)	(1.146)	(1.238)	(1.238)	(1.087)	(1.081)
Open Enrollment Charter	0.327	0.131	0.371	0.068	-	-
	(0.215)	(0.234)	(0.277)	(0.265)		
District Conversion Charter	-	-	-0.202	0.152	-	-
			(0.247)	(0.325)		
High School	-0.228 *	-0.271 **	0.030	0.126	-0.102	-0.184 *
	(0.130)	(0.129)	(0.180)	(0.141)	(0.117)	(0.111)
Middle School	-0.008	-0.074	0.149	0.191	-0.118	-0.185 *
	(0.113)	(0.115)	(0.181)	(0.144)	(0.112)	(0.106)
Average Math Z-score	-0.143	-0.133	0.165	0.165	0.096	0.121
	(0.119)	(0.112)	(0.164)	(0.138)	(0.148)	(0.143)
Average ELA Z-score	0.145	0.117	0.038	-0.010	0.143	0.042
	(0.140)	(0.130)	(0.235)	(0.194)	(0.188)	(0.184)
ALE Per 100 Students		0.054		-0.214		0.154 ***
		(0.113)		(0.209)		(0.059)
Corporal Punishment Per 100 Students		0.001		0.004 **		-0.003
		(0.002)		(0.002)		(0.004)
Expulsion Per 100 Students		-0.093		0.001		-0.259
		(0.117)		(0.099)		(0.271)
ISS Per 100 Students		0.006 **		0.002		0.000
		(0.003)		(0.002)		(0.001)
No Action Per 100 Students		0.149		-0.014		0.041 *
		(0.154)		(0.015)		(0.022)
Other Action Per 100 Students		0.000		0.003 ***		-0.003
		(0.001)		(0.001)		(0.002)
OSS Per 100 Students		-0.007 *		-0.014 ***		-0.011 ***
		(0.004)		(0.004)		(0.003)
Region Indicators	Y	Y	Y	Y	Y	Y
Observations	160	160	154	154	152	152

Note. Estimates are marginal effects. Robust standard errors, clustered at the district level, are in parentheses. All explanatory variables are lagged except for region and school type variables. Percent variables on a 0-1 scale. Variables omitted due to perfect collinearity include other school type (not elementary, middle, or high), and in some models, indicators for open enrollment and district conversion charter schools. Five region indicators with largest region, NW, as omitted group. Omitted school type is elementary.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8

Probit Models of Compliance (Schools with at Least Five Truancy Incidents)

	2013-14		2014-15		2015-16	
	(1)	(2)	(3)	(4)	(5)	(6)
Truancy Per 100 Students	-0.022 *	-0.033 ***	0.002	-0.005	-0.016 **	-0.018 **
	(0.013)	(0.011)	(0.004)	(0.006)	(0.008)	(0.007)
Percent of Students Disciplined	-0.030	-0.753	0.030	-0.383	0.136	1.115
	(0.366)	(0.779)	(0.334)	(0.450)	(0.358)	(0.720)
ln(school enrollment)	0.082	0.014	0.090	0.049	-0.007	-0.033
	(0.065)	(0.070)	(0.106)	(0.086)	(0.066)	(0.087)
School % FRL	0.482	0.089	1.263 **	0.925 **	0.357	-0.023
	(0.391)	(0.403)	(0.564)	(0.460)	(0.444)	(0.531)
School % Black	-0.436 **	-0.319	-0.769 ***	-0.123	-0.374	-0.091
	(0.206)	(0.217)	(0.285)	(0.280)	(0.246)	(0.268)
School % Hispanic	-1.212 ***	-0.549	-1.707 ***	-1.859 ***	-0.235	0.036
	(0.404)	(0.414)	(0.446)	(0.471)	(0.390)	(0.420)
School % Other Minority	-4.804 **	-4.155 **	-1.086	1.347	-2.350 *	-1.903
	(1.882)	(1.842)	(1.945)	(1.356)	(1.322)	(1.289)
School % Sped	1.112	1.016	0.886	3.058 *	-0.902	-1.796
	(0.954)	(0.962)	(1.657)	(1.775)	(1.154)	(1.215)
High School	-0.124	-0.212 *	1.530 ***	1.336 ***	-0.123	-0.253 *
	(0.133)	(0.126)	(0.187)	(0.137)	(0.108)	(0.139)
Middle School	-0.040	-0.155	1.688 ***	1.456 ***	-0.077	-0.207
	(0.108)	(0.118)	(0.184)	(0.161)	(0.121)	(0.153)
Average Math Z-score	-0.154	-0.172	0.280	0.206	0.004	0.041
	(0.121)	(0.105)	(0.195)	(0.158)	(0.157)	(0.145)
Average ELA Z-score	0.354 **	0.318 **	0.120	-0.034	0.173	0.017
	(0.157)	(0.147)	(0.265)	(0.215)	(0.194)	(0.185)
ALE Per 100 Students		0.377 ***		-0.068		0.133 **
		(0.105)		(0.114)		(0.064)
Corporal Punishment Per 100 Students		0.003		0.003 *		-0.001
		(0.002)		(0.002)		(0.004)
Expulsion Per 100 Students		-0.056		-0.246 **		-0.201
		(0.098)		(0.110)		(0.297)
ISS Per 100 Students		0.008 ***		0.001		0.000
		(0.003)		(0.002)		(0.001)
No Action Per 100 Students		-0.051		-0.005		0.041 **
		(0.100)		(0.016)		(0.020)
Other Action Per 100 Students		-0.005 **		0.003 ***		-0.003
		(0.002)		(0.001)		(0.002)
OSS Per 100 Students		-0.009 **		-0.017 ***		-0.012 ***
		(0.004)		(0.005)		(0.004)
Region Indicators	Y	Y	Y	Y	Y	Y
Observations	130	130	119	119	123	123

Note. Estimates are marginal effects. Robust standard errors, clustered at the district level, are in parentheses. All explanatory variables are lagged except for region and school type variables. Percent variables are on a 0-1 scale. Variables omitted due to perfect collinearity include other school type (not elementary, middle, or high) and indicators for open enrollment and district conversion charter schools. Five region indicators with largest region, NW, as omitted group. Omitted school type is elementary.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Another key relationship that gained significance in Table 8 is that schools with higher ELA scores positively predict compliance in 2013-14. Some of the coefficients on rates of different consequence types gained significance in Table 8 (relative to Table 7), such as the coefficients on ALE rates and "other" action rates in 2013-14, and the coefficient on expulsion rates in 2014-15. In addition, the coefficients on truancy rates and log of school enrollment in the 2014-15 model lost their significance in Table 8, relative to Table 7. The relationship between school percent black and compliance in column 2 was not significant in Table 8 (although it was in Table 7), and the relationship between school percent FRL and compliance in column 4 of Table 8 gained significance. Other estimates were not substantively changed.

Thus, summarizing the results from Tables 7 and 8, the most consistent results are that OSS rates and truancy rates generally negatively predict compliance. Minority rates also negatively predict compliance, but primarily in the first two outcome years. These results suggest that the schools failing to comply, even after three years, are likely the types of schools targeted by this policy (those with high truancy rates and high OSS rates).

Given that some schools return to non-compliance after a year of compliance, I also estimate ordered probit models predicting whether a school complied in zero, one, two, or three of the outcome years. Here, as elsewhere, compliance is defined as administering OSS in zero percent of truancy cases, and non-compliance is defined as administering OSS as a consequence for truancy in at least one truancy case. The results are in Table 9.

According to Table 9, schools with more frequent truancy and more frequent use of OSS in the baseline year are found to be more likely to comply in zero of the three outcome years and less likely to comply in two or three outcome years. These results are consistent with the findings from our earlier models, as truancy and OSS rates tend to negatively predict compliance.

Table 9

Ordered Probit Results (Schools with at Least one Truancy Incident)

	Marginal Effects Predicting Likelihood of:							
	Compliance in 0 Years		Compliance in 1 Year		Compliance in 2 Years		Compliance in 3 Years	
Truancy Per 100 Students	0.013 *	0.020 **	-0.002	-0.002	-0.005 *	-0.007 **	-0.006 *	-0.011 **
	(0.007)	(0.008)	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.005)
Pct of Students Disciplined	0.137	0.052	-0.019	-0.004	-0.051	-0.020	-0.068	-0.028
	(0.385)	(0.782)	(0.052)	(0.066)	(0.144)	(0.297)	(0.191)	(0.420)
ln(school enrollment)	0.077	0.084	-0.011	-0.007	-0.028	-0.032	-0.038	-0.045
	(0.059)	(0.059)	(0.009)	(0.007)	(0.023)	(0.023)	(0.030)	(0.033)
School % FRL	-0.693 **	-0.451	0.095	0.038	0.257 **	0.171	0.342 *	0.242
	(0.340)	(0.330)	(0.059)	(0.038)	(0.130)	(0.126)	(0.183)	(0.185)
School % Black	0.707 ***	0.452 **	-0.097 **	-0.038	-0.262 ***	-0.171 *	-0.348 ***	-0.242 *
	(0.204)	(0.215)	(0.045)	(0.030)	(0.088)	(0.089)	(0.120)	(0.125)
School % Hispanic	1.076 ***	0.911 ***	-0.148 **	-0.077	-0.398 ***	-0.346 ***	-0.530 ***	-0.489 ***
	(0.271)	(0.264)	(0.067)	(0.058)	(0.116)	(0.114)	(0.171)	(0.169)
School % Other Minority	1.011	0.964	-0.139	-0.081	-0.374	-0.366	-0.498	-0.517
	(0.820)	(0.776)	(0.117)	(0.082)	(0.315)	(0.310)	(0.416)	(0.423)
School % Sped	-1.212	-1.200	0.166	0.101	0.449	0.455	0.597	0.644
	(0.974)	(0.948)	(0.157)	(0.111)	(0.361)	(0.357)	(0.489)	(0.527)
High School	0.010	0.001	-0.001	-0.000	-0.004	-0.000	-0.005	-0.000
	(0.138)	(0.140)	(0.019)	(0.012)	(0.051)	(0.053)	(0.068)	(0.075)
Middle School	-0.179	-0.199	0.025	0.017	0.066	0.076	0.088	0.107
	(0.128)	(0.138)	(0.021)	(0.019)	(0.050)	(0.056)	(0.063)	(0.072)
Average Math Z-score	0.014	0.016	-0.002	-0.001	-0.005	-0.006	-0.007	-0.009
	(0.128)	(0.124)	(0.018)	(0.011)	(0.047)	(0.047)	(0.063)	(0.067)
Average ELA Z-score	-0.302 **	-0.139	0.041	0.012	0.112 *	0.053	0.149 *	0.075
	(0.145)	(0.160)	(0.026)	(0.016)	(0.057)	(0.060)	(0.076)	(0.088)
ALE Per 100 Students		0.230 **		-0.019		-0.087 **		-0.123 *
		(0.104)		(0.016)		(0.039)		(0.064)
Corporal Pun. Per 100 Stud.		-0.001		0.000		0.000		0.001
		(0.003)		(0.000)		(0.001)		(0.001)
Expulsion Per 100 Stud.		0.108		-0.009		-0.041		-0.058
		(0.077)		(0.009)		(0.031)		(0.042)
ISS Per 100 Stud.		-0.003		0.000		0.001		0.002
		(0.003)		(0.000)		(0.001)		(0.002)
No Action Per 100 Stud.		-0.008		0.001		0.003		0.004
		(0.148)		(0.012)		(0.056)		(0.079)
Other Action Per 100 Stud.		-0.000		0.000		0.000		0.000
		(0.001)		(0.000)		(0.000)		(0.000)
OSS Per 100 Stud.		0.010 **		-0.001		-0.004 **		-0.006 **
		(0.004)		(0.001)		(0.002)		(0.003)
Region Indicators	Y	Y	Y	Y	Y	Y	Y	Y
Observations	132	132	132	132	132	132	132	132

Note. Estimates are marginal effects. Explanatory variables are from the baseline year (2012-13). Robust standard errors, clustered at the district level, are in parentheses. Five region indicators with the largest region (NW) as the omitted group. Omitted school type is elementary school, and one K-12 school was included in this omitted group, as it was the only school in the sample that was not classified as an elementary, middle, or high school. The results are generally robust to whether this one school is indicated with a separate school type indicator (except that the significant marginal effects for School % FRL were not significant).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

In addition, schools with greater proportions of Black and Hispanic students are more likely to comply in zero years, and generally less likely to comply in one, two, or three years. These race percentage variables are on a 0-1 scale, so a ten percentage point increase in the share of Black students in the school is associated with a 4.5 to 7.1 percentage point higher likelihood of complying in zero of the outcome years and a 2.4 to 3.5 percentage point lower likelihood of complying all three years, for example. The marginal effects for Hispanic share are even larger, but given the lower Hispanic share in the state overall (12% versus 21% for Black students), it is important to note that moving Hispanic share by ten percentage points represents a very large change. As seen in previous models, school percent FRL appears to predict compliance in the opposite direction as these minority variables, but given the correlation between school percent FRL and school percent Black ($r = 0.4867$), and the relative inconsistency of this estimated relationship, I recommend interpreting this with caution.

Table 9 also indicates that higher ELA test scores are associated with a higher likelihood of compliance, a relationship suggested in Table 8, but not in Table 7. In Table 9, higher ELA test scores are associated with a lower likelihood of compliance in zero years and a higher likelihood of compliance in two or three years. However, these results are only significant in the models that do not control for baseline levels of each consequence type, which are correlated with school academic achievement. Still, this is suggestive evidence that underperforming schools, which tend to also be the ones with high discipline rates, struggle with compliance.

Finally, in Table 9, the baseline rate of ALE use is negatively associated with compliance. This is the only major result in Table 9 that is clearly inconsistent with the results from the year-by-year models in Table 7 and 8, in which I estimated that higher ALE rates positively predicted compliance in some years and models. This difference could be due, in part, to using 2012-13 ALE rates for the models in Table 9 and the prior year rate for each of the models in Tables 7 and 8. Regardless, given the inconsistency of this result across models, and the relatively low prevalence of ALE rates overall, I do not place much focus on this result.

So far, the results from these models suggest that the schools likely targeted by this policy continue to struggle with compliance in future years. However, there is evidence, in Table 3, that some schools do change how they report handling truancy. As a result, it is important to measure what school-level outcomes may have changed along with this policy.

Research Question 3: How did school-level outcomes such as average math and ELA test scores, school attendance, chronic absenteeism, and disciplinary outcomes change in "policy-affected" schools (i.e., those using OSS for truancy in 2012-13)?

The results of the CITS, used to test whether, after the policy, there was a differential change in key outcomes in treatment schools, relative to comparison schools, are in Tables 10 through 13. Panel A in each of these four tables shows the results for the simple definition of treatment status, T_s, indicating whether a school used OSS as a consequence for truancy at least once (T_s=1) or not at all (T_s=0) in 2012-13. Panel B shows the results for an alternative definition of treatment status defined as the school's percentage of truancy cases in 2012-13 that resulted in OSS. It is theorized that schools with a high percentage of truancy resulting in OSS would potentially have been affected more by the policy. I report β_5 and β_6, the coefficients on $(T_s \times POLICY_t)$ and $(T_s \times YR\_SINCE\_POLICY_t)$, respectively, as well as the linear combination $(\beta_5 + 3 \times \beta_6)$ which is an estimate of the total effect by 2015-16 (after three years). The models without the matched comparison group are in the left of each table, and the results from the propensity-score matched sample are in the right side of each table.

Table 10 presents the results from models predicting two measures of student absenteeism: the percent of days absent (the sum of all student days absent divided by the sum of total school days) as well as the percent of students chronically absent (missing 10% or more of school days) in a given year. Theoretically, the impact of this policy could be positive or negative: perhaps students are more likely to be absent if truancy is effectively decriminalized. Or, if the interventions replacing OSS are reengaging truant students in the school community, perhaps their effect would be positive. The estimated relationship of the policy to these two attendance-related outcomes after three years is null, except that, among the propensity-score matched sample, there was a marginally significant increase in the percent of days absent in treatment schools, relative to comparison schools. The dependent variable is on a 0-1 scale, so these estimates are small (about 0.7 days of additional absences per 100 days of school).

Table 11 indicates the policy-related change in the use of "other" consequences for truancy. Tables 2 and 3 indicate that schools were shifting away from OSS and ISS towards these "other" consequences, but the CITS analysis estimates the size of the policy-related change by using the comparison schools as the counterfactual for what theoretically would have happened in the treatment schools in the absence of the policy. The results in Panel A indicate a policy-related increase in the use of "other" consequences for truancy of about 11 percentage points in the non-matched sample, and 16 to 19 percentage points in the matched sample.

The total "effect" in Panel B of Table 11 represents the policy-related change, after three years, for a school using OSS for 100% of truancy cases in 2012-13, compared to a school using OSS for 0% of truancy cases in 2012-13. As hypothesized, the schools administering OSS for 100% of truancy cases appear to have a larger change post-policy than the average treatment school in Panel A.

Table 10

Estimated Policy-related Change in Attendance and Chronic Absenteeism

	Without Matching						With Propensity-Score Matched Smaple					
	Percent of Days Absent			Pct. of Stud. Chronically Absent			Percent of Days Absent			Pct. of Stud. Chronically Absent		
Panel A: T_s = used OSS for truancy at least once in 2012-13												
$T_s \times Policy_t$	0.006 **	0.006 ***	0.007 ***	0.013	0.015	0.016	0.010 ***	0.010 ***	0.011 ***	0.023 **	0.024 **	0.027 **
	(0.002)	(0.002)	(0.002)	(0.010)	(0.010)	(0.010)	(0.003)	(0.003)	(0.003)	(0.011)	(0.012)	(0.012)
$T_s \times$ Years since policy	-0.001	-0.001	-0.002	-0.005	-0.005	-0.006	-0.001	-0.001	-0.001	-0.004	-0.005	-0.004
	(0.001)	(0.001)	(0.001)	(0.005)	(0.005)	(0.005)	(0.001)	(0.001)	(0.001)	(0.006)	(0.006)	(0.006)
Total "effect" by 2015-16	0.002	0.002	0.002	-0.002	-0.002	-0.002	0.007 *	0.007 *	0.007 *	0.012	0.011	0.014
	(0.003)	(0.003)	(0.003)	(0.014)	(0.013)	(0.013)	(0.004)	(0.004)	(0.004)	(0.016)	(0.016)	(0.016)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1959	1959	1959	1959	1959	1959
Adjusted R-squared	0.032	0.050	0.057	0.051	0.068	0.078	0.041	0.057	0.066	0.059	0.085	0.101
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13												
$T_s \times Policy_t$	0.006 *	0.008 **	0.008 **	0.012	0.016	0.018	0.013 **	0.013 **	0.015 **	0.022	0.024	0.029
	(0.004)	(0.004)	(0.004)	(0.017)	(0.017)	(0.017)	(0.005)	(0.006)	(0.006)	(0.025)	(0.026)	(0.026)
$T_s \times$ Years since policy	-0.003	-0.002	-0.003	-0.016 **	-0.016 **	-0.017 **	-0.001	-0.001	-0.002	-0.014	-0.016 *	-0.018 *
	(0.002)	(0.002)	(0.002)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)	(0.010)	(0.010)	(0.009)
Total "effect" by 2015-16 [a]	-0.001	0.0005	0.0002	-0.037	-0.032	-0.034	0.010	0.010	0.010	-0.021	-0.024	-0.023
	(0.006)	(0.006)	(0.005)	(0.023)	(0.023)	(0.022)	(0.009)	(0.009)	(0.008)	(0.037)	(0.036)	(0.035)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-Squared	0.032	0.051	0.057	0.053	0.070	0.080	0.046	0.063	0.071	0.061	0.087	0.102
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Dependent variables are on a 0-1 scale. Sch. chars. predicting baseline compliance include: log of enrollment, percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group).

[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 11

Estimated Policy-related Change in the Use of "Other" Consequences for Truancy

	Without Matching			With Propensity-Score Matched Smaple		
	Pct of Truancy Resulting in "Other"			Pct of Truancy Resulting in "Other"		
Panel A: T_s = used OSS for truancy at least once in 2012-13						
$T_s \times Policy_t$	0.069 *	0.065 *	0.066 *	0.124 **	0.125 **	0.127 **
	(0.038)	(0.037)	(0.037)	(0.051)	(0.050)	(0.050)
$T_s \times$ Years since policy	0.015	0.016	0.016	0.012	0.020	0.020
	(0.020)	(0.019)	(0.019)	(0.027)	(0.026)	(0.026)
Total "effect" by 2015-16	0.113 ***	0.113 ***	0.115 ***	0.159 ***	0.186 ***	0.187 ***
	(0.043)	(0.041)	(0.041)	(0.058)	(0.054)	(0.054)
Number of Schools	374	374	374	228	228	228
Observations	2,631	2,631	2,631	1,645	1,645	1,645
Adjusted R-squared	0.177	0.249	0.249	0.186	0.241	0.240
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13						
$T_s \times Policy_t$	0.058	0.063	0.066	0.173	0.162	0.168
	(0.061)	(0.059)	(0.059)	(0.111)	(0.109)	(0.109)
$T_s \times$ Years since policy	0.025	0.027	0.028	0.047	0.053	0.053
	(0.033)	(0.031)	(0.032)	(0.056)	(0.054)	(0.055)
Total "effect" by 2015-16 [a]	0.134 *	0.143 **	0.150 **	0.314 ***	0.321 ***	0.326 ***
	(0.070)	(0.067)	(0.067)	(0.110)	(0.106)	(0.106)
Number of Schools	374	374	374	228	228	228
Observations	2,631	2,631	2,631	1,645	1,645	1,645
Adjusted R-Squared	0.173	0.245	0.245	0.192	0.245	0.243
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Dependent variables are on a 0-1 scale. Sch. chars. predicting baseline compliance include: log of enrollment, percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group).
[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

I also estimate whether the policy was related to changes in reports of truancy or "other" infractions not included in a state-reported infraction category[31] (Table 12). There is suggestive evidence (null or marginally significant) that there was a reduction of 2-3 truancy cases per 100

[31] Like "other" consequences, "other" infractions are coded by schools, but when reported by the state, grouped into an "other" category. This is not a researcher-created group of infractions.

students, depending on the model and sample. Fewer truancy reports could indicate either a real reduction in truancy behavior, or a change in reporting behavior. Unfortunately, I cannot directly test this with the available data. The relationship between the policy and truancy rates was not significant in Panel B, however, so the estimated effect is sensitive to treatment definition, and there is not consistent evidence that the policy was related to a decline in reports of truancy.

I also test whether there was a policy-related change in reports of “other” infractions. Some schools may code truancy-type incidents as something else in order to comply with this policy. I estimate, in some models, a policy-related increase in “other” infractions of about 10 incidents per 100 students after three years. Given the size of this estimate, relative to the small reduction in truancy, it appears something else is happening besides the new policy. Further, the same estimates are not significant in Panel B, suggesting this increase may not be policy-related. Indeed, the rise in “other” infractions was occurring even before the policy change. “Other” infractions represented 19.1% of total infractions in 2008-09 and had risen to 24.0% by the baseline year (2012-13). Overall, these results suggest that, around the time of the policy change, treatment schools - more so than comparison schools - started having more of these types of infractions or started increasing their reporting of these types of behaviors.

Finally, I estimate whether there was a policy-related change in test scores. Perhaps when schools use OSS as a consequence for truancy, there is a dampening effect on student achievement. If this is the case, and if the policy keeps truant students in school, it is possible that student achievement could improve. Alternatively, if truant and potentially disruptive students remain in schools as a result of this policy, these students could disrupt their peers and negatively impact overall test scores. I estimate, across both definitions of treatment status and both samples, no policy-related change in school-level test scores (see Table 13).

Table 12

Estimated Policy-related Change in Disciplinary Infractions (Truancy and "Other")

	Without Matching						With Propensity-Score Matched Smaple					
	Truancy Per 100 Students			"Other" Infractions Per 100 Students			Truancy Per 100 Students			"Other" Infractions Per 100 Students		
Panel A: T_s = used OSS for truancy at least once in 2012-13												
$T_s \times Policy_t$	0.208	0.118	0.019	1.265	2.097	1.980	0.047	0.004	-0.114	4.757	3.931	3.855
	(0.799)	(0.771)	(0.762)	(4.880)	(2.972)	(2.989)	(0.818)	(0.967)	(0.939)	(4.429)	(3.389)	(3.423)
$T_s \times$ Years since policy	-0.766	-0.744	-0.746	2.522	2.594 *	2.706 *	-1.049	-0.880	-0.850	1.366	2.208	2.248
	(0.494)	(0.487)	(0.483)	(2.221)	(1.474)	(1.463)	(0.646)	(0.691)	(0.695)	(2.862)	(1.930)	(1.968)
Total effect by 2015-16	-2.089	-2.113 *	-2.220 *	8.831	9.880 **	10.100 **	-3.099 *	-2.635 *	-2.663	8.854	10.550 *	10.600 *
	(1.355)	(1.165)	(1.184)	(6.457)	(4.693)	(4.678)	(1.740)	(1.556)	(1.633)	(7.862)	(5.627)	(5.832)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-squared	0.012	0.181	0.183	0.097	0.588	0.589	0.008	0.133	0.134	0.105	0.447	0.447
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13												
$T_s \times Policy_t$	-1.101	-0.457	-0.579	-10.940 **	-3.600	-3.779	-0.850	-1.066	-1.285	0.557	-0.520	-0.780
	(0.806)	(0.855)	(0.846)	(5.232)	(3.315)	(3.332)	(1.046)	(1.369)	(1.332)	(5.629)	(4.131)	(4.159)
$T_s \times$ Years since policy	0.056	0.057	0.064	0.778	0.858	1.077	0.038	0.001	0.101	2.584	1.741	1.958
	(0.515)	(0.486)	(0.484)	(2.502)	(1.822)	(1.805)	(0.777)	(0.802)	(0.780)	(3.433)	(2.553)	(2.595)
Total effect by 2015-16 [a]	-0.933	-0.287	-0.386	-8.606	-1.027	-0.547	-0.735	-1.064	-0.982	8.309	4.702	5.095
	(1.351)	(1.093)	(1.106)	(7.758)	(4.905)	(4.926)	(1.854)	(1.489)	(1.530)	(8.726)	(6.427)	(6.749)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-Squared	0.009	0.179	0.181	0.097	0.588	0.590	0.005	0.131	0.132	0.105	0.446	0.447
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Sch. chars. predicting baseline compliance include: log of enrollment, percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group).

[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 13

Estimated Policy-related Change in School-Level Average Test Scores

	Without Matching						With Propensity-Score Matched Smaple					
	Math Z-scores			ELA Z-scores			Math Z-scores			ELA Z-scores		
Panel A: T_s = used OSS for truancy at least once in 2012-13												
$T_s \times Policy_t$	0.018	0.018	0.016	0.038	0.035	0.032	0.044	0.042	0.036	0.023	0.019	0.015
	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)	(0.038)	(0.038)	(0.039)	(0.035)	(0.035)	(0.036)
$T_s \times$ Years since policy	-0.008	-0.007	-0.006	-0.008	-0.007	-0.005	-0.035	-0.032	-0.032	-0.006	-0.003	-0.002
	(0.016)	(0.017)	(0.017)	(0.015)	(0.015)	(0.015)	(0.022)	(0.022)	(0.022)	(0.019)	(0.019)	(0.019)
Total "effect" by 2015-16	-0.006	-0.004	-0.002	0.015	0.015	0.018	-0.061	-0.054	-0.061	0.004	0.011	0.007
	(0.040)	(0.040)	(0.040)	(0.032)	(0.032)	(0.032)	(0.052)	(0.052)	(0.052)	(0.040)	(0.040)	(0.040)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	2,863	2,863	2,863	2,860	2,860	2,860	1,752	1,752	1,752	1,752	1,752	1,752
Adjusted R-squared	0.017	0.019	0.023	0.029	0.037	0.053	0.022	0.029	0.037	0.034	0.051	0.066
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13												
$T_s \times Policy_t$	0.034	0.033	0.028	0.062	0.057	0.052	0.096	0.095	0.083	0.079	0.076	0.071
	(0.042)	(0.043)	(0.043)	(0.048)	(0.047)	(0.047)	(0.070)	(0.071)	(0.072)	(0.059)	(0.059)	(0.059)
$T_s \times$ Years since policy	-0.008	-0.008	-0.006	-0.006	-0.006	-0.003	-0.048	-0.043	-0.040	-0.025	-0.020	-0.017
	(0.026)	(0.026)	(0.026)	(0.022)	(0.021)	(0.021)	(0.043)	(0.043)	(0.044)	(0.028)	(0.028)	(0.027)
Total "effect" by 2015-16 [a]	0.010	0.008	0.011	0.045	0.039	0.044	-0.048	-0.035	-0.037	0.003	0.016	0.021
	(0.062)	(0.062)	(0.062)	(0.049)	(0.048)	(0.048)	(0.097)	(0.097)	(0.097)	(0.065)	(0.063)	(0.062)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	2,863	2,863	2,863	2,860	2,860	2,860	1,752	1,752	1,752	1,752	1,752	1,752
Adjusted R-Squared	0.020	0.023	0.027	0.028	0.037	0.053	0.023	0.030	0.037	0.035	0.051	0.067
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Sch. chars. predicting baseline compliance include: log of enrollment, percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group).

[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Robustness Checks

A key threat to causal inference is the assumption that deviations from trend within the comparison schools serve as a valid counterfactual for what would have happened in the treatment schools in the absence of the policy. This is a strong assumption and would be invalidated if unobservable factors changed around the time of the policy in different ways in the treatment and comparison schools. For example, if, independent of this policy, treatment schools or comparison schools were making other changes to discipline policies (or both types of schools were making changes but in different ways), this would prohibit a causal interpretation.

While it is impossible to directly test any unobservable factors, I can test whether observable factors appear to be "affected" by the policy. I estimate regressions as in Equation 5 but predict school-by-year measures of observable traits that may influence our outcome measures. Specifically, I estimate whether the policy "impacted" school-by-year percent of students who are FRL, percent of students who are special education, percent of students who are Black, percent of students who are Hispanic, and school enrollment size. The explanatory variables are the same, except that, for each model, the dependent variable (or a derivation there of) is not included as an explanatory variable. For example, in the model predicting school enrollment, I remove the log of enrollment as an explanatory variable.

These models provide evidence on whether factors that may determine important school outcomes vary along with the policy change in ways that would confound the estimates of the policy-related change. If these models indicate the policy "impacts" these measures, I would have evidence against the identifying assumption. However, passing this test is not sufficient proof that unobservable characteristics are not changing around the time of the policy.

Table 14

Robustness Check: Policy-related Change in Other Outcomes

	School % Special Education			School % FRL			School % Special Education			School % FRL		
Panel A: T_s = used OSS for truancy at least once in 2012-13												
$T_s \times Policy_t$	-0.003	-0.002	-0.002	0.001	-0.001	-0.003	-0.002	-0.001	-0.000	-0.009	-0.008	-0.010
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)
$T_s \times$ Years since policy	0.001	0.001	0.001	0.003	0.004	0.004	0.003	0.002	0.002	-0.006	-0.006	-0.006
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
Total effect by 2015-16	0.001	0.001	0.001	0.010	0.010	0.008	0.006	0.006	0.007	-0.028 *	-0.027 *	-0.028 **
	(0.004)	(0.004)	(0.004)	(0.014)	(0.014)	(0.013)	(0.006)	(0.006)	(0.006)	(0.015)	(0.014)	(0.014)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-squared	0.0593	0.0647	0.0731	0.155	0.183	0.210	0.0718	0.0882	0.0938	0.220	0.222	0.244
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13												
$T_s \times Policy_t$	-0.006	-0.005	-0.005	-0.002	-0.005	-0.007	-0.003	-0.001	-0.000	-0.015	-0.014	-0.018 *
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)	(0.010)	(0.010)	(0.010)
$T_s \times$ Years since policy	0.002	0.002	0.002	0.005	0.004	0.006	0.003	0.002	0.002	-0.008	-0.007	-0.005
	(0.003)	(0.003)	(0.003)	(0.006)	(0.007)	(0.006)	(0.005)	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)
Total effect by 2015-16 [a]	-0.000	0.000	-0.000	0.014	0.008	0.009	0.006	0.006	0.006	-0.039	-0.036	-0.033
	(0.009)	(0.009)	(0.009)	(0.020)	(0.019)	(0.018)	(0.015)	(0.014)	(0.014)	(0.024)	(0.024)	(0.022)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-Squared	0.0686	0.0733	0.0814	0.155	0.183	0.210	0.0708	0.0873	0.0932	0.218	0.220	0.241
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Dependent variables are on a 0-1 scale. Sch. chars. predicting baseline compliance include: log of enrollment, percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group). However, percent special education is not a covariate in the model predicting percent special education.
[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 14, Cont'd

Robustness Check: Policy-related Change in Other Outcomes, Cont'd

	School % Black			School % Hispanic			School % Black			School % Hispanic		
Panel A: T_s = used OSS for truancy at least once in 2012-13												
$T_s \times Policy_t$	-0.001	-0.001	0.001	0.003	0.004	0.004 *	0.003	0.004	0.006 *	0.003	0.003	0.004
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$T_s \times$ Years since policy	-0.002	-0.003	-0.003 *	-0.000	-0.000	-0.001	-0.004 *	-0.005 **	-0.004 **	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Total effect by 2015-16	-0.008 *	-0.008 *	-0.007	0.002	0.002	0.002	-0.009	-0.011 *	-0.007	0.001	0.001	0.003
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-squared	0.004	0.021	0.101	0.323	0.330	0.353	0.006	0.045	0.132	0.343	0.348	0.373
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13												
$T_s \times Policy_t$	-0.003	-0.002	0.001	0.003	0.004	0.005	0.007	0.007	0.011 *	0.007	0.008	0.010
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
$T_s \times$ Years since policy	-0.004	-0.004	-0.005	-0.001	-0.002	-0.002	-0.009 *	-0.009 **	-0.010 **	-0.003	-0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)
Total effect by 2015-16 [a]	-0.014	-0.013	-0.013	-0.001	-0.001	-0.002	-0.020	-0.021	-0.020	-0.003	-0.002	-0.002
	(0.010)	(0.010)	(0.009)	(0.005)	(0.005)	(0.005)	(0.014)	(0.014)	(0.014)	(0.008)	(0.008)	(0.008)
Number of Schools	374	374	374	374	374	374	228	228	228	228	228	228
Observations	3,205	3,205	3,205	3,205	3,205	3,205	1,959	1,959	1,959	1,959	1,959	1,959
Adjusted R-Squared	0.004	0.021	0.100	0.325	0.332	0.353	0.012	0.050	0.139	0.346	0.352	0.377
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Dependent variables are on a 0-1 scale. Sch. chars. predicting baseline compliance include: log of enrollment, percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group). However, dependent variables (or their corresponding variables) are not included as covariates.
[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 14, Cont'd

Robustness Check: Policy-related Change in Other Outcomes, Cont'd

	School Enrollment			School Enrollment		
Panel A: T_s = used OSS for truancy at least once in 2012-13						
$T_s \times Policy_t$	5.454	4.048	2.729	8.646	7.446	6.504
	(9.068)	(8.932)	(8.785)	(11.38)	(11.38)	(11.22)
$T_s \times$ Years since policy	1.414	2.250	2.363	-1.077	-0.301	0.505
	(5.500)	(5.533)	(5.513)	(5.985)	(6.013)	(6.005)
Total effect by 2015-16	9.696	10.800	9.818	5.416	6.544	8.018
	(14.63)	(14.81)	(14.67)	(18.19)	(18.35)	(18.22)
Number of Schools	374	374	374	228	228	228
Observations	3,205	3,205	3,205	1,959	1,959	1,959
Adjusted R-squared	0.013	0.017	0.021	0.019	0.020	0.024
Panel B: T_s = percent of truancy cases resulting in OSS in 2012-13						
$T_s \times Policy_t$	-2.857	-4.475	-6.226	-0.680	-1.869	-4.161
	(9.410)	(9.352)	(9.294)	(13.72)	(13.78)	(13.26)
$T_s \times$ Years since policy	7.469	7.729	8.306	9.822	10.690	13.020 *
	(6.426)	(6.433)	(6.401)	(7.760)	(7.812)	(7.766)
Total effect by 2015-16 [a]	19.550	18.710	18.690	28.790	30.200	34.890
	(17.89)	(17.83)	(17.63)	(22.94)	(23.06)	(23.03)
Number of Schools	374	374	374	228	228	228
Observations	3,205	3,205	3,205	1,959	1,959	1,959
Adjusted R-Squared	0.017	0.020	0.024	0.019	0.021	0.025
Sch. chars. predicting baseline compliance	Y	Y	Y	Y	Y	Y
Seven cons. types (freq. per 100 stud.)		Y	Y		Y	Y
Full controls for school-level chars.			Y			Y

Note. Each panel-column combination represents a separate regression. All models include school fixed effects. Standard errors are clustered at the school level. Dependent variables are on a 0-1 scale. Sch. chars. predicting baseline compliance include: percent Black, region indicators, and truancy frequency (incidents per 100 students). Full controls for school-level chars. add: percent special education, FRL, Hispanic, and other minority, as well as indicators for middle and high schools (with elementary schools as the omitted group).

[a] Total "effect" by 2015-16 represents the "effect" for schools with 100% OSS for truancy, compared to a school with 0% OSS for truancy, in 2012-13.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

In general, these results (in Table 14) indicate that the policy did not "impact" these characteristics. The exception is slight evidence (only marginally significant) that the school percent black may have been "impacted" as a result. Given that this was the only variable "affected" by the policy, we do not have a clear rejection of the identifying assumption.

Discussion and Conclusions

The state of Arkansas pursued a policy prohibiting the use of OSS for truancy, theoretically in an effort to positively impact students. For example, it is logical that part of the state's goal was to encourage reasonable consequences so that punishments "fit the crime" and ultimately prevent future misbehavior. In general, this policy was not associated with differential changes in important school outcomes such as percent of total student days absent, the percent of students chronically absent, or math and ELA test scores. The only estimated relationship of the policy to one of these four outcomes (percent of days absent) was small, marginally significant, and sensitive to model specification. However, the lack of changes in school-level outcomes may not be very surprising given that in each year, less than 2% of students are written up for truancy, so future research could estimate the impact on individual students instead.

I do find that, after the policy, there was an increase in the use of "other" consequences for truancy in treatment schools, relative to comparison schools. It is unclear what schools are really using within this "other" category, so in order to understand the impact of this policy change, the state needs to increase reporting capabilities and accountability for what consequences or interventions schools are actually using.

There is also suggestive evidence that treatment schools had a slight decline in truancy reports per 100 students, relative to comparison schools, but it is unclear whether the estimated policy-related decline in truancy is a reflection of truancy prevention or a change in reporting practices. In addition, I estimate a policy-related increase in “other” infractions per 100 students, although the pattern of evidence suggests this rise in “other” infractions is not solely due to this policy change. Rather, the findings suggest there may be other (non-policy related) changes in reports of these “other” infractions.

Why did this policy not affect school-level absenteeism or student achievement? First of all, there was much variation in compliance across schools, and the types of schools that the state was likely intending to impact with this policy (schools with high rates of truancy, high proportions of minority students, and frequent use of OSS), are also the types of schools that fail to comply with the policy. A few factors may have contributed to a lack of a clear relationship between the policy and school outcomes: 1) insufficient communication to schools including the reasons for the change, an indication of how schools will be held accountable, and suggested alternatives to using OSS 2) a lack of accountability for adherence to the policy and 3) a lack of capacity or resources for schools to comply. I was unable in this study to empirically test whether these factors were present and influenced the policy-related outcomes, but theoretically, these three items (communication, accountability, and resources) are necessary - but not necessarily sufficient - for producing the intendent effect.

While there was an apparent lack of accountability for compliance with this policy, it is unclear what the ideal form of accountability would be. A set of rules distributed by the Arkansas Department of Education (2012) indicates that the Department “shall monitor compliance” with its rules, and that if schools fail to file their policies with the state, the school district “shall have all state aid funds withheld until such disciplinary policy is filed with the Department of Education.” This set of rules, however, does not explicitly indicate that state funds could be withheld for failure to comply with other parts of the state code, although perhaps it is implied. Certainly, if the Arkansas Department of Education has the right to withhold state funds for failure to comply with this policy, there is no evidence that it has done so or has plans to do so. Rather, it appears the state has taken a very hands-off approach to compliance so far.

Another type of accountability could come from parents taking legal action against the school district, although parents would have to be aware of this law, which is unlikely given that there was apparently very little communication about the policy, and taking legal action could potentially be very costly as well. Legal action against a school could be quite burdensome in some of the communities where schools are failing to comply.

More research should be conducted to understand the impacts of this policy on students. It could be that there are impacts on particular types of students or in particular types of schools, but that there is no apparent overall impact when looking at outcomes in the entire school. Future research could address whether there are impacts only on the truant students, for example, or on the peers of truant students who now remain in school.

Arkansas may need to pursue more in depth, qualitative research to understand how schools are reacting to this policy, and more broadly, what alternatives to OSS schools are using. It is not clear that the desired impact will occur if the only change is a reduction in the use of OSS, without replacing it with other supports for students (Anderson, Ritter, & Zamarro, 2017).

It is also possible that focusing on consequences is not as effective as preventing truancy in the first place, so more work should be done to study the effectiveness of truancy interventions in the state. For example, there are "promising interventions" available for reducing truancy (Sutphen, Ford, & Flaherty, 2010). There are a variety of school-based, court-based, and community-based approaches to handling truancy, and evidence suggests that educators should focus more on preventative and social-based supports, rather than punitive approaches (Smith &

Weilbrunn, 2005). OSS may not be an effective practice for dealing with truant students because schools need to address the underlying reasons behind the absence in order to interrupt the student's decline towards delinquency (Baker et al., 2001). However, large-scale experimental evaluations on the effectiveness of truancy interventions are lacking. Two experimental studies, limited in size and scope, found truancy could be reduced through contingency management including a token economy and group contracts (Brooks, 1975) or through threat of losing public assistance for frequent non-attendance (Jones, Harris, & Finnegan, 2002). However, in one of these studies, attendance declined over the course of the intervention (Jones, Harris, & Finnegan, 2002).

Another school-based approach that focuses on a variety of student behaviors is School-Wide Positive Behavior Interventions and Supports (SWPBIS). SWPBIS includes three tiers of support. Tier I is school-wide, Tier II is targeted, small group support that can focus on study skills, social skills, behavior, attendance, or dropout prevention, and Tier III is individualized support for the students with greatest need. Experimental evidence finds that implementing SWPBIS can improve student perceptions of school safety and test scores in elementary schools (Horner et al., 2009). Experimental studies at the high school level are lacking, but when implemented with fidelity, this framework has been linked to improvement in disciplinary incidents (Flannery, Fenning, Kato, & McIntosh, 2014; Freeman et al., 2015) and attendance (Freeman et al., 2015) in high schools.

In conclusion, this study provides evidence on implementation fidelity and outcomes related to a state-level policy prohibiting the use of OSS as a consequence for truancy, theoretically with the goal of aligning disciplinary consequences to infractions and improving student outcomes. It is a cautionary tale that hints at the inability of states to use high-level policy to influence student discipline in ways that improve student outcomes, at least if there is insufficient communication, accountability, and school resource capacity. The schools that theoretically were the key targets of the policy are found to be the ones that fail to comply even three years after the policy changed, and perhaps as a result, there was no apparent improvement in school-level test scores, attendance, and chronic absenteeism. While this conclusion is troubling, is may not be surprising, as it is often difficult to change human behavior. For example, we might also expect that following a ban on sugary soft drinks, people who previously consumed large amounts of these drinks would likely be the ones that struggle to comply.

Of course, I acknowledge that these three necessary conditions (communication, accountability, and school resource capacity) are not necessarily sufficient to produce improvements in student outcomes, and more work is needed to fully understand the key factors for ideal policy design. Looking forward, it will be interesting to see how schools in Arkansas react to a recent policy change banning OSS and expulsion in kindergarten through fifth grade except in cases of "physical risk" or "serious disruption that cannot be addressed through other means" (Arkansas Act 1059, 2017).

References

Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record, 103*(5), 760-822.

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.

Anderson, K. P., & Ritter, G. W. (November 10, 2016). *Assessing the Effect of Exclusionary Discipline on Student Academic Outcomes.* Presentation for the Arkansas State Board of Education. Fayetteville, AR: Office for Education Policy, University of Arkansas. Retrieved from http://adecm.arkansas.gov/Attachments/COM-17-036--State_Board_Discipline_Pres_Nov_10_Final.pdf

Anderson, K. P., & Ritter, G. W. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. state. *Education Policy Analysis Archives, 25*(49). Retrieved from http://dx.doi.org/10.14507/epaa.25.2787

Anderson, K. P., Ritter, G. W., & Zamarro, G. (2017). *Understanding a vicious cycle: Do out-of-school suspensions impact student test scores?* EDRE Working Paper 2017-09. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944346

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*(2014), 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 1059. (2017). An Act to Amend Provisions of Title 6 of the Arkansas Code Concerning Discipline of Students in Public Schools; And for Other Purposes. Retrieved from http://www.arkleg.state.ar.us/assembly/2017/2017R/Acts/ACT1059.pdf

Arkansas Code § 6-18-507. Suspension – Expulsion. Retrieved from http://law.justia.com/codes/arkansas/2016/title-6/subtitle-2/chapter-18/subchapter-5/section-6-18-507/

Arkansas Department of Education (January 10, 2017). *Student discipline report, reminder and committee information.* ADE Commissioner’s Memo. Retrieved from http://adecm.arkansas.gov/ViewApprovedMemo.aspx?Id=2127

Baker, M. L., Sigmon, J. N., & Nugent, M. E. (2001). *Truancy reduction: Keeping students in school.* Washington, D.C.: Department of Justice, Office of Juvenile Justice and Delinquency Prevention. Retrieved from https://www.ncjrs.gov/pdffiles1/ojjdp/188947.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., & Byrnes, V. (2012). *The importance of being in school: A report on absenteeism in the nation's public schools.* Baltimore, MD: Johns Hopkins University. Retrieved from http://new.every1graduates.org/wp-content/uploads/2012/05/FINALChronicAbsenteeismReport_May16.pdf

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.

Bell, A. J., Rosen, L. A., and Dynlacht, D. (1994). Truancy intervention. *The Journal of Research and Development in Education, 57*(3), 203-211.

Bloom, H. S., & Riccio, J. A. (2005). Using place-based random assignment and comparative interrupted time-series analysis to evaluate the Jobs-Plus employment program for public housing residents. *The ANNALS of the American Academy of Political and Social Science, 599*(1), 19-51.

Brooks, B. D. (1975). Contingency management as a means of reducing school truancy. *Education, 95*(3), 206-211.

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.

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/

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *The Annals of Mathematical Statistics, 31*(2), 385-391.

Dryfoos, J. G. (1990). *Adolescents at Risk: Prevalence and Prevention.* New York, NY: Oxford University Press.

Epstein, J. L., & Sheldon, S. B. (2002). Present and accounted for: Improving student attendance through family and community involvement. *Journal of Educational Research, 95*(5), 308-318.

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.* New York, NY: Council of State Governments Justice Center. Retrieved from https://csgjusticecenter.org/wp-content/uploads/2012/08/Breaking_Schools_Rules_Report_Final.pdf

Flannery, K. B., Fenning, P., Kato, M. 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.

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2015). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions, 18*(1).

Garry, E. M. (1996). *Truancy: First step to a lifetime of problems.* Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. Retrieved from https://www.ncjrs.gov/pdffiles/truncy.pdf

Geis, G., & Binder, A. (1991). Sins of their children: Parental responsibility for juvenile delinquency. *Notre Dame Journal of Law, Ethics, and Public Policy, 5*(2), 303–322.

Gerety, R. M. (October 25, 2016). School 'suspensions' continue despite Miami-Dade's no suspension policy. *Miami Herald.* Retrieved from http://www.miamiherald.com/news/local/education/article110343972.html

Gottfried, M. A. (2015). Chronic absenteeism in the classroom context: Effects on achievement. *Urban Education.* Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0042085915618709

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*(4),455 – 475.

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. Retrieved from https://doi.org/10.1177/1098300709332067

Huizinga, D., Loeber, R., & Thornberry, T. (1995). *Urban delinquency and substance abuse: Initial findings.* Washington, D.C.: U.S. Department of Justice, Office of Justice

Programs, Office of Juvenile Justice and Delinquency Prevention. Retrieved from https://www.ncjrs.gov/pdffiles/urdel.pdf

Huizinga, D., Loeber, R., Thornberry, T. P., & Cothern, L. (2000). Co-occurrence of delinquency and other problem behaviors. *Juvenile Justice Bulletin.* Retrieved from https://pdfs.semanticscholar.org/5622/95f8d85bd7bcbda8d0c8f068d0db10213c0a.pdf

Ingersoll, S., & LeBoeuf, D. (1997). *Reaching out to youth out of the education mainstream.* Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. Retrieved from https://eric.ed.gov/?id=ED408667

Jones, L. P., Harris, R., & Finnegan, D. (2002). School Attendance Demonstration Project: An evaluation of a program to motivate public assistance teens to attend and complete school in an urban school district. *Research on Social Work Practice, 12*(2), 222-237.

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*(4), 217 – 226.

Kube, B., & Ratigan, G. (1992). Does your school have a clue?: Putting the attendance policy to the test. *The Clearing House*, *65*(6), 348–350.

Lee, M. I., & Miltenberger, R. G. (1996). School refusal behavior: Classification, assessment, and treatment issues. *Education and Treatment of Children, 19*(4), 474-486.

Locker, L. A. (2001). *Some "modern" statistics: A primer and demonstration.* Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA. Retrieved from http://files.eric.ed.gov/fulltext/ED449221.pdf

Losen, D., Hodson, C., Keith, M., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles, CA: The Civil Rights Project, University of California – Los Angeles. Retrieved from https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/are-we-closing-the-school-discipline-gap

Losen, D., & Skiba, R. (2010). *Suspended education: Urban middle schools in crisis.* Los Angeles, CA: The Civil Rights Project, University of California – Los Angeles. Retrieved from http://www.saferfoundation.org/files/documents/Suspended_Education.pdf

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).

Nauer, K., Mader, N., Robinson, G., & Jacobs, T. (2014). *A better picture of poverty: What* *chronic absenteeism and risk load reveal about NYC's lowest income elementary schools.* New York, NY: Center for New York City Affairs. Retrieved from http://www.attendanceworks.org/wordpress/wp-content/uploads/2014/11/BetterPictureofPoverty_PA_FINAL_001.pdf

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.

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

O'Donnell, P. (2016). *Ohio will "decriminalize" truancy under new law.* Cleveland.com Retrieved from http://www.cleveland.com/metro/index.ssf/2016/12/ohio_will_decriminalize_truanc.html

Pell, C. E. (2000). Pre-offense monitoring of potential juvenile offenders: An examination of the Los Angeles County Probation Department's novel solution to the interrelated problems of truancy and (juvenile) crime. *Southern California Law Review, 73*, 879–925.

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,* 2003(99), 17-33.

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.

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. Retrieved from http://www.indiana.edu/~atlantic/wp-content/uploads/2014/03/Academic-Cost-of-School-Discipline.pdf

Ritter, G. W., & Anderson, K. P. (2016). *School discipline and inequalities in Arkansas.* Presentation for the Arkansas State Board of Education. Fayetteville, AR: Office for Education Policy, University of Arkansas. Retrieved from https://posting.arktimes.com/media/pdf/state_board_presentation.pdf

Rohrman, D. (1993). Combating truancy in our schools: A community effort. *National Association of Secondary School Principals Bulletin, 76*(549), 40-51.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

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.* Chicago, IL: The University of Chicago Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/publications/suspending-chicagos-students-differences-discipline-practices-across-schools

Siegel, L. (2002). *Juvenile delinquency: The core*. Belmont, CA: Wadsworth.

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 dolor of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review, 34*(4), 317-342.

Skiba, R. J., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores.* Bloomington, IN: Center for Evaluation and Education Policy, Indiana University.

Smink, J., & Heilbrunn, J. Z. (2005). *The legal and economic implications of truancy*. Clemson, SC: National Dropout Prevention Center/Network. Retrieved January 15, 2007, from http://files.eric.ed.gov/fulltext/ED491290.pdf

Somers, M., Zhu, P., Jacob, R., & Bloom, H. (September 2013). *The validity and precision of the comparative interrupted time series design and the difference-in-difference design in educational evaluation.* MDRC Working Paper on Research Methodology. Retrieved from https://www.mdrc.org/sites/default/files/validity_precision_comparative_interrupted_time_series_design.pdf

Stearns, E., & Glennie, E. J. (2006). When and why dropouts leave school. (2006). *Youth & Society, 38*(1), 29-57.

Steinberg, M. P., & Lacoe, J. (2017). What do we know about school discipline reform? *Education Next, 17*(1), 44-52. Retrieved from http://educationnext.org/what-do-we-know-about-school-discipline-reform-suspensions-expulsions/

Stevens, W. D., Sartain, L., Allensworth, E. M., and Levenstein, R. *with* Guiltinan, S., Mader, N., Huynh, M. H., & Porter, S. (March 2015). *Discipline practices in Chicago schools: trends in the use of suspensions and arrests.* Chicago, IL: University of Chicago Consortium of Chicago School Research. Retrieved from https://consortium.uchicago.edu/sites/default/files/publications/Discipline%20Report.pdf

Sutphen, R. D., Ford, J. P., & Flaherty, C. (2010). Truancy interventions: A review of the research literature. *Research on Social Work Practice, 20*(2), 161-171.

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

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

U.S. Department of Education. (June 7 2016). *Chronic absenteeism in the nation's schools.* Retrieved from https://www2.ed.gov/datastory/chronicabsenteeism.html#intro

Wald, J., & Kurlaender, M. (2003). Connected in Seattle? An exploratory study of student perceptions of discipline and attachments to teachers. *New Directions for Youth Development, 2003*(99), 35-54.

Wong, M., Cook, T. D., & Steiner, P. M. (2011). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series.* Working Paper WP-09-11. Institute for Policy Research, Northwestern University. Retrieved from http://www.ipr.northwestern.edu/publications/docs/workingpapers/2009/IPR-WP-09-11.pdf

Appendix

Appendix A – Arkansas Timeline of Events

March 11, 2013: Arkansas legislature passes Act 1329, "An Act to Evaluate the Impact of School Discipline on Student Achievement; and for Other Purposes," which is codified as §6-18-507 (b). This law states "The board of directors of a school district may suspend or expel any student from school for violation of the school district's written discipline policies, except that a school district shall not use out-of-school suspension as a discipline measure for truancy."

Feb 11, 2016: University of Arkansas researchers present a report at the Arkansas State Board of Education (SBE) (Ritter & Anderson, 2016), indicating that as of 2014-15, 9% of truancy cases in the state (and 14% of truancy cases for African American students) resulted in OSS. The report also reminds the SBE that schools shall not use OSS as a consequence for truancy according to Act 1329 of 2013, but that over 100 districts were still doing this as of 2014-15. The report notes that 29 schools used OSS for 100% of truancy cases during 2014-15.

November 10, 2016: University of Arkansas researchers present to the SBE a report on the impact of exclusionary discipline on students (Anderson & Ritter, 2016).

January 10, 2017: The Arkansas Department of Education (ADE, 2017) distributed Memo Number COM-17-036 to the attention of superintendents and principals stating that during the November 10 meeting, "State Board members requested the department remind districts that Ark. Code Ann. § 6-18-507(b) prohibits school districts from imposing out-of-school suspension as a discipline measure for truancy." Attached to this memo was a copy of the November 10, 2016 presentation (Anderson & Ritter, 2016).

Chapter 5:

Conclusion

The three studies within this dissertation have helped fill several distinct gaps in the evidence base regarding the implementation and outcomes of student discipline policies. Yet, there are limitations to these studies, and there is still a need for more, particularly causal, evidence on these topics. In this concluding chapter, I reiterate some of the key findings from each paper, the limitations surrounding the data and analytic approach, and what policy implications follow as a result.

Summary of Findings

In the first paper (Chapter 2), Dr. Gary Ritter and I find that, at least in Arkansas, racial disparities in the use of exclusionary discipline, even after accounting for the nature and frequency of disciplinary actions, as well as a variety of student characteristics, are quite large and evident primarily across schools rather than within schools. This work also shows that the racial composition of schools is a stronger driver of the severity of punishments than is free-and reduced-price lunch (FRL) status. These findings suggest that policies designed to encourage fair administration of discipline may need to focus on particular schools and contexts.

Findings from Chapters 3 and 4 suggest it is not necessarily the case that reducing or eliminating exclusionary discipline will improve student- or school-level outcomes. In Chapter 3, Dr. Gary Ritter, Dr. Gema Zamarro, and I attempt the first causal estimate of the impact of out-of-school suspension (OSS) on student test scores in math and English Language Arts (ELA). While a vast amount of prior literature has shown troubling correlations between exclusionary discipline and a variety of negative student outcomes, we find virtually zero evidence of a negative causal impact. Rather, the impacts are either very small positive impacts or null. This indicates that, without other interventions or supports, a reduction in OSS may not have the intended effect of improving student achievement, at least for the suspended students themselves.

The third paper focuses on the implementation and outcomes from a state-level policy prohibiting the use of OSS as a disciplinary consequence for truancy. There was a partial decline in the use of OSS as a consequence for truancy from 13.8% in 2012-13, the year the policy was passed, to 8.7% in 2015-16, but this was also accompanied by a large increase in the use of "other" non-specified consequences for truancy, which makes it difficult to understand or interpret what this means for students. Further, I find that the schools that the policy likely intended to target, those with high truancy rates and high OSS rates in general, were the types of schools that still fail to fully comply after three years. I speculate that this implementation failure is due, in large part, to three key factors: 1) a lack of communication from the state to school leaders about the importance of the policy change and how it would be carried out, 2) failure to hold schools accountable for this change, and 3) a lack of information or resources to carry out other alternatives to OSS. Perhaps as a result, this lack of implementation was also accompanied by no measurable policy-related change in school outcomes such as average math and ELA test scores, student attendance, and chronic absenteeism. There was a small policy-related reduction in reported truancy rates and a somewhat larger policy-related increase in "other" infraction rates, but this latter finding was sensitive to treatment definition, and thus I conclude it is likely not entirely driven by the policy.

Study Limitations

While these three papers fill gaps in the knowledge base, there are several limitations to the generalizability of these findings and the implications that follow. First, all three of these studies rely on administrative data from a single state. The disciplinary measures in all three of these papers are only the disciplinary infractions and consequences that were reported by schools. Thus, in Chapter 2, in which my co-author and I estimate the across- and within-school racial disproportionalities in exclusionary discipline, we are not accounting for any racial disparities in how student behaviors are actually observed and/or reported. For example, if certain students misbehave but never receive an office referral, we do not see that in our dataset. In addition, even if an infraction is reported to the office, it is not guaranteed that the school will record the incident in a way that is officially reported to the state. While the within-school analysis utilizes school fixed effects, and thus controls for time-invariant school characteristics which may include reporting practices, the across-school estimates may be influenced by differences in how schools or school districts report their disciplinary infractions.

This data limitation is also a factor in Chapter 3, in which we estimate the impact of OSS on student test scores. To account for differences in reporting across school districts, we utilize district fixed effects. While this helps account for constant district-level differences, if schools within districts report disciplinary incidents very differently, this would affect the interpretation of these results.

In Chapter 4, to account for school-level differences in how incidents are reported over time, I include lagged measures of reported disciplinary incidents in the descriptive analyses predicting whether a school used OSS as a consequence for truancy in the baseline year and predicting compliance with the policy in future years. When I estimate the policy-related change in school-level outcomes, the fullest model specifications also include these disciplinary measures for each time period, to account for differential trends in disciplinary practices in treatment and comparison schools. To further account for time-invariant school characteristics, these models also utilize school fixed effects.

Each paper has its own additional limitations. A key limitation in Chapter 2, which measured racial disproportionalities in exclusionary discipline, relates to how we can interpret the results and use them to inform policy. While we advocate for data transparency as a way to make key stakeholders aware of disciplinary outcomes in their local public schools and perhaps create advocates for change, there is potential for unintended consequences if a rise in public reporting encourages school personnel to simply under-report or code disciplinary infractions in vague or non-transparent ways. In addition, even if student discipline data are publicly reported, whether high numbers of disciplinary referrals are a bad thing or a good thing is likely very context dependent, making comparisons across schools difficult.

The key limitations of Chapter 3, which assesses the impact of OSS on student test scores, relate to the outcome measures available, as well as the key identifying assumptions for causal inference. This study only looks at the impact of exclusionary discipline on student math and ELA test scores, so it could be that there is a causal impact on other important outcomes that we are not able to measure. The estimates are based only on within-student variation for students who have varied amounts of exposure to OSS over time. While this may seem like a limitation, this was an intentional choice because it allows us to control for student time-invariant unobservable characteristics, and additionally, these students comprise what is likely the most policy-relevant population. A remaining concern for a causal interpretation of these results is that there may be time-varying shocks to students that affect both discipline and student test scores over time. For example, if there are unobservable events in a student's life such as divorce or a death in the family, this could constitute a time-persistent shock. Another limitation of Chapter 3 is that it focuses only on the outcomes for students suspended rather than the systemic impacts on the entire school.

Finally, Chapter 4, which assesses the implementation of and school-level outcomes following a state-level policy, utilizes data from a single state where all schools were technically subject to the policy change. Thus, a key limitation relates to the selection of a comparison group. Ideally, a very similar state, but one not subject to the policy change, could have been used as a comparison group. However, using the data available, I created a set of treatment and comparison schools in the state, assuming that schools using OSS as a consequence for truancy in the baseline year would theoretically be affected, while others would not. School participation in these groups is clearly not exogenous, so the comparison schools are likely not a good counterfactual for what would have happened in the treatment schools in the absence of the policy. Thus, it is unlikely that the estimated relationships between the policy and school-level outcomes are causal. A further limitation of Chapter 4 is that I cannot directly test my hypothesis that a lack of implementation fidelity was due, in part, to three factors: 1) a lack of communication from the state, 2), a lack of accountability, and 3) a lack of information about reasonable and effective alternatives to exclusionary discipline.

Important Lessons for Policy

Regardless of the limitations, these three studies still make meaningful contributions to the policy debate surrounding the use of exclusionary discipline. In fact, combining the results of all three indicates the type of policies or reforms that states may want to pursue. The persistence of racial gaps in exclusionary discipline, coupled with the failure of a state-level policy to eliminate exclusionary discipline as intended and a lack of causal evidence that this exclusionary discipline actually harms student achievement, provides support for the idea that simply aiming to reduce OSS, without providing additional interventions or supports for at-risk students, may be ineffective. In addition, given that I find racial disparities are primarily driven by differences across schools and that high-discipline schools tend to struggle the most with compliance following a state-level policy change, these findings suggest the need for a more targeted approach to reform.

Further, the findings from these three studies suggest a need for more careful use of evidence in the debate over discipline reform. Some researchers have decried the recent efforts as ineffective and potentially harmful for teacher morale (Eden, 2017; Loveless, 2017), but others use the troubling disproportionalities as their battle cry for reform (Losen, Hodson, Keith, Morrison, & Belway, 2015). Other commentators are seeing the need for a "third way that integrates a school's approach to discipline with high-quality, culturally competent school cultures, teaching and learning practices, and student supports," while also building school capacity to actually achieve the intended outcomes (Anderson, 2017).

The policy implications from these three papers suggest the need for a middle ground. The evidence on racial disproportionalities in Chapter 2 highlights the disparate exposure to harsh and exclusionary discipline, but is also useful for helping design state policies, as it suggests that efforts need to focus on between-school differences, and thus must take into account the types of schools that minority students are residentially assigned to attend. In addition, there may be a middle ground needed in data transparency and public reporting. Schools likely need to be able to compare how their discipline outcomes stack up with other similar (or dissimilar) schools in the state if they are to identify whether there is a problem or a need for change. However, Campbell's Law warns us that we also should be concerned about the potential for unintended consequences if we focus too much on any particular measure (Campbell, 1979).

The results of Chapter 3 suggest what we should expect to happen to student test scores if OSS is reduced: not very much. Here, our estimate of the impact of exclusionary discipline on student test scores attempts to isolate the impact of the suspension itself by carefully controlling for the frequency and type of behaviors a student is reportedly engaging in. We find a null to slightly positive impact of OSS on student test scores, depending on the model and analytic sample, which indicates that reductions in OSS, without any additional supports, interventions, or resources, are not expected to improve student achievement, as measured by test scores. This result does not mean that we should not care about reducing the misbehavior that led to the suspension to begin with. In fact, it may be more beneficial to focus on interventions that build cultural competency for teachers and staff, improve school climate and safety, and generally prevent disruptive or dangerous behaviors in schools.

This same implication follows from the findings of Chapter 4, in which a high-level policy reform seeking to eliminate the use of OSS for truancy not only failed to improve student outcomes, but also was not implemented with fidelity. In addition, we learn from Chapter 4 that implementation may be low, and outcomes will likely not be as intended, if policies are implemented in a way that does not encourage real change. It is a cautionary tale about what we might expect from state-level policies that prohibit certain behaviors, without additional communication, accountability, or supports.

Looking forward, it is unclear how Arkansas and other states will focus their future policy initiatives related to student discipline. Under the Every Student Succeeds Act (ESSA) of 2015, state education agencies are now required by federal law to describe in their Title I state plan how they plan to support local education agencies (LEAs) in reducing "the overuse of discipline practices that remove students from the classroom" as well as "the use of aversive behavioral interventions that compromise student health and safety" (Sec. 1111(g)(1)(C)). Further, under ESSA, schools are encouraged to implement a "schoolwide tiered model to prevent and address problem behavior" which could reasonably include frameworks such as School-wide Positive Behavioral Interventions and Supports (SWPBIS) (Sec. 1114(b)(7)(A)(iii)(III)).

Arkansas has described, in the March 2017 draft of its ESSA state plan, efforts to improve the usefulness of discipline data in schools and the development of SWPBIS modules and training materials (Arkansas Department of Education, 2017). In addition, the state has passed another state-level policy to reduce reliance on exclusionary discipline. In April 2017, Arkansas Act 1059, "An Act to Amend Provisions of Title 6 of the Arkansas Code Concerning Discipline of Students in Public Schools; And for Other Purposes" further limited the use of out-of-school suspension in Arkansas public school districts. In particular, the law states: "the school district shall not use out-of-school suspension or expulsion for a student in kindergarten through grade five (K-5) except in cases when a student's behavior: poses a physical risk to himself or herself or to others; or causes a serious disruption that cannot be addressed through other means" (Arkansas Code § 6-18-507). As with Act 1329, which banned OSS for truancy, there is little guidance about the reasons for this change, the accountability system to ensure compliance, or guidance on disciplinary alternatives to meet this requirement. As of August 2017, there is no indication that the Arkansas Department of Education has distributed any further guidance on these changes. As a result, I am not very hopeful about what impact this will actually have on Arkansas students. While certain high-functioning school districts with excellent leadership will likely find and use the resources available, the types of schools most in need of reform are likely to continue to be the schools left behind without the knowledge or resources about available alternatives.

As state and local education agencies continue to design policy, three things are abundantly clear. First, policy needs to be informed by facts and by causal evidence whenever possible, and asking the right questions is an important first step in the right direction. Secondly, the impact of high-level policies may be limited, and unintended consequences are a real possibility, so policy design, implementation, and assessment of progress are critically important. Finally, removing ideology and moving towards a middle ground based on facts and evidence is critical if we hope to improve outcomes for America's students.

References

Arkansas Code § 6-18-507. (2017). Arkansas Act 1059. An Act to Amend Provisions of Title 6 of the Arkansas Code Concerning Discipline of Students in Public Schools; And for Other Purposes. Retrieved from http://www.arkleg.state.ar.us/assembly/2017/2017R/Acts/ACT1059.pdf

Anderson, C. (March 26, 2017). *The left-right school discipline debate misses the point. We need a third way*. The 74 Million. Retrieved from https://www.the74million.org/article/cami-anderson-the-left-right-school-discipline-debate-misses-the-point-we-need-a-third-way

Arkansas Department of Education. (March 1, 2012). *Guidelines for the development, review and revision of school district student discipline and school safety policies.* Retrieved from http://www.arkansased.gov/public/userfiles/Legal/Legal-Current%20Rules/ade_253-discipline%20and%20safety%20policies-March%202012.pdf

Arkansas Department of Education. (2017). The Arkansas Every Student Succeeds Act Draft Consolidated State Plan. Retrieved from https://www.arkansasessa.org/

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67-90.

Eden, M. (2017). *School discipline reform and disorder: Evidence from New York City Public Schools, 2012-16.* Manhattan Institute. Retrieved from https://www.manhattan-institute.org/html/school-discipline-reform-and-disorder-evidence-nyc-schools-10103.html

Every Student Succeeds Act of 2015. S.1177. (114th Congress). Retrieved from https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf

Losen, D., Hodson, C., Keith, M., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles, CA: The Civil Rights Project, University of California – Los Angeles. Retrieved from https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/are-we-closing-the-school-discipline-gap

Loveless, T. (2017). *The 2017 Brown Center Report on American Education. How well are American Students Learning?* Washington, DC: Brookings Institution. Retrieved from https://www.brookings.edu/multi-chapter-report/the-2017-brown-center-report-on-american-education/

Appendix A – Institutional Review Board Approval Letter

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

February 27, 2017

MEMORANDUM

TO:	Sarah McKenzie Kaitlin Anderson Sivan Tuchman Malachi Nichols Heidi Holmes Elise Swanson Gary Ritter
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT CONTINUATION
IRB Protocol #:	15-02-511
Protocol Title:	*Arkansas School Discipline Study*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Previous Approval Period:	Start Date: 03/02/2015 Expiration Date: 03/01/2017
New Expiration Date:	03/01/2018

Your request to extend the referenced protocol has been approved by the IRB. If at the end of this period you wish to continue the project, you must submit a request using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. Failure to obtain approval for a continuation on or prior to this new expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

This protocol has been approved for 2,000,000 total participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu

The University of Arkansas is an equal opportunity/affirmative action institution.