ENSURING SUPPORTIVE EDUCATIONAL ENVIRONMENTS FOR MARGINALIZED STUDENTS

A Dissertation

Presented to

The Faculty of Education and Human Development

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree Doctor of Philosophy

---

by Todd Madison Hall

May 2025

UNIVERSITY of VIRGINIA | SCHOOL of EDUCATION and HUMAN DEVELOPMENT

Doctoral Dissertation Approval Form

Hall, Todd, M	tmh4xa@virginia.edu	[redacted]
Name* **Last, First, MI*	**UVA Computing ID**	[redacted]
03/05/2025	Educ (PhD) Ed Policy Studies	Signed by: Todd Hall A609C39E3409E4E2...
Defense Date	**Degree Program**	**Student Signature**

This doctoral dissertation has been approved by the Graduate Faculty of the School of Education and Human Development in partial fulfillment for the degree of Doctor of Philosophy.

Approved Title of Doctoral Dissertation*

Ensuring Supportive Educational Environments for Marginalized Students

**Please ensure the title matches the final approved title on the dissertation document exactly.*
**Please follow appropriate approved title format. Here is a helpful tool, an APA title case converter.*

Approved Dissertation Committee

Daphna Bassok **Dissertation Chair**	Education Leadership, Foundations, & Policy Education and Human Development, University of Virginia	Signed by: Daphna Bassok F4805C8D3A1A45E... 4/4/2025
Co-Chair **(If applicable.)**		
Brendan Bartanen **Committee Member**	Education Leadership, Foundations, & Policy Education and Human Development, University of Virginia	Signed by: Brendan Bartanen [redacted] 4/3/2025
Beth Schueler **Committee Member**	Education Leadership, Foundations, & Policy Education and Human Development, University of Virginia	Signed by: Beth Schueler [redacted] 4/2/2025
Vivian Wong **Committee Member**	Education Leadership, Foundations, & Policy Education and Human Development, University of Virginia	DocuSigned by: Vivian Wong [redacted] 4/2/2025
Committee Member		
Committee Member		
Committee Member		

ACKNOWLEDGEMENTS

Completing my doctoral degree would not have been possible on my own. First, I want to thank God. Through no merit of my own, God provided the clarity and resources to pursue this PhD program, gave me a host family in Victory Church before I set foot in Charlottesville, and sustained me through challenging times. To Him be all honor and glory and praise.

I am exceedingly grateful for the EdPolicyWorks community, which has nurtured my development as a researcher and scholar. My dissertation committee—Daphna Bassok, Vivian Wong, Beth Schueler, and Brendan Bartanen—offered critical feedback and support. As my primary advisor, Daphna has tirelessly given wise counsel, nurtured my potential, and advocated for me over the last five years. I thank Allison Atteberry, Jim Wyckoff, Julie Cohen, and Wintre Foxworth Johnson for insightful comments in seminars and on papers. The EPW graduate students have further enriched my doctoral experience, making it easy to be a cohort of one. I also thank several EPW alumni—Justin Doromal, Kylie Anglin, Arielle Boguslav, Brian Kim, Walter Herring, and Anandita Krishnamachari—who I have had the pleasure of working with or informational interviewing. A key mentor and collaborator, Justin Doromal met with me monthly to answer questions about the hidden curriculum and post-PhD life.

Outside of UVA, several researchers have invested in my development. Jon Valant, Laura Bellows, Jane Lincove, and Lucy Sorenson listened to my early thoughts on studying Louisiana's ban on corporal punishment for students with disabilities. Their encouragement at an early stage helped lay the foundation for Chapter 2. As a discussant and a recurring friendly face at research conferences, Christopher Cleveland has offered constructive feedback on this research. Naniette Coleman, organizer of the Summer Institute in Computational Social Science (SICCS) Howard-Mathematica, affirmed my potential as a scholar, helping form (and fund) the research team behind Chapter 3. Space does not allow me to list every researcher who contributed to my growth. I am grateful for the collective generosity of the Association for Education Finance and Policy Scholars of Color and Graduate Student Communities.

I thank the Institute of Education Sciences, the National Science Foundation, the American Educational Research Association, SICSS Howard-Mathematica, the Social Science Research Council, and the Russell Sage Foundation for generous financial support at various stages of the PhD program. This dissertation's content does not represent the views of these organizations.

I thank my parents for inculcating the value of education and providing unwavering support for my professional pursuits. When my brother and I were young, my father would always tell us that we could be anything if we worked hard. He said that education is something no one could ever take from us. My mother made an "executive decision" to send me to McNair Academic High School, a magnet school in Jersey City. She warned me that I had better pass the entrance exam because she knew I could; not being accepted would be considered an act of rebellion. McNair set me on a path to Williams College. There, opportunities to conduct research under Lucie Schmidt and Lara Shore-Sheppard set me on a path to pursuing a PhD.

Finally, to my wife, Shannon: being your husband is the honor of honors. When I doubted, you had faith. When I got COVID, you took care of me. When I had to work late, you gave me space and caught up with your friends, especially Emily. You left me encouraging Post-it notes—"You got this!"—at just the right times. You proofread my resumes and cover letters for job applications. You also made sure I took some weekends off to rest and recover.

TABLE OF CONTENTS

DISSERTATION OVERVIEW

Marginalized students represent a growing share of public school enrollment. The share of students with disabilities has grown in the past decades to 15% in recent years (Schaeffer, 2023). Public school students are also increasingly children of color (National Center for Education Statistics, 2023), who accounted for the majority of public school enrollment for the first time in 2016 (C. Lindsay et al., 2022). Given these trends, the performance of public education hinges increasingly on how well we serve marginalized students.

Better serving marginalized students is not just a recent necessity but a longstanding imperative for public education. The 1975 Individuals with Disabilities in Education Act (IDEA) guaranteed students with disabilities a free education suited to their needs. However, students with disabilities have had persistently lower academic achievement, lower graduation rates, higher discipline rates, and higher juvenile detention rates than their peers without disabilities (MacSuga-Gage et al., 2021; Morando Rhim et al., 2017; Welsh & Little, 2018a). One of the most enduring and discussed issues in education are the achievement gaps between Black and White students, Latinx and White students, and Native American and White students in test scores, advanced course enrollment, and graduation rates (Ladson-Billings, 2006). Black and Latinx students have long been disproportionately represented among students with disabilities (Annamma et al., 2013; Cruz, Kulkarni, et al., 2021), and students of color with disabilities often experience worse educational outcomes than students of color without disabilities or White students with disabilities (Annamma et al., 2018).

The poor educational outcomes of marginalized students may be understood as a function of inadequate or inequitable support rather than as individual student deficits. For example, scholars have reframed racial achievement gaps as educational debts, the logical and cumulative

result of generations of unequal educational opportunities (Ladson-Billings, 2006). Researchers have found that racial achievement gaps were related to racial discipline gaps (Gregory et al., 2010; Morris & Perry, 2016; Pearman et al., 2019). Although it is difficult to perfectly control for student behavior, prior ratings of problem behavior did not account for the Black-White discipline gap (Huang, 2020), and teachers punished Black and Latinx students more harshly even when they were involved in the same incident alongside White students (Liu, Hayes, et al., 2022a). Evidence from one district suggests that a small share of more inexperienced teachers accounted for a substantial share of the Black-White suspension gap (Liu, Penner, et al., 2022). Thus, these discipline gaps likely reflect biases and shortcomings of adults (Welsh, 2023a; Welsh & Little, 2018a). Beyond discipline and achievement, it matters if marginalized students experience unsupportive schools. Students who perceived worse school climate and had weaker relationships with teachers felt less safe and had lower psychological well-being (Cornell & Mayer, 2010; Roorda et al., 2017; Sanders et al., 2018).

Since beginning graduate school, I have moved my research agenda upstream from academic performance to focus on supportive environments for marginalized students. I applied to graduate school partly motivated by concerns about the Black-White achievement gap. As I learned more about the adverse experiences of not only Black students, but also multiply marginalized students—e.g., the disproportionate use corporal punishment against students of color with disabilities (MacSuga-Gage et al., 2021)—I became more focused on understanding efforts to improve their lived experiences in schools.

My dissertation uses quantitative methods to explore possibilities for educators, parents, and policymakers to ensure supportive environments for marginalized students. This conceptual link loosely unifies the three chapters, which span methodological approaches (descriptive,

experimental, and quasi-experimental) and education levels (early childhood education and K-12). I focus on different units of analysis that shape the educational experiences of marginalized students: teachers, parents, and schools. To better understand factors that shape the experiences of multiply marginalized students, I conduct subgroup analyses where possible. This includes subgroup analyses by teacher race (Chapter 1), by race and disability status (Chapter 2), and by parent and student gender (Chapter 3).

The first chapter focuses on sustaining a more racially and ethnically diverse teacher workforce, a promising support for children of color. About 79% of K-12 teachers are White (Taie & Goldring, 2020). This mismatch between teacher demographics and student demographics represents a missed opportunity for support. A sizeable literature on Black students and a nascent literature on Latinx students has found that they benefitted from same-race teachers in having higher expectations, higher test scores, and fewer suspensions (Lindsay et al., 2022; Lindsay & Hart, 2017; Redding, 2019). In both K-12 and early education settings, same-race teachers perceived children's behavior as significantly less disruptive or defiant compared to ratings of the same child by teachers of different races (Accavitti & Williford, 2020; Dee, 2005).

Chapter one of my dissertation acknowledges that supporting racially marginalized students by recruiting a more diverse workforce also requires supporting and retaining those teachers. Several K-12 studies have found that racial/ethnic match between teachers and principals is related to teachers' levels of trust with their leader, job satisfaction, and job retention (Bartanen & Grissom, 2021; Brezicha & Fuller, 2019; Grissom & Keiser, 2011; Lindsay & Egalite, 2020; Olsen & Huang, 2018; Viano & Hunter, 2017; Vinopal, 2018). However, teacher-leader racial match had not yet been examined in child care. I analyzed how

having a same-race/ethnicity center director relates to child care teacher job satisfaction and turnover in a sample of 1,011 teachers at in Virginia. In contrast with the K-12 evidence, I did not find significant associations between racial/ethnic match and teacher job outcomes after adjusting for covariates. That said, Black and Hispanic teachers were much less likely to have a same-race/ethnicity director.

The second chapter analyzes the effects of a 2017 ban on corporal punishment for students with disabilities in Louisiana. Fifteen states still expressly allow corporal punishment—paddling, spanking, and other physical discipline—in public schools (M. Cardona, 2023). In 2017-2018, the last year of national data collection, public schools reported nearly 100,000 incidents of corporal punishment (Keierleber, 2021). Students with disabilities, especially Black students with disabilities, were disproportionately subjected to corporal punishment (Losen et al., 2019; MacSuga-Gage et al., 2021). Six Southern states have recently restricted corporal punishment for students with disabilities while allowing the practice for other students (M. Cardona, 2023). We do not know how closely schools complied and whether they replaced corporal punishment with in-school or out-of-school suspensions, two other punitive consequences.

Chapter two of my dissertation fills this gap. I use a quadruple difference identification strategy to evaluate schools' compliance with Louisiana's 2017 ban on corporal punishment for students with disabilities and to assess unintended consequences for suspensions, overall and for Black students with disabilities. This chapter makes specific contributions to the literature on corporal punishment and broad contributions to the literature on discipline reform. We have little evidence on compliance with U.S. state corporal punishment bans because historically state bans were enacted after districts stopped using corporal punishment almost entirely (Curran &

Kitchin, 2018). Evaluating a ban specific to students with disabilities allows me to estimate effects in a context where schools were still using corporal punishment. In the broader discipline reform literature, studies have typically focused on suspensions and rarely examined policy tradeoffs arising from potential substitution across discipline practices (Anderson et al., 2019; Khafaji-King, 2024; Sorensen et al., 2022). This chapter provides evidence on discipline bans beyond suspension and examines unintended consequences across discipline outcomes.

The third chapter aims to understand how discipline and test score disparities affect Black parents' school choice preferences and expectations of school racial climate. In 2019, nearly half of Black parents (46%) indicated that public school choices were available to them (National Center for Education Statistics, 2021). Proponents have long hailed school choice as a way to support marginalized students by empowering low-income families and families of color to access higher quality schools (Chubb & Moe, 1990; Friedman & Friedman, 1980; Schneider et al., 1998). Critics have argued that school choice policies do not overcome structural barriers such as residential segregation and lack of access to transportation segregation (Cooper, 2005; Eisenlohr et al., 2023; Lareau et al., 2021; Saporito & Lareau, 1999; Simms & Talbert, 2019; Ukanwa et al., 2022). Moreover, qualitative studies have found that Black parents hesitate to send their children to higher-performing schools due to concerns that these schools have poor racial climates (Butler & Quarles, 2024; Lareau et al., 2021; Posey-Maddox et al., 2021). Yet because we rarely collect large-scale data on Black parents' school preferences, we cannot distinguish the effects of structural barriers from expectations that schools conventionally considered high-performing would be stigmatizing and unwelcoming for their Black children (Eisenlohr et al., 2023). Survey experiments have found that Black parents are much less deterred than White parents by not having a majority same-race school (Hailey, 2022; Mellon &

Siegler, 2023), but there is no such evidence on whether racial gaps in student outcomes deter Black parents.

Chapter three of my dissertation fills this gap. I conduct a survey experiment in a national sample of 1,925 Black parents that randomly assigns them to school profiles with or without disparities in test scores and suspension rates. I hypothesize these disparities proxy for school racial climate: "perceptions of interracial interactions and the socialization around race and culture in a school" Byrd (2017, p. 1). I assess how test score and suspension gaps separately and jointly affect Black parents' desire to enroll and anticipation of school racial climate using hypothetical school profiles modeled after state-run school search websites.

In different ways, each of these chapters contributes to our understanding of possibilities for supporting marginalized students. The first focuses on supporting diverse educators who serve them. The second chapter evaluates a policy intended to protect students with disabilities from corporal punishment, one of the harshest legal school experiences. The last helps clarify why Black parents may or may not pursue conventionally higher achievement schools when given the chance, asking what gaps in student outcomes signal and for whom schools are perceived to be supportive.

These loosely affiliated chapters may become more coherent in light of my positionality. My family, in particular, has shaped my research questions. I dedicate the first chapter to my mother, a Black woman and former child care center director. She retired early from the center, calling it the most difficult job she ever held, not because of the children but because of the frequent staffing challenges. I dedicate the second chapter to my brother and sister-in-law. They asked me to research the schools in the area where they want to start a family, hoping for a school where their future Black children will be supported. I dedicate the third chapter to my

loving father. When I was six years old and lied to him, he gave me my first and only spanking. He did what he thought was best, having been raised in a church that taught "spare the rod, spoil the child." Until learning about the counterfactual in undergraduate econometrics, I, too, thought that that spanking had supported my development. I recognize that even well-intentioned parents, educators, and policymakers may not know how best to support marginalized students—but we can and must find out. This dissertation places the onus on adults to improve the experiences and outcomes of marginalized and multiply marginalized students.

References

Accavitti, M. R., & Williford, A. P. (2020). Teacher perceptions of externalizing behaviour subtypes in preschool: Considering racial factors. *Early Child Development and Care*, *0*(0), 1–15. https://doi.org/10.1080/03004430.2020.1825405

American Psychological Association Zero Tolerance Task Force. (2008). Are zero tolerance policies effective in the schools?: An evidentiary review and recommendations. *American Psychologist*, *63*(9), 852–862. https://doi.org/10.1037/0003-066X.63.9.852

Anderson, K. P., Egalite, A. J., & Mills, J. N. (2019). Discipline Reform: The Impact of a Statewide Ban on Suspensions for Truancy. *Journal of Education for Students Placed at Risk (JESPAR)*, *24*(1), 68–91. https://doi.org/10.1080/10824669.2018.1537794

Annamma, S. A., Connor, D., & Ferri, B. (2013). Dis/ability critical race studies (DisCrit): Theorizing at the intersections of race and dis/ability. *Race Ethnicity and Education*, *16*(1), 1–31. https://doi.org/10.1080/13613324.2012.730511

Annamma, S. A., Ferri, B. A., & Connor, D. J. (2018). Disability Critical Race Theory: Exploring the Intersectional Lineage, Emergence, and Potential Futures of DisCrit in Education. *Review of Research in Education*, *42*(1), 46–71. https://doi.org/10.3102/0091732X18759041

Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). *The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime* (Working Paper 26257; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w26257

Billingham, C. M., & Hunt, M. O. (2016). School Racial Composition and Parental Choice: New Evidence on the Preferences of White Parents in the United States. *Sociology of Education*, *89*(2), 99–117. https://doi.org/10.1177/0038040716635718

Butler, A., & Quarles, B. (2024). “It’s a Chance, Not a Choice”: Black Families, School Choice, and Gentrification in Washington, D.C. *Teachers College Record*, 01614681241242878. https://doi.org/10.1177/01614681241242878

Cardona, M. (2023, March 24). *March 24, 2023—Letter from Secretary Cardona calling for an end to corporal punishment in schools* [Letters (Correspondence)]. US Department of Education (ED). https://www2.ed.gov/policy/gen/guid/secletter/230324.html

Cardona, M. A., & Rodríguez, R. J. (2023). *The State of School Diversity in the United States*. U.S. Department of Education. https://www2.ed.gov/rschstat/eval/resources/diversity.pdf

Chapman, B. (2021, December 9). Schools Confront a Wave of Student Misbehavior, Driven by Months of Remote Learning. *Wall Street Journal*. https://www.wsj.com/articles/schools-student-misbehavior-remote-learning-covid-11639061247

Cornell, D. G., & Mayer, M. J. (2010). Why Do School Order and Safety Matter? *Educational Researcher*, *39*(1), 7–15. https://doi.org/10.3102/0013189X09357616

Cruz, R. A., Firestone, A. R., & Rodl, J. E. (2021). Disproportionality Reduction in Exclusionary School Discipline: A Best-Evidence Synthesis. *Review of Educational Research*, *91*(3), 397–431. https://doi.org/10.3102/0034654321995255

Cruz, R. A., Kulkarni, S. S., & Firestone, A. R. (2021). A QuantCrit Analysis of Context, Discipline, Special Education, and Disproportionality. *AERA Open*, *7*, 23328584211041354. https://doi.org/10.1177/23328584211041354

Curran, F. C., & Kitchin, J. (2018). Estimating the Relationship between Corporal Punishment Use and School Suspensions: Longitudinal Evidence from the Civil Rights Data Collection. *Peabody Journal of Education*, *93*(2), 139–160. https://doi.org/10.1080/0161956X.2018.1435036

Dee, T. S. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American Economic Review*, *95*(2), 158–165.

Gershoff, E. (2010). More Harm Than Good: A Summary of Scientific Research on the Intended and Unintended Effects of Corporal Punishment on Children. *Law and Contemporary Problems*, *73*(2), 31–56.

Gershoff, E., Sattler, K. M. P., & Holden, G. W. (2019). School corporal punishment and its associations with achievement and adjustment. *Journal of Applied Developmental Psychology*, *63*, 1–8. https://doi.org/10.1016/j.appdev.2019.05.004

Gershoff, E. T. (2017). School corporal punishment in global perspective: Prevalence, outcomes, and efforts at intervention. *Psychology, Health & Medicine*, *22*(sup1), 224–239. https://doi.org/10.1080/13548506.2016.1271955

Gershoff, E. T., & Font, S. A. (2016). Corporal Punishment in U.S. Public Schools: Prevalence, Disparities in Use, and Status in State and Federal Policy. *Social Policy Report*, *30*, 1.

Gilliam, W. S. (2005). *Prekindergarteners Left Behind: Expulsion Rates in State Prekindergarten Systems*. Yale University Child Study Center. https://medicine.yale.edu/childstudy/policy/Images/National%20Prek%20Study_expulsion_tcm942-34954.pdf

Gilliam, W. S., & Reyes, C. R. (2018). Teacher Decision Factors That Lead to Preschool Expulsion. *Infants & Young Children*, *31*(2), 93–108. https://doi.org/10.1097/IYC.0000000000000113

Giordano, K., Interra, V. L., Stillo, G. C., Mims, A. T., & Block-Lerner, J. (2021). Associations Between Child and Administrator Race and Suspension and Expulsion Rates in

Community Childcare Programs. *Early Childhood Education Journal*, *49*(1), 125–133. https://doi.org/10.1007/s10643-020-01049-1

Goldring, E. B., & Hausman, C. S. (1999). Reasons for parental choice of urban schools. *Journal of Education Policy*, *14*(5), 469–490. https://doi.org/10.1080/026809399286161

Granja, M. R., Smith, S., Nguyen, U. S., & Grifa, B. (2018). *Learning about Young Children's Challenging Behavior and Impacts on Programs and Families: A State-wide Survey of Virginia's Early Care and Education Teachers*. 18.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The Achievement Gap and the Discipline Gap: Two Sides of the Same Coin? *Educational Researcher*, *39*(1), 59–68. https://doi.org/10.3102/0013189X09357621

Hailey, C. A. (2022). Racial Preferences for Schools: Evidence from an Experiment with White, Black, Latinx, and Asian Parents and Students. *Sociology of Education*, *95*(2), 110–132. https://doi.org/10.1177/00380407211065179

Haney, J. E. (1978). The Effects of the Brown Decision on Black Educators. *The Journal of Negro Education*, *47*(1), 88–95. https://doi.org/10.2307/2967104

Hashim, A. K., Strunk, K. O., & Dhaliwal, T. K. (2018). Justice for All? Suspension Bans and Restorative Justice Programs in the Los Angeles Unified School District. *Peabody Journal of Education*, *93*(2), 174–189. https://doi.org/10.1080/0161956X.2018.1435040

Heekes, S.-L., Kruger, C. B., Lester, S. N., & Ward, C. L. (2022). A Systematic Review of Corporal Punishment in Schools: Global Prevalence and Correlates. *Trauma, Violence, & Abuse*, *23*(1), 52–72. https://doi.org/10.1177/1524838020925787

Holmes Erickson, H. (2017). How do parents choose schools, and what schools do they choose? A literature review of private school choice programs in the United States. *Journal of School Choice*, *11*(4), 491–506. https://doi.org/10.1080/15582159.2017.1395618

Houston, D. M., & Henig, J. R. (2021). The Effects of Student Growth Data on School District Choice: Evidence from a Survey Experiment. *American Journal of Education*, *127*, 33.

Houston, David M. & Henig, Jeffrey R. (2022). *The 'Good' Schools: Academic Performance Data, School Choice, and Segregation*. https://doi.org/10.26300/7N5S-KM15

Huang, F. L. (2020). Prior Problem Behaviors Do Not Account for the Racial Suspension Gap. *Educational Researcher*, *49*(7), 493–502. https://doi.org/10.3102/0013189X20932474

Keierleber, M. (2021, May 19). Kids Keep Getting Hit, Even in Schools Where It's Banned. *The74Million*. https://www.the74million.org/article/kids-keep-getting-hit-at-school-even-where-corporal-punishment-is-banned/

Khafaji-King, J. A. (2024). *Disability as Discipline? Effects of the New York City Suspension Ban on Identification of Students with Disabilities*. https://jk7317.github.io/jakwebsite/disabilitydiscipline_01052024.pdf

King, Jr., J. B. (2016, November 22). *Letter to Governors and Chief School Officers Re: Corporal Punishment*. https://www2.ed.gov/documents/press-releases/11212016-corporal-punishment.pdf

Kurtz, H. (2022, January 12). Threats of Student Violence and Misbehavior Are Rising, Many School Leaders Report. *Education Week*. https://www.edweek.org/leadership/threats-of-student-violence-and-misbehavior-are-rising-many-school-leaders-report/2022/01

Lacoe, J., & Steinberg, M. P. (2018). Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes. *Peabody Journal of Education*, *93*(2), 207–227. https://doi.org/10.1080/0161956X.2018.1435047

Lacoe, J., & Steinberg, M. P. (2019). Do Suspensions Affect Student Outcomes? *Educational Evaluation and Policy Analysis*, *41*(1), 34–62. https://doi.org/10.3102/0162373718794897

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher*, *35*(7), 3–12. https://doi.org/10.3102/0013189X035007003

Lareau, A., Weininger, E. B., & Warner-Griffin, C. (2021). Structural constraints and the school choice strategies of black American middle-class parents. *British Journal of Sociology of Education*, *42*(4), 490–505. https://doi.org/10.1080/01425692.2021.1872367

Lindsay, C., & Hart, C. M. D. (2017). Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina. *Educational Evaluation and Policy Analysis*, *39*(3), 485–510.

Lindsay, C., Monarrez, T., & Luetmer, G. (2022). *The effects of teacher diversity on hispanic student achievement in texas* (p. 45). Urban Institute. https://www.urban.org/sites/default/files/publication/105325/the-effects-of-teacher-diversity-on-hispanic-student-achievement-in-texas_0.pdf

Liu, J., Hayes, M. S., & Gershenson, S. (2022a). JUE insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline. *Journal of Urban Economics*, 103453. https://doi.org/10.1016/j.jue.2022.103453

Liu, J., Hayes, M. S., & Gershenson, S. (2022b). JUE insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline. *Journal of Urban Economics*, 103453. https://doi.org/10.1016/j.jue.2022.103453

Liu, J., Penner, E., & Gao, W. (2022). Troublemakers? The Role of Frequent Teacher Referrers in Expanding Racial Disciplinary Disproportionalities. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://edworkingpapers.com/ai22-616

Losen, D., Whitaker, A., Kizzire, J., Savitsky, Z., & Dunn, K. (2019). *The Striking Outlier: The Persistent, Painful, and Problematic Practice of Corporal Punishment in Schools*. Southern Poverty Law Center. https://www.splcenter.org/20190611/striking-outlier-persistent-painful-and-problematic-practice-corporal-punishment-schools

MacSuga-Gage, A. S., Gage, N. A., Katsiyannis, A., Hirsch, S. E., & Kisner, H. (2021). Disproportionate Corporal Punishment of Students With Disabilities and Black and Hispanic Students. *Journal of Disability Policy Studies*, *32*(3), 212–223. https://doi.org/10.1177/1044207320949960

Meek, S., Smith, L., Allen, R., Catherine, E., Edyburn, K., Williams, C., Fabes, R., McIntosh, K., Garcia, E., Takanishi, R., Gordon, L., Jimenez-Castellanos, O., Hemmeter, M. L., Gilliam, W., & Pontier, R. (2020). *Start with Equity: From the Early Years to the Early Grades | Center for Child and Family Success*. Children's Equity Project and Bipartisan Policy Center. https://childandfamilysuccess.asu.edu/cep/start-with-equity

Morando Rhim, L., Sutter, J., & Campbell, N. (2017, January 31). Improving Outcomes for Students with Disabilities. *Center for American Progress*. https://www.americanprogress.org/article/improving-outcomes-for-students-with-disabilities/

Morris, E. W., & Perry, B. L. (2016). The Punishment Gap: School Suspension and Racial Disparities in Achievement. *Social Problems*, *63*(1), 68–86. https://doi.org/10.1093/socpro/spv026

National Center for Education Statistics. (2021). *Public school choice programs* (Fast Facts). Institute of Education Sciences. https://nces.ed.gov/fastfacts/display.asp?id=6

National Center for Education Statistics. (2023). *Racial/Ethnic Enrollment in Public Schools* (Conditions of Education). Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cge/racial-ethnic-enrollment#suggested-citation

Nishioka, V., Merrill, B., & Hanson, H. (2021). *Changes in Exclusionary and Nonexclusionary Discipline in Grades K–5 Following State Policy Reform in Oregon* (REL 2021 061). Institute of Education Sciences. https://files.eric.ed.gov/fulltext/ED610682.pdf

Novoa, C., & Rasheed, M. (2018). *Suspensions Are Not Support*. Center for American Progress. https://www.americanprogress.org/article/suspensions-not-support/

Organda Portela, M. J., & Pells, K. (2015). *Corporal Punishment in Schools: Longitudinal Evidence from Ethiopia, India Peru and Viet Nam* (Innocenti Discussion Paper No. 2015-02; Young Lives: An International Study of Childhood Poverty). UNICEF Office of Research. https://www.unicef-irc.org/publications/pdf/CORPORAL%20PUNISHMENTfinal.pdf

Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are Achievement Gaps Related to Discipline Gaps? Evidence From National Data. *AERA Open*, *5*(4), 2332858419875440. https://doi.org/10.1177/2332858419875440

Pendharkar, E. (2023, November 21). Disparities, Bullying, and Corporal Punishment: The Latest Federal Discipline Data. *Education Week*. https://www.edweek.org/leadership/disparities-bullying-and-corporal-punishment-the-latest-federal-discipline-data/2023/11

Posey-Maddox, L., de Royston, M. M., Holman, A. R., Rall, R. M., & Johnson, R. A. (2021). No Choice Is the "Right" Choice: Black Parents' Educational Decision-Making in Their Search for a "Good" School. *Harvard Educational Review*, *91*(1), 38–61. https://doi.org/10.17763/1943-5045-91.1.38

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, *89*(4), 499–535. https://doi.org/10.3102/0034654319853545

Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, *116*(17), 8255–8260. https://doi.org/10.1073/pnas.1808307116

Rodriguez, L. A., & Welsh, R. O. (2022). The Dimensions of School Discipline: Toward a Comprehensive Framework for Measuring Discipline Patterns and Outcomes in Schools. *AERA Open*, *8*, 23328584221083669. https://doi.org/10.1177/23328584221083669

Roorda, D. L., Jak, S., Zee, M., Oort, F. J., & Koomen, H. M. Y. (2017). Affective Teacher–Student Relationships and Students' Engagement and Achievement: A Meta-Analytic Update and Test of the Mediating Role of Engagement. *School Psychology Review*, *46*(3), 239–261. https://doi.org/10.17105/SPR-2017-0035.V46-3

Sanders, S. M., Durbin, J. M., Anderson, B. G., Fogarty, L. M., Giraldo-Garcia, R. J., & Voight, A. (2018). Does a rising school climate lift all boats? Differential associations of perceived climate and achievement for students with disabilities and limited English proficiency. *School Psychology International*, *39*(6), 646–662. https://doi.org/10.1177/0143034318810319

Schaeffer, K. (2023, July 24). What federal education data shows about students with disabilities in the U.S. *Pew Research Center*. https://www.pewresearch.org/short-reads/2023/07/24/what-federal-education-data-shows-about-students-with-disabilities-in-the-us/

Schneider, M., & Buckley, J. (2002). What Do Parents Want From Schools? Evidence From the Internet. *Educational Evaluation and Policy Analysis*, *24*(2), Article 2. https://doi.org/10.3102/01623737024002133

Skiba, R. J. (2015). Interventions to Address Racial/Ethnic Disparities in School Discipline: Can Systems Reform Be Race-Neutral? In R. Bangs & L. E. Davis (Eds.), *Race and Social Problems: Restructuring Inequality* (pp. 107–124). Springer. https://doi.org/10.1007/978-1-4939-0863-9_7

Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race Is Not Neutral: A National Investigation of African American and Latino Disproportionality in School Discipline. *School Psychology Review*, *40*(1), 85–107. https://doi.org/10.1080/02796015.2011.12087730

Sorensen, L. C., Bushway, S. D., & Gifford, E. J. (2022). Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes. *Education Finance and Policy*, *17*(2), 255–284. https://doi.org/10.1162/edfp_a_00341

Southward, L. H., Blanchard, T., Costello, H., Hanna, H., & Buffum, L. (2006). *Economic volatility and expulsion in US child care settings: A multi-state survey. Mississippi State, MS.* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.399.1223&rep=rep1&type=pdf

Taie, S., & Goldring, R. (2020). Characteristics of public and private elementary and secondary school teachers in the United States: Results from the 2017-18 national teacher and principal survey. First look. NCES 2020-142. *National Center for Education Statistics*.

US Department of Education Office of Civil Rights & US Department of Justice Civil Rights Division. (2014). *Dear Colleague Letter on the Nondiscriminatory Administration of School Discipline.* https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.pdf

Vinh, M., Strain, P., Davidon, S., & Smith, B. J. (2016). One State's Systems Change Efforts to Reduce Child Care Expulsion: Taking the Pyramid Model to Scale. *Topics in Early Childhood Special Education*, *36*(3), 159–164. https://doi.org/10.1177/0271121415626130

Wang, R. (2022). The Impact of Suspension Reforms on Discipline Outcomes: Evidence From California High Schools. *AERA Open*, *8*, 23328584211068067. https://doi.org/10.1177/23328584211068067

Welsh, R. O. (2022). School Discipline in the Age of COVID-19: Exploring Patterns, Policy, and Practice Considerations. *Peabody Journal of Education*, *97*(3), 291–308. https://doi.org/10.1080/0161956X.2022.2079885

Welsh, R. O. (2023a). Up the down escalator? Examining a decade of school discipline reforms. *Children and Youth Services Review*, *150*, 106962. https://doi.org/10.1016/j.childyouth.2023.106962

Welsh, R. O. (2023b, October 6). I Combed Through 81 Studies on School Discipline. Here's What Educators Need to Know. *Education Week*. https://www.edweek.org/leadership/opinion-i-combed-through-81-studies-on-school-discipline-heres-what-educators-need-to-know/2023/10

Welsh, R. O., & Little, S. (2018a). The School Discipline Dilemma: A Comprehensive Review of Disparities and Alternative Approaches. *Review of Educational Research*, *88*(5), 752–794. https://doi.org/10.3102/0034654318791582

Welsh, R. O., & Little, S. (2018b). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research*, *88*(5), 752–794. https://doi.org/10.3102/0034654318791582

Whitehurst, G. J. (2017). *Education Choice and Competition Index 2016: Summary and Commentary*. Center on Children and Families at Brookings. https://www.brookings.edu/articles/the-2016-education-choice-and-competition-index/

CHAPTER 1

A director like me: Teacher-leader racial/ethnic match and job outcomes in child care centers

Todd Hall*
University of Virginia

Daphna Bassok
University of Virginia

Justin B. Doromal
Brown University

Anna J. Markowitz
University of California, Los Angeles

Hall, T., Bassok, D., Doromal, J. B., & Markowitz, A. J. (2023). A director like me: Teacher-leader racial/ethnic match and job outcomes in child care centers. *Early Childhood Research Quarterly*, *62*, 369-384. https://doi.org/10.1016/j.ecresq.2022.10.004

Acknowledgements: The research-policy partnership described in this manuscript was supported through a federal Preschool Development Birth-Five grant awarded to the Virginia Department of Education (#90TP0039.02.00). The authors wish to acknowledge the Virginia Department of Education and the Virginia Early Childhood Foundation for their partnership. Todd Hall's time for analysis and writing was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200005 to the University of Virginia. The contents of this article are those of the authors and do not represent the official views of these organizations or institutions.

*Corresponding author. Contact: PO Box 400879, Charlottesville, VA 22904. tmh4xa@virginia.edu

Abstract

The leaders of child care centers shape the everyday experiences of child care teachers, who play a critical role in supporting young children. While research has linked leadership to policy-relevant outcomes such as teacher job satisfaction and turnover, little is known about specific leader characteristics linked to teacher job outcomes. In K-12, one key factor associated with increased job satisfaction and lower turnover is teacher-principal racial/ethnic match, however there is no evidence to date on how teacher-leader racial/ethnic match relates to teacher job outcomes in child care settings. Using data from a sample of 1,011 teachers at child care centers in Virginia, this study documents the prevalence of teacher-leader racial/ethnic match. We estimate associations between racial/ethnic match and teachers' views of leadership, job satisfaction, turnover intentions, and turnover after six months. Although two-thirds of child care teachers experienced racial/ethnic match with their child care center leader, Black and Hispanic teachers were far less likely to experience racial/ethnic match than White teachers. We did not find significant associations between racial/ethnic match and teacher job outcomes after adjusting for covariates. We discuss potential explanations, directions for future research, and implications for policy.

Leaders of child care centers can support the teachers at their site, who in turn play a critical role in children's lives. Like K-12 school principals, child care leaders can serve as instructional leaders, coaches, and mentors, providing professional development to teachers at their centers (Grissom & Keiser, 2011; Zinsser & Curby, 2014). Leaders may also shape teachers' experiences through their role managing staffing and budget decisions and cultivating positive workplace climate (Doromal & Markowitz, 2021; Zinsser & Curby, 2014). Indeed, when child care teachers rate their leaders as more supportive, they report greater job satisfaction and leave their jobs at lower rates (Doromal & Markowitz, 2021). However, research on the importance of leaders in child care settings—independently-run centers not funded by federal Head Start and not operated by schools—is sparse. We know little about factors that influence how child care teachers perceive their leaders or about the relationship between these views and policy relevant outcomes, such as job satisfaction and retention (Muijs et al., 2004).

One salient characteristic that may contribute to both child care teachers' views of their leader and their job outcomes is having a shared racial/ethnic background. In the K-12 context, teacher-principal racial/ethnic match has been linked to improved teacher job satisfaction and reduced teacher turnover, perhaps by enhancing trust, support, and open communication (Bartanen & Grissom, 2021; Grissom & Keiser, 2011; Lindsay & Egalite, 2020).

It is possible that teacher-leader racial/ethnic match in early childhood settings plays a similar role. Indeed, a recent National Academies of the Sciences report emphasized the importance of leaders who mirror their staff (IOM/NRC, 2015). Likewise, the largest professional organization of early childhood educators, NAEYC, has advocated for more representation of women of color in leadership roles (NAEYC, 2019). Policymakers, practitioners, and researchers have also voiced concern about what ideas are reinforced when

Black and Hispanic women are underrepresented in child care leadership (Johnson-Staub, 2017; Sandstrom & Schilder, 2021). There is no evidence, however, on whether teacher-leader racial/ethnic match is associated with key teacher outcomes such as views of leadership, job satisfaction, and turnover in child care or in early childhood education (ECE) more broadly.

Examining the relationship between teacher-leader racial/ethnic match and job outcomes in child care matters because the findings may highlight ways to better support and retain child care teachers from diverse backgrounds (Bartanen & Grissom, 2021; Brezicha & Fuller, 2019; Lindsay & Egalite, 2020). The child care workforce includes a large proportion of teachers of color (Austin et al., 2019); if racial/ethnic match is associated with job satisfaction and retention, as in K-12, this should inform workforce recruitment and retention policies.

However, patterns observed in K-12 may not transfer to child care. First, racial/ethnic match might not counteract the poverty-level wages, challenging working conditions, and high turnover rates common to these settings (Bassok, Smith, et al., 2021; McDonald et al., 2018; McLean et al., 2021).

Further, having a leader of the same race/ethnicity might be less salient in child care, where the workforce is more diverse than in K-12. K-12 teachers are predominantly White (Spiegelman, 2020), which can result in greater isolation for teachers of color, and in turn less job satisfaction (Grissom & Bartanen, 2022; Kemper Patrick & Arturo Santelli, 2022; Ravenell et al., 2018). In 2017, 79% of all traditional public school teachers and 85% of all private school teachers in the United States were non-Hispanic White (Taie & Goldring, 2020). In Virginia, the context for this study, 80% of public school educators in 2020-2021 were non-Hispanic White (Virginia Department of Education, 2022). In contrast, a 2019 survey of early educators in Virginia found that 54% of lead teachers in child care centers were White (Bassok et al., 2020).

Nationally in 2019, 24% of educators in child care centers identified with a race other than White and 13% identified as Hispanic, regardless of race (Greenberg & Luetmer, 2022).

This paper is the first we know of to explore teacher-leader racial/ethnic match in child care centers. Using survey and state administrative data from Virginia, we address two research questions: 1) what is the prevalence of teacher-leader racial/ethnic match in child care centers, and 2) what is the relationship between racial/ethnic match and teacher job outcomes, including teachers' views of their leader, job satisfaction, turnover intentions, and observed six-month turnover.

Why Are Child Care Leaders Important?

Stable, well-supported teachers are essential for high-quality child care. Unfortunately, child care teachers face difficult jobs with few supports and low compensation (McLean et al., 2021). These conditions likely contribute to high annual turnover in child care—nationally, 16% at for-profit, independent centers and 27% at chains—and weaken the teacher-child interactions that support child development (Tran & Winsler, 2011; Whitebook et al., 2014; Yoshikawa et al., 2013). Turnover disrupts teacher-child relationships, wastes investments in teacher training and professional development, and diverts effort from quality improvement (Cassidy et al., 2011; Kwon et al., 2020; Whitebook & Sakai, 2003). In contrast, job satisfaction is positively linked to warm and stimulating teacher-child interactions (Thomason & La Paro, 2013).

In child care centers, leaders (e.g., center directors) fulfill multiple roles that support or hinder teachers in managing the challenges of their jobs, potentially increasing satisfaction and reducing turnover (IOM/NRC, 2015). Leaders shape centers' organizational culture and teacher experiences in a variety of ways. Some leaders may fulfill an instructional leadership role by providing feedback, coaching, mentoring, and professional development; others may focus on

staffing issues and creating positive work climates that support teacher well-being and job satisfaction (Zinsser et al., 2016; Zinsser & Curby, 2014). They can establish trust, foster a sense of community, counsel struggling teachers, and articulate a unified vision for the center (Doromal & Markowitz, 2021; Zinsser & Curby, 2014; Zulauf & Zinsser, 2019). Leaders also oversee staffing and budget decisions, including the provision of wages and benefits, which directly impact their teachers' satisfaction and turnover decisions (IOM/NRC, 2015; LeeKeenan & Chin Ponte, 2018). Finally, child care leaders face substantial administrative demands related to staff documentation and compliance with licensing regulations, which may temper their ability to support teaching and learning (Sims et al., 2019). For all these reasons, leaders are central to most aspects of teachers' experiences at their center.

Although there is no agreed-upon measure of leader effectiveness (Muijs et al., 2004), the way teachers view their leaders is likely an important metric. Qualitative research finds that teachers reported higher child care quality when they felt leaders cultivated a shared vision for quality (e.g., inputs, process, or child outcomes) (Andersen et al., 2018). One quantitative study finds that teachers who rated their leaders as supportive, respectful, and trustworthy were less likely to turn over (Doromal & Markowitz, 2021). Studies have also linked teacher-leader relationships to turnover in Head Start centers (Jeon & Wells, 2018; Wells, 2015). Together, this research suggests that teachers' views of their leaders could be linked to job outcomes.

How Might Teacher-Leader Racial/Ethnic Match Matter in Child Care?

There are both theoretical and empirical reasons to hypothesize that racial/ethnic match may influence teachers' views of their leaders, job satisfaction, and turnover. Representative bureaucracy theory, a public adminstration theory that has informed K-12 racial/ethnic match research, predicts that a shared leader-constituent demographic background is associated with

shared socialization experiences, and thereby decisions that align with constituent values and preferences (Lim, 2006; Meier & Stewart Jr, 1992; Selden, 1997). Consistent with this theory, a sizeable body of K-12 research has demonstrated that racial/ethnic match between teachers and principals is related to teachers' levels of trust with their leader, job satisfaction, and job retention (Bartanen & Grissom, 2021; Brezicha & Fuller, 2019; Grissom & Keiser, 2011; Lindsay & Egalite, 2020; Olsen & Huang, 2018; Viano & Hunter, 2017; Vinopal, 2018).

To date, teacher-leader racial/ethnic match has not been explored in child care, but analogous research suggests that match matters for both young children and for family engagement. For example, in pre-k and Kindergarten, teacher-child racial/ethnic match has been associated with more positive assessments of learning and behavior (Bates & Glick, 2013; Downer et al., 2016; Downey & Pribesh, 2004). Outside the classroom, racial/ethnic match between social service providers and families has been associated with home visiting attendance, participation in Head Start events, and more favorable perceptions of mental health counselors (Huang & Zane, 2016; Markowitz et al., 2020; McCurdy et al., 2003). It is thus plausible that racial/ethnic match may also matter for child care teachers. Below, we outline two hypothesized mechanisms by which teacher-leader racial/ethnic match might improve job outcomes in child care centers: leader provision of supports and shared communication style.

Leader-Provided Supports

Representative bureaucracy theory predicts that leaders may secure benefits specifically for people with a common social origin, to either show partiality or counteract existing discrimination or disadvantage (Lim, 2006; Meier & Stewart Jr, 1992). In the child care context, a same-race leader may do this by providing concrete supports for managing the day-to-day challenges of working with children—for example supporting teachers in managing child

behavior, an area that has been called out as crucial in previous research (Granja et al., 2018; Martin et al., 2018; Wymer et al., 2020; Zulauf & Zinsser, 2019). Leaders may also secure resources such as staff mental health trainings and consultants (Zinsser et al., 2016), and may more readily pursue or find more effective resources for same-race/ethnicity teachers. Leaders who share their racial/ethnic background with a teacher might also devote more time to coaching that teacher or communicating with parents on their behalf (Grissom & Jones, 2020; Martin et al., 2018). This is particularly relevant for job satisfaction and turnover as some teachers leave child care because they can find better paid, less challenging work elsewhere, and a supportive leader may counteract those challenges (McDonald et al., 2018; Whitebook & Sakai, 2003).

Leaders may also offer more encouragement and recognition to teachers with a shared racial or ethnic background (Grissom & Keiser, 2011; Viano & Hunter, 2017); and, when leaders offer formal recognition of teacher efforts or acknowledge teacher achievements (Zinsser et al., 2016), it may lead teachers to feel more respected by, supported by, and satisfied with leaders. This may matter in child care because teachers often cite the rewarding nature of the work itself as a primary reason to stay, despite low pay (McDonald et al., 2018; Vinopal, 2018).

Shared Communication Style and Values

Leaders may also be more likely to share values, preferences, and communication styles with people from a common social origin (Lim, 2006). Shared backgrounds and communication styles may make it easier for leaders and teachers to interact and work productively (Bartanen & Grissom, 2021; Grissom & Jones, 2020; Viano & Hunter, 2017). In child care, this may make teachers more comfortable asking for help and ease difficult conversations because the leader reacts in an expected and relatable manner. Having a same-race leader may also elicit more enthusiastic cooperation from teachers because they seek the approval of a leader with whom

they identify (Grissom et al., 2015; Lim, 2006). K-12 research provides some support for this theory: Teachers placed more trust in same-race principals and were more likely to apply to and remain in positions under same-race principals (Brezicha & Fuller, 2019; Goff et al., 2018).

Racial/ethnic match may also alleviate naivete about workplace racial tensions, particularly for teachers of color. This may in turn reduce burdensome expectations that teachers of color should solve race-related issues that may arise, such as racial disparities in discipline (Accavitti & Williford, 2020; Brezicha & Fuller, 2019). This open communication may also translate to a sense of trust, respect, and commitment to an anti-racist learning environment.

Evolving Evidence on Racial/Ethnic Match

Despite the evidence that racial/ethnic match matters, there are also contexts in which the hypothesized benefits of match have not been observed. For example, one K-12 study found a significant negative association between teacher-principal racial/ethnic match and teacher transfers to other schools, but not for exits from teaching (Ravenell et al., 2018). Another failed to find significant associations between match and job outcomes among inexperienced teachers (Lindsay & Egalite, 2020). This finding may be particularly relevant for child care, which has a high proportion of new teachers. In mental health counseling, there is evidence that clients prefer same-race counselors, but not that having a same-race counselor predicts treatment outcomes (Huang & Zane, 2016). There is also evidence that the salience of teacher-leader racial/ethnic match in K-12 may be declining over time: A study of multiple waves of the School Administrator Staffing Survey found a significant association between racial/ethnic match and job satisfaction in the earliest wave of data (1999-2000), but in no other year up to 2011 (Viano & Hunter, 2017). Some K-12 researchers have also suggested that associations between racial/ethnic match and job outcomes may be driven by White teachers' dissatisfaction with

Black principals (Olsen & Huang, 2018; Viano & Hunter, 2017), rather than by benefits for teachers of color. Given the differences in both workforces and work contexts between child care and K-12, research focused on racial/ethnic match in child care is warranted.

Present Study

This quantitative study explores teacher-leader racial/ethnic match and its association with job outcomes in child care for the first time. We ask two research questions:

1. What is the prevalence of teacher-leader racial/ethnic match in child care centers, and does it vary by teacher race/ethnicity?
2. What is the relationship between racial/ethnic match and teacher job outcomes, including teachers' views of their leader, job satisfaction, and observed six-month turnover, and does it vary by teacher race/ethnicity?

We hypothesize that racial/ethnic match will be common and associated with more positive views of leadership, greater job satisfaction, and lower six-month turnover. Our investigation has two key strengths. First, we analyze both common survey-reported job outcomes and unique administrative data on observed six-month teacher turnover. Second, our sample includes multiple teachers within centers, so we can compare teachers at the same center who do and do experience racial/ethnic match. Doing so better isolates the association between match and outcomes, as it accounts for other center-level characteristics that may influence teacher job outcomes. Our study thus provides new insights for policymakers looking to improve child care stability through more effective leadership, particularly in the wake of COVID-19.

Method

Data and Sample

Data were collected through an ongoing research policy partnership with the University

of Virginia, the Virginia Department of Education (VDOE), and a nonprofit partner, the Virginia Early Childhood Foundation. This partnership was formed to support Virginia's Preschool Development Grant Birth-5 (PDG) initiative, which aims to improve access to high-quality ECE throughout the state. In fall 2020, the PDG included 70 cities and counties in Virginia grouped into 15 PDG communities. All ECE programs in these communities receiving child care subsidies, state funds, or federal funds had the opportunity to participate, but this study only includes community-based child care centers (i.e., not Head Start, Early Head Start, or school-system affiliated programs). The PDG-participating communities were both rural and urban and accounted for about two-thirds of Virginia's population. They largely mirrored the demographics of the state but had slightly lower median incomes than the state average. In fall 2020, 294 child care centers, or about 29% of all subsidy-accepting centers in these communities, participated in the PDG. Within those sites all teachers working at least 30 hours per week were eligible for the Teacher Recognition Program (TRP), a financial incentive related to PDG paraticipation (Bassok, Doromal, et al., 2021).

We combined three data sources from the PDG initiative. First, we used data from a survey we administered to child care teachers from October to December 2020. All teachers at PDG-participating centers were invited to take the survey, which was available in English and Spanish. About 73% did so and received a $20 gift card for their time.

Second, we used data from a survey we fielded to leaders working at PDG-participating centers at the same time. The leader survey, which was also available in English and Spanish, aimed to reach at least one leader at each center, and did so in 85% of centers. Leader respondents also received a $20 gift card. For the present study, we retained centers where at least one leader identified their role as director, owner/director, owner, assistant director, or

principal/assistant principal. Of these, most centers (84%) had exactly one leader take the survey; in the 30 centers where multiple leaders took the survey, we retained the leader who had the most responsibility for teachers based on their title. That is, we prioritized directors, followed by owners/directors, principals, owners, and assistant directors or assistant principals. Results were not sensitive to the exclusion of centers with multiple surveyed leaders. In our final analtyic sample, 93% of leaders identified as directors, owners/directors, or owners.

Third, we used administrative teacher turnover data collected by our state partners. As part of TRP, they contacted centers to determine which of the teachers who were working 30 hours per week or more at the beginning of the study were still doing so six months later. The administrative data provide our turnover outcome measure.

Our analyses depended on knowing both teacher and leader race, so we limited the sample to child care teachers who completed the fall 2020 teacher survey and whose leader also completed a survey. Given the demographic makeup of Virginia early educators, we also restricted to teachers who self-reported Black or White as their only race or who reported Hispanic as their ethnicity regardless of race, and whose leader also reported their race/ethnicity as one of these groups. Among leader survey takers who worked at a site with teacher respondents, 89% identified as Black, White, or Hispanic. Likewise, about 87% of the teacher survey takers who had a completed leader survey identified as Black, White, or Hispanic. While teacher-leader racial/ethnic match may be important for teachers of other racial/ethnic groups, we only had sufficient samples to explore match for these three categories. Finally, we dropped teachers who had missing information on all outcome variables.

The final analytic sample included 1,011 teachers at 188 centers, with an average of five teachers per center. While our sample is large relative to most existing research on child care

leadership, it is small compared to studies of racial/ethnic match in K-12 settings, which often include tens of thousands of teachers (Grissom & Jones, 2020; Lindsay & Egalite, 2020; Ravenell et al., 2018; S. L. Viano & Hunter, 2017), raising the concern that our analyses may have low power, a point we return to in the discussion.

Measures

Race/Ethnicity and Racial/Ethnic Match

Teachers and leaders were both asked the following question: "What is the best description of your race/ethnicity? Mark all that apply." Respondents could select "American Indian or Alaskan Native, non-Hispanic," "Asian, non-Hispanic," "Black, non-Hispanic," "Hispanic, regardless of race," "Native Hawaiian or other Pacific Islander, non-Hispanic," or "White, non-Hispanic," or they could write in a preferred identification. Our question included Hispanic alongside racial categories to align with the standard approach of our VDOE partners and because asking separately can lead to misidentification for Hispanic individuals (Viano & Baker, 2020). Because we limited our sample to Black (27%), Hispanic (14%), and White (59%) respondents, no teachers or leaders had missing race/ethnicity information. White teachers are the omitted category in regressions. To operationalize teacher-leader racial/ethnic match, we created a binary variable where teachers were coded as 1 if the teacher reported the same race/ethnicity as their leader, and 0 otherwise.

Self-identified race/ethnicity can serve as a marker for shared background (Viano & Baker, 2020)—and has been used in most quantitative studies of racial/ethnic match—but does have limitations. Effects of racial/ethnic match likely depend on how teachers and leaders perceive each other, but this is not captured in our measure, and may differ from how individuals perceive themselves. This measurement error would bias estimated match effects towards zero,

as we may have coded individuals who do have perceived match as non-matched, and vice versa.

Teacher Job Outcomes

Teacher job outcomes included teacher views of their leader, job satisfaction, and turnover. We first present items used to measure these constructs, and, for views of leaders and job satisfaction, then describe how they were combined using factor analysis.

Views of Leadership. Our items assessing teachers' views of leadership initially came from the "School Leadership" subscale of the Virginia School Climate Survey, a biennial working conditions survey of licensed professionals working in traditional public schools. The survey was designed for Virginia, though some items were taken from similar surveys in North Carolina and Tennessee (Brezicha & Fuller, 2019; Kemper Patrick & Arturo Santelli, 2022). The scale is reliable among K-12 teachers (Miller, (2020), with a Cronbach's alpha for the school leadership items of 0.961, and factor loadings ranging from 0.880-0.966.

Because our study was conducted in partnership with VDOE, we started with this existing measure and made modifications as necessary to make items appropriate for child care settings (i.e., changing the phrase "school administrators" to "site leader" and "school" to "site"). Teachers responded on a five-point agreement scale from strongly disagree to strongly agree to items such as "I feel respected by the site leader" and "I feel comfortable raising issues and concerns that are important to me with the site leader" (see Appendix Table 1 for full scale). Items are similar to those of other ECE leadership scales (e.g., Ehrlich et al., 2019). We also added one new item given the racial reckoning of summer 2020: "The site leader is committed to creating an anti-racist learning environment for the children we serve."

To understand the proportion of total variance accounted for at the center level, we computed intraclass correlations (ICCs) using Stata's mixed command. ICCs ranged from 0.07 to

0.12. Missingness for views of leadership items ranged from 4-5% (see Appendix Table 1).

Job Satisfaction. This measure came from three survey items. First, teachers rated their overall satisfaction with their job on a four-point scale (*not at all, a little bit, somewhat, or very*). Second, teachers rated their satisfaction with interactions with site leaders on the same scale. Third, teachers indicated their agreement with the statement "I really enjoy my present job" on a five-point scale (strongly disagree to strongly agree). ICCs for these items were 0.06, 0.10, and 0.00, respectively. Each item was missing for roughly 4-5% of teachers (see Appendix Table 1).

Factor Analysis. Because of the large number of correlated items, we ran exploratory factor analyses on all leadership and satisfaction items in their original ordinal scales. Results supported a two-factor solution that mapped on to the original item categories based on factor loadings and eigenvalues greater than one (root mean squared error = 0.072). We then used categorical confirmatory factor analysis (CCFA) to generate factor scores to serve as dependent variables in our models (Kuhfeld & Soland, 2020). The loadings for the seven views of leadership items ranged from 0.85 to 0.93 (Cronbach's alpha = 0.95). The factor score for views of leadership was missing if the teacher did not respond to any one of the seven items (8%). The loadings for the three job satisfaction items ranged from 0.55 to 1.04 (Cronbach's alpha = 0.78). This factor was missing for teachers who did not respond to any one of the three items (6%).

Turnover. Between November of 2020 and May of 2021, a roughly six-month period, our partners contacted each center leader up to three times to check on each teacher's employment status (i.e., whether they were still employed at their site, and whether they were working 30 hours a week or more). Using this information, we constructed a turnover variable that captured whether or not a teacher was still working at their site six months after the study started. We coded turnover as 0 if the center leader reported that the teacher was still working at

the center at the end of the six-month period. We coded as 1 if at any of the three checkpoints the leader reported that the teacher was no longer working at the center. Turnover was coded as missing if our state partner was unable to collect employment data, which was the case for less than 5% of teachers. Turnover was also coded as missing for 33 teachers for whom leaders reported working fewer than 30 hours per week in the first or second check, because these teachers were not followed through May 2021. However, turnover was coded as 0 for eight teachers who leaders reported began working fewer than 30 hours in May 2021 (the third check), because they had not left their center.

Turnover Intentions. Given the missingness in our administrative turnover data, we complemented our analysis by analyzing teacher-reported *turnover intentions* from the fall survey as a separate, secondary outcome. Teachers were asked in fall 2020 how likely they were to be working at their current center in May 2021 on a four-point scale (*not likely, a little bit likely, moderately likely, and very likely*). Because most teachers (89%) selected moderately or very likely, we reverse coded and dichotomized this item, such that 1 indicated that teachers reported being not likely or a little bit likely to be working at their center. Turnover intentions were missing for 4.35% of teachers.

Covariates

Additional teacher, leader, and center characteristics may be correlated with racial/ethnic match and teachers' views of leadership, job satisfaction, and turnover. To account for this possibility, we include several of these characteristics as covariates in our regression models.

At the teacher level, we account for: age; gender; teacher's highest degree earned (high school diploma or less, Associate's degree, or Bachelor's degree or higher, with high school or less as the omitted group); whether a teacher holds a Child Development Associate (CDA)

credential; teacher role (1 = lead teacher); self-reported hourly wages; years of experience in early childhood (continuous); and age level taught (infant and toddler teacher or preschool-age teachers, with preschool-age as the omitted category).

Panel A of Table 1 provides descriptive information on teachers in our sample. About 75% were the lead teacher of their classroom. Just under half (46%) worked primarily with infants and toddlers. Almost all (98%) were female. Nearly one in four (23%) held a Bachelor's degree or higher, but most (63%) held a high school degree or less. One in four (25%) held a CDA. On average, teachers had worked in ECE for nine years, though more than one-third (36%) had three years of experience or less. On average, teachers' wages were $12.50 per hour.

At the leader level, we controlled for race/ethnicity, years of experience in ECE, and highest degree earned (high school degree or less, Associate's degree, Bachelor's degrees, or Master's degree or higher, with high school or less as the omitted category). Panel B of Table 1 presents these leader characteristics. Just over half had either a Bachelor's degree (33%) or a Master's degree or higher (18%). On average, leaders had 16.5 years of experience in ECE, and the majority (89%) had more than three years of experience.

At the center level, we controlled for the number of assistant and lead teachers employed and the number of enrolled children five and under. We also included indicators for each of the 15 PDG communities to account for local economic conditions. Panel C of Table 1 shows that the average center had about five assistant teachers and seven lead teachers and enrolled about 43 children five and under.

Table 1 also shows rates of missingness for all covariates. Missingness was less than 4% for all teacher characteristics except for years of experience (26%) and age group taught (31%). Missingness for leader education was less than 1%, and for experience, 13%. Missingness for

center characteristics ranged from 3% to 9%. We had full data for the PDG community variable.

Analysis

Our first research question was: how prevalent is teacher-leader racial/ethnic match, and does it vary by race/ethnicity? To address this, we first described the distribution of Black, Hispanic, and White teachers and leaders in our sample, then showed the prevalence of teacher-leader racial/ethnic match overall and separately for Black, Hispanic, and White teachers.

Our second research question was: how does teacher-leader racial/ethnic match relate to teacher job outcomes, including views of leadership, job satisfaction, and turnover, and does it vary by race/ethnicity? To begin, we compared the means for each outcome across teachers who did and did not experience racial/ethnic match with their leader. Although we used factor scores for our regression analyses (described later), we dichotomized individual survey items for mean comparisons, both for ease of interpretation and because most teachers responded affirmatively. We created binary outcomes, where 1 represents the affirmative—either agree/strongly agree or somewhat/very satisfied—and 0 represents the negative—disagreeing or not being satisfied. We used two-sample t-tests to determine whether the means for each outcome differed for teachers by racial/ethnic match, overall and separately for Black, Hispanic, and White teachers.

We next ran linear models that estimated the association between racial/ethnic match and our four outcomes of interest: the views of leadership factor score, the job satisfaction factor score, observed turnover, and turnover intentions. Our first model was:

$$Y_{ij} = \beta_0 + \beta_1 \text{RaceMatch}_{ij} + \beta_2 \text{Black}_i + \beta_3 \text{Hispanic}_i + \Gamma \boldsymbol{X} + \epsilon_{ij} \qquad \text{(Equation 1)}$$

where Y_{ij} is a job outcome of interest for teacher i at center j. The coefficient of interest, β_1, represents the association between teacher-leader racial/ethnic match and job outcomes, controlling for differences in each outcome across teacher racial/ethnic groups and all covariates.

We ran standard OLS models and present standardized betas for the factor score regressions; we ran linear probability models for turnover and turnover intentions, and interpreted β_1 as the average percentage-point change in the probability that a teacher turns over within six months.

The vector $\boldsymbol{X}$ represents the teacher and center covariates included in our model. We included all covariates in all models, as well as missing indicators. That is, for each covariate we included a set of dichotomous variables where 1 indicated missingness, 0 otherwise. To keep all observations in the model, we then replaced the missing values of continuous variables (age, wages, and years of experience) with the mean for each variable, and replaced missing categorical covariates with an added level. These changes allowed the indicator variables to control for any teacher, leader, or center characteristics that may be associated with missingness. Standard errors were clustered at the center level.

Next, we explored whether the associations between teacher-leader racial/ethnic match and job outcomes varied by race for each of our outcomes. Similar to Grissom and Keiser (2011), we allowed the effect of teacher-leader racial/ethnic match to vary across Black, Hispanic, and White teachers by including interaction terms between teacher-leader racial/ethnic match and teacher race/ethnicity, with the largest racial group, White, omitted:

$$Y_{ij} = \beta_0 + \beta_1 \text{RaceMatch}_{ij} + \beta_2 \text{Black}_i + \beta_3 \text{Hispanic}_i + \beta_4 \text{RaceMatch*Black}_{ij} + \beta_5 \text{RaceMatch*Hispanic}_{ij} + \Gamma \boldsymbol{X} + \epsilon_{ij} \quad \text{(Equation 2)}$$

With the addition of the interaction terms, β_1 now reflects the association between racial/ethnic match and job outcomes for the omitted group (White teachers); β_1 is the difference in the outcome for White teachers who experienced match compared to those who did not. The difference in the outcome for Black teachers who experienced match compared to Black teachers who did not is represented by $\beta_1 + \beta_4$. For Hispanic teachers, this association is $\beta_1 + \beta_5$.

Including Hispanic teachers provided enough degrees of freedom to estimate associations with match separately for Black and White teachers, but we did not interpret the coefficient for Hispanic teachers due to small sample size (only 51 Hispanic teachers experienced match).

Our center-level covariates likely do not account for all center characteristics that could be correlated with both racial/ethnic match and outcomes. To account for this, we also ran center fixed effects models, which control for any center-level factor that impacts all teachers in the same way. These models allowed us to compare the outcomes of teachers with and without racial/ethnic match who worked at the same center.

By necessity, the center fixed effects models held the leader's race constant—all teachers within a center have the same leader and the same leader race. The variation in racial/ethnic match in these models came solely from comparing teachers of different races/ethnicities who work at the same center. Since teacher race fully determined racial/ethnic match in the fixed effects regressions, we dropped the teacher race indicators and the interaction terms from Equation 2. Likewise, we could not control for director and center characteristics, which are constant within center; we did, however, include the same teacher and job characteristics as in models 1 and 2, because they varied within centers. We estimated the following model:

$$Y_{ij} = \beta_0 + \beta_1 \text{RaceMatch}_{ij} + \Gamma \boldsymbol{X} + \alpha_j + \epsilon_{ij} \quad \text{(Equation 3)}$$

where the vector $\boldsymbol{X}$ represents the remaining covariates and α_j represents the center fixed effects.

We examined the relationship between teacher-leader racial/ethnic match and job outcomes first overall (that is, in the full sample) and then in subsamples split by leader race (e.g., Grissom & Keiser, 2011; Viano & Hunter, 2017). In the full sample, β_1 is the relationship between match and outcomes, pooled across all teacher and leader races.

Finally, we split our sample by leader race and re-estimated Equation 3 within each

subsample (e.g., Black leader, White leader). In splitting the sample by leader race, we retained within-center variability in teacher race and thus could assess whether associations varied by teacher race. In the Black leader subsample models, β_1 is the relationship between match and job outcomes for Black teachers, that is, it is the difference in outcomes between Black teachers at this site and Hispanic and White teachers; in the White leader subsample, it is the difference for White teachers as compared to Black and Hispanic teachers.

Results

Prevalence of Teacher-Leader Racial/Ethnic Match

We first documented the prevalence of teacher-leader racial/ethnic match overall and by racial/ethnic group. Table 2 shows that 27% of teachers in our sample were Black, 14% were Hispanic, and 59% were White. Leaders in our sample were more likely to be White (66%) and less likely to be Hispanic (6%) than sample teachers. Overall, 66% of teachers experienced racial/ethnic match with their leader, though rates varied by race/ethnicity. Just 42% of Black teachers and 36% of Hispanic teachers experienced match, compared to 84% of White teachers.

Relationships between Racial/Ethnic Match and Views of Leadership

Mean Comparisons: Views of Leadership

Prior to looking at the role of racial/ethnic match, we examined mean outcomes overall and by teacher/race ethnicity (Table 3, Panel A). Overall, teachers viewed their leaders positively: most felt respected by their leader (81%) and supported in managing child behavior (79%); affirmed their leader's commitment to an anti-racist learning environment (90%); and agreed that their leader supports professional development (82%). Black teachers were less likely than White teachers to feel supported in managing challenging behavior and to agree that their leader supports professional development. Black and Hispanic teachers were less likely than

White teachers to agree that their leader was committed to an anti-racist learning environment.

We next explored the relationship between teacher-leader racial/ethnic match and job outcomes. First, we used t-tests to compare unadjusted differences in individual item means for teachers who did and did not experience racial/ethnic match (Table 4). In the full sample, across all outcomes, teachers who experienced match reported more positive views of leadership than teachers who did not, but differences were statistically significant only for two of eight variables: feeling supported by the leader in managing challenging behavior and agreeing that the leader was committed to an anti-racist learning environment. By race, Black teachers who experienced racial/ethnic match reported more favorable views for all outcomes, but differences were only significant in two cases: feeling supported by the leader in managing challenging behavior (82% compared to 69%) and agreeing that the leader was committed to an anti-racist learning environment (90% compared to 77%). Hispanic teachers who experienced racial/ethnic match also reported more favorable views of leaders for all but one outcome. These differences were not statistically significant for any outcome, although our sample sizes here are small. There were no statistically significant differences for White teachers.

Regression Results: Views of Leadership

Table 5 presents results from regression models predicting the views of leadership factor score. Models include all covariates and regional indicators. Columns 1 presents results from Equation 1, which estimated the association between teacher-leader racial/ethnic match and job outcomes for all teachers. The match coefficients in column 1 are statistically and practically insignificant, suggesting no relationship between match and views of leadership.

Columns 2 present the results from Equation 2, which added interactions between teacher race/ethnicity and racial/ethnic match. These interactions test whether the associations between

match and views of leadership vary by teacher race. We again found no significant relationships.

Columns 3-5 show results from center fixed effects models in the full sample (Column 3), the Black leaders subsample (Column 4), and the White leaders subsample (Column 5). Recall that center fixed effects models estimate the relationship between racial/ethnic match and outcomes *within* a center. For this reason, each of these models omitted teachers in centers where teacher survey takers were racially homogenous, that is, where there was no within-center variability in match. In the full sample model (Column 3), this amounted to 327 teachers (N= 94 centers) omitted. These 94 centers were smaller than the heterogenous centers, so they accounted for less than half of our teacher sample (32%). Additionally, we cannot know if these centers are in fact racially homogenous or if teachers of other races/ethnicities did not respond to the survey. The racially heterogenous centers included 684 teachers who contributed to fixed effects estimates (also N=94 centers). Findings remained null in these models (Equation 3).

Relationships between Racial/Ethnic Match and Job Satisfaction and Turnover

Before analyzing associations among match, job satisfaction, and turnover, we examined dichotomized outcome means both overall and by teacher/race ethnicity (Panel B of Table 3). More than 80% of teachers responded affirmatively to all three satisfaction outcomes. About 16% of the teachers in our sample left their center over a six-month period and 11% reported that they were not likely to stay at their centers. There were no statistically significant differences in these outcomes between Black and White or between Hispanic and White teachers.

Mean Comparisons: Job Satisfaction and Turnover

Table 6 presents means and t-tests for our job satisfaction and turnover measures by experience of racial/ethnic match. We found no statistically significant differences in job satisfaction and turnover by racial/ethnic match, either overall or by race/ethnicity.

Regression Results: Job Satisfaction and Turnover

Tables 7, 8, and 9 present regression results for job satisfaction, observed turnover, and turnover intentions, respectively. We found no statistically significant associations between racial/ethnic match and any outcome across specifications, overall or by teacher race/ethnicity.

Discussion

Child care leaders can serve as source of support for early educators, and many policymakers consider effective leadership a promising approach to strengthening the workforce and ultimately better supporting young children (IOM & NRC, 2015). The empirical literature on leadership in child care settings is underdeveloped, however, and there is little research linking leader characteristics to teachers' views of their leaders, job satisfaction, or turnover. In particular, while research has established the importance of teacher-leader racial/ethnic match for teachers working in K-12 schools, no prior study has explored the role of match in child care settings. Given the large share of teachers of color working in child care (Austin et al., 2019; Bassok et al., 2020), and the high rate of turnover among these educators (Bassok, Markowitz, et al., 2021), it is worthwhile to examine teacher-leader racial/ethnic match in this context.

Using a sample of over 1,000 early educators in Virginia, we found that two-thirds of child care teachers experience racial/ethnic match, but that match was far more common for White teachers (84%) than Black (42%) or Hispanic (36%) teachers. This is on par with national data on match in K-12, where 83% of White teachers experienced principal racial/ethnic match compared to 43% of Black teachers (Viano & Hunter, 2017). The particularly low match rate for Hispanic teachers in our sample likely reflects that just 6% of leaders were Hispanic.

In our sample, most teachers reported they were satisfied and viewed their leaders positively; we found limited evidence that teachers of color viewed their leaders more favorably

when they share their race/ethnicity. Four out of five teachers felt respected by their leader. The only significant differences by racial/ethnic match were that Black teachers who experienced match were more likely to report that their leaders support them in managing challenging behavior and show commitment to an anti-racist learning environment. After adjusting for covariates, we found no evidence that teacher-leader racial/ethnic match was linked to any of our job outcomes in linear or fixed effects models

Exploring Null Findings

Our null findings were contrary to both our hypotheses and findings from K-12 settings (e.g., Bartanen & Grissom, 2021; Grissom & Keiser, 2011; Lindsay & Egalite, 2020). Below we explore plausible explanations for our null findings and pose aims for future research. We first consider how features of child care settings and the timing of our data relative to the COVID-19 pandemic may have led to a weaker relationship between racial/ethnic match and teacher job outcomes than observed in K-12 settings; then we consider the study limitations and whether our method was able to detect relationships should they exist.

Weak Link Between Racial/Ethnic Match and Outcomes

One possibility is that racial/ethnic match between teachers and leaders may be less salient in child care centers, in part because of the higher level of racial/ethnic diversity in the workforce (Austin et al., 2019; Greenberg & Luetmer, 2022; Paschall et al., 2020). That is, if child care teachers of color have a large number of colleagues of color, perhaps having a leader of color is less important for job outcomes. Through the lens of representative bureaucracy theory, child care leaders of color may not provide unique support to teachers of color or share notably different communication styles in this context. Indeed, Black K-12 teachers experience especially high turnover, lower satisfaction, and more biased teacher observations when they are

racially isolated, with few Black colleagues (Grissom & Bartanen, 2022; Kemper Patrick & Arturo Santelli, 2022; Ravenell et al., 2018)—and our data suggest that far fewer child care teachers experience this racial isolation. The greater racial diversity in child care, paired with a potential decline in the salience of racial/ethnic match in K-12 since the early 2000s (Viano & Hunter, 2017) may lead to a lack of associations relative to prior K-12 studies.

Another possibility is that the low wages and high turnover common to child care settings make teacher-leader racial/ethnic match less predictive of job outcomes. Nationally, teachers of children ages 0-5 are paid substantially less than K-8 school-based teachers, despite similar job demands, and they are nearly eight times more likely to live in poverty (McLean et al., 2021). These wages may be a main driver of our key job outcomes. Indeed, we did find a consistent, stable association between higher wages and more positive views of leadership across four of our five specifications, as well as some evidence of an association between wages and turnover intentions. We also found that teachers' age consistently predicted lower turnover, such that a teacher's life stage may be salient for job outcomes. Of course, these associations are only suggestive and warrant research explicitly focused on probing them more carefully. Finally, it may be that for new teachers—of which child care has a high share—leader racial-ethnic match is a less salient characteristic, as suggested by some K-12 research (Lindsay & Egalite, 2020). Future research could test these hypotheses.

In a similar vein, our data were collected in 2020, when the COVID-19 pandemic dramatically destabilized the child care sector, leading to unprecedented challenges with staffing (Bassok, Smith, et al., 2021). The recency of our data enhances its policy relevance, yet the pandemic may have attenuated associations that would have been observed at a different historical time, particularly given the differential impacts of COVID-19 across racial groups

(Centers for Disease Control and Prevention, 2022). Our universally favorable views of leaders and high levels of job satisfaction may have been a by-product of the pandemic-related challenges child care centers faced, and the need for teachers to rationalize their continued work amid the pandemic. This may have eliminated match-linked gaps in teachers' views of their leaders. Likewise, the pandemic may have loomed large in turnover decisions, drowning out other factors, again highlighting the importance of future research.

Notably, we were able to analyze similar data on teachers and leaders from the same research policy partnership in 2019—that is, *prior to the COVID-19 pandemic*. This sample is much smaller (619 teachers) and is only suitable as a robustness check. However, consistent with the results of the present study, we found no significant associations between racial/ethnic match and job outcomes in the pre-pandemic data.

Methodological Limitations

It is also possible that associations between racial/ethnic match and teacher job outcomes do exist in child care, and are of a similar magnitude as in K-12, but that we were underpowered to detect them. A key limitation of the current study is that we have a small sample compared to K-12 research, including a small number of Black teachers and Black leaders. Similarly, we were unable to estimate specific relationships at all for Hispanic teachers or teachers from other racial/ethnic groups, including any Asian or Pacific Islander groups. We had only 4 Asian or Pacific Islander leaders and 22 teacher survey respondents at their centers who would have met our inclusion criteria—fewer than the 11 Hispanic leaders with 112 teacher survey respondents at their sites in our sample. A recurring issue in the racial/ethnic match literature is its inability to conduct estimates for these racial and ethnic groups. Future research should focus on gathering large, diverse samples to better understand the potential role of leader diversity in supporting

both teachers and high quality ECE. Growing calls for better administrative data systems in early education (e.g., Whitebook et al., 2018) could enhance the potential for such research.

Uncertainty in measuring views of leadership and racial/ethnic match for teachers with multiple leaders might also have attenuated the associations we observed. Although the proportion of teachers with multiple leaders was modest (20%)—and the proportion of teachers with multiple leaders where the leaders are of different races is even smaller (9%)—we cannot know if teachers' survey responses corresponded to the leader we retained for our analysis (i.e., the leader with the highest responsibility). If we incorrectly coded match, this would bias our estimates toward zero. To address this concern, we conducted a robustness test where we dropped teachers with multiple leaders and reran all models. Findings did not change.

This study also only examined a small set of outcomes. Future research should explore associations between teacher-leader racial/ethnic match and outcomes that may be more closely tied to children's ECE experiences such as teacher self-efficacy, stress, and the quality of teacher-child interactions (Silver & Zinsser, 2020; Thomason & La Paro, 2013). Likewise, outcomes such as disproportionate exclusionary discipline across racial/ethnic groups might be examined in relation to racial/ethnic match between both teachers and leaders and leaders and families. In addition, future research with larger samples may be able to separately examine associations between outcomes and teacher-leader racial/ethnic match by leader type (e.g., director versus owner), exploring how different roles and staff interactions may matter. Future research might also expand to other program types (e.g., school or Head Start).

Finally, we did not find associations between teacher-leader racial/ethnic match and 6-month turnover but see potential for future research on match and staffing over a longer time horizon and broader set of outcomes. We found a 16% turnover rate over a short period that did

not include the summer months. To put this turnover in context, prior research found that a similar short-term turnover rate in child care settings (18% from fall 2016 to spring 2017) more than doubled to 41% by the following fall (Bellows et al., 2021). Prior research in K-12 found not only that turnover was lower for teachers of color with same-race principals, but also that principals of color were more likely to hire teachers of color (Bartanen & Grissom, 2021; Grissom & Keiser, 2011). Turnover is just one part of workforce development; it is possible that Black and Hispanic child care leaders may be more likely to recruit teachers of color than White leaders. It is also possible that having a leader of color may inspire teachers of color to envision themselves as a leader or help them stay in the field over a longer period. Exploring multiple points along the child care staffing pipeline is an important direction for future research.

Implications and Conclusions

The mounting evidence that teacher-child racial/ethnic match benefits children suggests that supporting the racial diversity of the childcare workforce is important (Downer et al., 2016; Markowitz et al., 2020; Meek et al., 2020a; Vinopal, 2018). Our results do not provide evidence for associations between teacher-leader racial/ethnic match and job outcomes, but many open questions remain. As the first study on teacher-leader racial/ethnic match in child care, it would be premature to conclude that match does not matter. Moreover, creating opportunities for more racial/ethnic representation in leadership may be inherently valuable or matter for outcomes that we did not examine, such as teacher-child interactions, child outcomes, long-term retention, or leadership development for teachers of color. More research is needed to better understand how match may support teachers and children, and to identify ways to enhance diversity in child care.

At the same time, policymakers can work to improve the capacity of current leaders to support the diverse teachers already working in child care settings. We found that most Black

and Hispanic teachers did not have leaders who shared their race/ethnicity. Before adjusting for covariates, Black teachers who experienced match were more likely to report their leader was committed to an anti-racist learning environment. Efforts to support White leaders' cultural competency, perhaps through training and professional development, may be beneficial.

Policymakers should also focus on addressing other predictors of job satisfaction, especially in light of COVID-19. We found no association between teacher-leader race match and teacher turnover after six months, but we did observe a 16% turnover rate. This exceeds the 10.5% one-year turnover rate for K-12 teachers in Virginia, despite covering a considerably shorter time period (Virginia Department of Education, 2021). This turnover undermines stable and high-quality care, and was likely exacerbated by the COVID-19 pandemic (Bassok, Smith, et al., 2021). Research in Virginia suggests that compensation levels predict turnover (Bassok, Hall, et al., 2021), and experimental evidence shows that financial incentives can significantly reduce turnover in child care centers (Bassok, Doromal, et al., 2021). Moreover, in the present study, we found that wages predict teachers' views of their leaders. Efforts to restructure compensation—and to address wage inequities specifically faced by early educators of color (Austin et al., 2019)—with continued investigation of how to best support teachers are likely essential.

Together, findings suggest that policymakers should invest in efforts to help all leaders engage effectively with the diverse ECE workforce, and that efforts should prioritize communities of color, including centers owned by Black and Hispanic women and those serving a high proportion of Black and Hispanic children.

References

Accavitti, M. R., & Williford, A. P. (2020). Teacher perceptions of externalizing behaviour subtypes in preschool: Considering racial factors. *Early Child Development and Care*, *0*(0), 1–15. https://doi.org/10.1080/03004430.2020.1825405

Andersen, L. B., Bjørnholt, B., Bro, L. L., & Holm-Petersen, C. (2018). Achieving high quality through transformational leadership: A qualitative multilevel analysis of transformational leadership and perceived professional quality. *Public Personnel Management*, *47*(1), 51–72. https://doi.org/10.1177/0091026017747270

Austin, L. J. E., Edwards, B., Chávez, R., & Whitebook, M. (2019, December 19). Racial wage gaps in early education employment. *Center for the Study of Child Care Employment*. https://cscce.berkeley.edu/racial-wage-gaps-in-early-education-employment/

Bartanen, B., & Grissom, J. A. (2021). School principal race, teacher racial diversity, and student achievement. *Journal of Human Resources*, 0218. https://doi.org/10.3368/jhr.58.4.0218-9328R2

Bassok, D., Doromal, J. B., Holland, A., & Michie, M. (2020). *Who teaches Virginia's youngest children?* EdPolicyWorks at the University of Virginia. https://vecf.org/wp-content/uploads/2021/06/SEE_Partnerships_VAPDG-Report_Racial-Composition_Revised.pdf

Bassok, D., Doromal, J. B., Michie, M., & Wong, V. C. (2021). *The effects of financial incentives on teacher turnover in early childhood settings: Experimental evidence from Virginia*. EdPolicyWorks at the University of Virginia. https://files.elfsightcdn.com/022b8cb9-839c-4bc2-992e-cefccb8e877e/6de6fd54-e921-4c88-a452-ad7cabccc362.pdf

Bassok, D., Hall, T., Markowitz, A. J., & Doromal, J. B. (2021). *Teacher turnover in child care: Pre-pandemic evidence from virginia* (p. 15). EdPolicyWorks at the University of Virginia. https://files.elfsightcdn.com/022b8cb9-839c-4bc2-992e-cefccb8e877e/86b10e35-84cf-4501-9aa8-2aed26adfa0f.pdf

Bassok, D., Markowitz, A. J., Bellows, L., & Sadowski, K. (2021). New Evidence on Teacher Turnover in Early Childhood. *Educational Evaluation and Policy Analysis*, *43*(1), 172–180. https://doi.org/10.3102/0162373720985340

Bassok, D., Smith, A. E., Markowitz, A. J., & Doromal, J. B. (2021). *Child care staffing challenges during the pandemic: Lessons from child care leaders in Virginia*. https://files.elfsightcdn.com/022b8cb9-839c-4bc2-992e-cefccb8e877e/a8274e3d-67bd-4abf-83e1-a8f99274df25.pdf

Bates, L. A., & Glick, J. E. (2013). Does it matter if teachers and schools match the student? Racial and ethnic disparities in problem behaviors. *Social Science Research*, *42*(5), 1180–1190. https://doi.org/10.1016/j.ssresearch.2013.04.005

Bellows, L., Bassok, D., & Markowitz, A. J. (2021). *Teacher Turnover in Early Childhood Education: Longitudinal Evidence from the Universe of Publicly-Funded Programs in Louisiana*. https://doi.org/10.26300/F9BZ-FS97

Brezicha, K. F., & Fuller, E. J. (2019). Building teachers' trust in principals: Exploring the effects of the match between teacher and principal race/ethnicity and gender and feelings of trust. *Journal of School Leadership*, *29*(1), 25–53. https://doi.org/10.1177/1052684618825087

Cassidy, D. J., Lower, J. K., Kintner-Duffy, V. L., Hegde, A. V., & Shim, J. (2011). The Day-to-Day Reality of Teacher Turnover in Preschool Classrooms: An Analysis of Classroom

Context and Teacher, Director, and Parent Perspectives. *Journal of Research in Childhood Education*, *25*(1), 1–23. https://doi.org/10.1080/02568543.2011.533118

Centers for Disease Control and Prevention. (2022, January 25). *Health equity*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html

Dee, T. S. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American Economic Review*, *95*(2), 158–165.

Doromal, J. B., & Markowitz, A. J. (2021). Following the leader: Associations between leader support and teacher retention in child care settings. *[Manuscript Submitted for Publication]. School of Education and Human Development. University of Virginia.*

Downer, J. T., Goble, P., Myers, S. S., & Pianta, R. C. (2016). Teacher-child racial/ethnic match within pre-kindergarten classrooms and children's early school adjustment. *Early Childhood Research Quarterly*, *37*, 26–38. https://doi.org/10.1016/j.ecresq.2016.02.007

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, *77*(4), 267–282. https://doi.org/10.1177/003804070407700401

Ehrlich, S. B., Pacchiano, D., Stein, A. G., Wagner, M. R., Park, S., Frank, E., Luppescu, S., & Young, C. (2019). Early education essentials: Validation of surveys measuring early education organizational conditions. *Early Education and Development*, *30*(4), 540–567.

Goff, P., Rodriguez-Escutia, Y., & Yang, M. (2018). Through the labor market looking glass: An inquiry into principal-teacher race congruence. In *Wisconsin Center for Education Research* (No. 2018–13; WCER Working Paper). Wisconsin Center for Education Research. https://eric.ed.gov/?id=ED593630

Granja, M. R., Smith, S., Nguyen, U. S., & Grifa, B. (2018). *Learning about Young Children's Challenging Behavior and Impacts on Programs and Families: A State-wide Survey of Virginia's Early Care and Education Teachers*. 18.

Greenberg, E., & Luetmer, G. (2022). *State policies shape the racial and ethnic diversity of the prekindergarten workforce*. Urban Institute. https://www.urban.org/sites/default/files/2022-04/State%20Policies%20Shape%20the%20Racial%20and%20Ethnic%20Diversity%20of%20the%20Prekindergarten%20Workforce.pdf

Grissom, J. A., & Bartanen, B. (2022). Potential race and gender biases in high-stakes teacher observations. *Journal of Policy Analysis and Management*, *41*(1), 131–161. https://doi.org/10.1002/pam.22352

Grissom, J. A., & Jones, A. (2020). Racial and ethnic diversity in the public sector workforce: Insights from public education. In *Race and Public Administration*. Routledge.

Grissom, J. A., & Keiser, L. R. (2011). A Supervisor Like Me: Race, Representation, and the Satisfaction and Turnover Decisions of Public Sector Employees. *Journal of Policy Analysis and Management*, *30*(3), 557–580.

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The "representative bureaucracy" in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*. https://doi.org/10.3102/0013189X15580102

Huang, C. Y., & Zane, N. (2016). Cultural influences in mental health treatment. *Current Opinion in Psychology*, *8*, 131–136. https://doi.org/10.1016/j.copsyc.2015.10.009

Institute of Medicine and National Reserach Council. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. https://doi.org/10.17226/19401

Jeon, L., & Wells, M. B. (2018). An organizational-level analysis of early childhood teachers' job attitudes: Workplace satisfaction affects Early Head Start and Head Start teacher turnover. *Child & Youth Care Forum*, *47*(4), 563–581. https://doi.org/10.1007/s10566-018-9444-3

Johnson-Staub, C. (2017). *Equity starts early: Addressing racial inequities in child care and early education policy*. Center for Law and Social Policy. https://www.clasp.org/publications/report/brief/equity-starts-early-addressing-racial-inequities-child-care-and-early

Kemper Patrick, S., & Arturo Santelli, F. (2022). Exploring the relationship between demographic isolation and professional experiences of black and latinx teachers. *Journal of Education Human Resources*, *40*(2), 138–168. https://doi.org/10.3138/jehr-2021-0042

Kuhfeld, M., & Soland, J. (2020). Avoiding bias from sum scores in growth estimates: An examination of IRT-based approaches to scoring longitudinal survey responses. *Psychological Methods*, *27*(2), 234. https://doi.org/10.1037/met0000367

Kwon, K.-A., Malek, A., Horm, D., & Castle, S. (2020). Turnover and retention of infant-toddler teachers: Reasons, consequences, and implications for practice and policy. *Children and Youth Services Review*, *115*, 105061. https://doi.org/10.1016/j.childyouth.2020.105061

LeeKeenan, D., & Chin Ponte, I. (2018). What does it mean to be a director? In *From Survive to Thrive: A Director's Guide for Leading an Early Childhood Program* (pp. 1–12). National Association for the Education of Young Children. https://www.naeyc.org/resources/pubs/books/from-survive-to-thrive

Lim, H.-H. (2006). Representative bureaucracy: Rethinking substantive effects and active representation. *Public Administration Review*, *66*(2), 193–204. https://doi.org/10.1111/j.1540-6210.2006.00572.x

Lindsay, C., & Egalite, A. J. (2020). *The effects of principal-teacher demographic matching on teacher turnover in North Carolina (working paper)*. https://www.newschools.org/wp-content/uploads/2020/10/Constance-Lindsay-Research-Paper-1.pdf

Markowitz, A. J. (2019). *Within-Year Teacher Turnover in Head Start and Children's School Readiness (Working Paper)*. https://curry.virginia.edu/sites/default/files/uploads/epw/70_Teacher_Turnover_in_Head_Start.pdf

Markowitz, A. J., Bassok, D., & Grissom, J. A. (2020). Teacher-Child Racial/Ethnic Match and Parental Engagement With Head Start. *American Educational Research Journal*, *57*(5), 2132–2174. https://doi.org/10.3102/0002831219899356

Martin, K. A., Bosk, E., & Bailey, D. (2018). Teachers' Perceptions of Childcare and Preschool Expulsion. *Children & Society*, *32*(2), 87–97. https://doi.org/10.1111/chso.12228

McCurdy, K., Gannon, R. A., & Daro, D. (2003). Participation patterns in home-based family support programs: Ethnic variations. *Family Relations*, *52*(1), 3–11. https://doi.org/10.1111/j.1741-3729.2003.00003.x

McDonald, P., Thorpe, K., & Irvine, S. (2018). Low pay but still we stay: Retention in early childhood education and care. *Journal of Industrial Relations*, *60*(5), 647–668. https://doi.org/10.1177/0022185618800351

McLean, C., Austin, L. J. E., Whitebook, M., & Olson, K. L. (2021). *Early childhood workforce index 2020*. Center for the Study of Child Care Employment, University of California, Berkeley. https://cscce.berkeley.edu/workforce-index-2020/

Meier, K. J., & Stewart Jr, J. (1992). The impact of representative bureaucracies: Educational systems and public policies. *The American Review of Public Administration*, *22*(3), 157–171.

Miller, L. C. (2020). *2019 Virginia Working Conditions survey: Descriptive analysis of statewide patterns*. University of Virginia. https://education.virginia.edu/sites/default/files/uploads/epw/2019%20Virginia%20Working%20Conditions%20Survey.pdf

Muijs, D., Aubrey, C., Harris, A., & Briggs, M. (2004). How do they manage?: A review of the research on leadership in early childhood. *Journal of Early Childhood Research*, *2*(2), 157–169. https://doi.org/10.1177/1476718X04042974

NAEYC. (2019). *Leading with equity—Early childhood educators make it personal: Summary of the leading and working toward equity leadership summit*. National Association for the Education of Young Children. https://www.naeyc.org/sites/default/files/globally-shared/downloads/PDFs/our-work/initiatives/equity_summit_final.pdf

Olsen, A. A., & Huang, F. L. (2018). Teacher job satisfaction by principal support and teacher cooperation: Results from the Schools and Staffing Survey. *Education Policy Analysis Archives*, *27*(11).

Paschall, K., Madill, R., & Halle, T. (2020). *Professional characteristics of the early care and education workforce: Descriptions by race, ethnicity, languages spoken, and nativity status* (OPRE Report #2020-107.). Office of Planning, Research, and Evaluation,

Administration for Children and Families, U.S. Department of Health and Human Services.

Ravenell, A., Grissom, J. A., & Bartanen, B. (2018). *Exploring turnover and retention patterns among Tennessee's teachers of color: A research brief on strengthening Tennessee's education labor market*. Tennessee Education Research Alliance. https://peabody.vanderbilt.edu/TERA/files/Retention_Patterns_Among_Teachers_of_Color_FINAL.pdf

Sandstrom, H., & Schilder, D. (2021). *Strengthening the diversity and quality of the early care and education workforce* (p. 54). Urban Institute. https://www.urban.org/sites/default/files/publication/104998/strengthening-the-diversity-and-quality-of-the-early-care-and-education-workforce.pdf

Selden, S. C. (1997). Representative bureaucracy: Examining the linkage between passive and active representation in the farmers home administration. *The American Review of Public Administration*, *27*(1), 22–42.

Silver, H. C., & Zinsser, K. M. (2020). The Interplay among Early Childhood Teachers' Social and Emotional Well-Being, Mental Health Consultation, and Preschool Expulsion. *Early Education and Development*, *31*(7), 1133–1150. https://doi.org/10.1080/10409289.2020.1785267

Sims, M., Waniganayake, M., & Hadley, F. (2019). What makes good even better? Excellent EC leadership. *International Journal of Educational Management*, *33*(4), 573–586. https://doi.org/10.1108/IJEM-01-2018-0032

Spiegelman, M. (2020). Race and ethnicity of public school teachers and their students. Data point. NCES 2020-103. In *National Center for Education Statistics*. National Center for Education Statistics. https://eric.ed.gov/?id=ED607758

Taie, S., & Goldring, R. (2020). Characteristics of public and private elementary and secondary school teachers in the United States: Results from the 2017-18 national teacher and principal survey. First look. NCES 2020-142. *National Center for Education Statistics*.

Thomason, A. C., & La Paro, K. M. (2013). Teachers' commitment to the field and teacher–child interactions in center-based child care for toddlers and three-year-olds. *Early Childhood Education Journal*, *41*(3), 227–234. https://doi.org/10.1007/s10643-012-0539-4

Tran, H., & Winsler, A. (2011). Teacher and center stability and school readiness among low-income, ethnically diverse children in subsidized, center-based child care. *Children and Youth Services Review*, *33*(11), 2241–2252. https://doi.org/10.1016/j.childyouth.2011.07.008

Viano, S., & Baker, D. J. (2020). How administrative data collection and analysis can better reflect racial and ethnic identities. *Review of Research in Education*, *44*(1), 301–331. https://doi.org/10.3102/0091732X20903321

Viano, S. L., & Hunter, S. B. (2017). Teacher-principal race and teacher satisfaction over time, region. *Journal of Educational Administration*, *55*(6), 624–639. https://doi.org/10.1108/JEA-10-2016-0122

Vinopal, K. (2018). Understanding individual and organizational level representation: The case of parental involvement in schools. *Journal of Public Administration Research and Theory*, *28*(1), 1–15. https://doi.org/10.1093/jopart/mux036

Virginia Department of Education. (2021). *Annual report on the condition & needs of public schools in Virginia.* https://www.doe.virginia.gov/boe/reports/index.shtml

Virginia Department of Education. (2022). *VDOE:: Education workforce data & reports: 2020-2021 Virginia educator ethnicity and race data*. https://www.doe.virginia.gov/teaching/workforce_data/index.shtml

Wells, M. B. (2015). Predicting preschool teacher retention and turnover in newly hired Head Start teachers across the first half of the school year. *Early Childhood Research Quarterly*, *30*, 152–159. https://doi.org/10.1016/j.ecresq.2014.10.003

Whitebook, M., McLean, C., & Austin, L. J. E. (2018). *The workforce data deficit* (p. 20). https://cscce.berkeley.edu/the-workforce-data-deficit/

Whitebook, M., Philipps, D., & Howes, C. (2014, November 1). Worthy Work, STILL Unlivable Wages: The Early Childhood Workforce 25 Years after the National Child Care Staffing Study. *Center for the Study of Child Care Employment*. https://cscce.berkeley.edu/worthy-work-still-unlivable-wages/

Whitebook, M., & Sakai, L. (2003). Turnover begets turnover: An examination of job and occupational instability among child care center staff. *Early Childhood Research Quarterly*, *18*(3), 273–293. https://doi.org/10.1016/S0885-2006(03)00040-1

Wymer, S., C., Williford, A. P., & Lhospital, Ann. S. (2020). Exclusionary Discipline Practices in Early Childhood—ProQuest. *Young Children*, *75*(3), 36–44.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development. https://www.fcd-us.org/the-evidence-base-on-preschool/

Zinsser, K. M., Christensen, C. G., & Torres, L. (2016). She's supporting them; who's supporting her? Preschool center-level social-emotional supports and teacher well-being. *Journal of School Psychology*, *59*, 55–66. https://doi.org/10.1016/j.jsp.2016.09.001

Zinsser, K. M., & Curby, T. W. (2014). Understanding preschool teachers' emotional support as a function of center climate. *SAGE Open*, *4*(4), 2158244014560728. https://doi.org/10.1177/2158244014560728

Zulauf, C. A., & Zinsser, K. M. (2019). Forestalling Preschool Expulsion: A Mixed-Method Exploration of the Potential Protective Role of Teachers' Perceptions of Parents. *American Educational Research Journal*, *56*(6), 2189–2220. https://doi.org/10.3102/0002831219838236

Table 1. Teacher, Leader, and Center Characteristics

Panel A: *Teacher Characteristics*	Overall	Standard Deviation	% Missing
Lead Teacher	75.1%		0.2%
Infant/Toddler Teacher	46.1%		30.9%
Female	98%		0.4%
Age	36.8	13.5	3.9%
Education			0.49%
High School Degree or Less	63.4%		
Associate's Degree	13.8%		
Bachelor's degree or higher	22.8%		
Child Development Associates	25.1%		1.9%
Experience in ECE			26.2%
Years experience in ECE	8.5	8.63	
Less than one year	12.2%		
One to three years	24.0%		
More than three years	63.8%		
Hourly Wage ($)	$12.56	3.55	1.0%
N	1011		

Panel B: *Leader Characteristics*	Overall	Standard Deviation	% Missing
Education			0.5%
High School Degree or Less	23.5%		
Associate's Degree	26.2%		
Bachelor's Degree	32.6%		
Master's Degree or Higher	17.6%		
Experience in ECE			13.3%
Years experience in ECE	16.4	9.7	
Less than one year	1.8%		
One to three years	9.8%		
More than three years	88.3%		
Role			0%
Director	65.4%		
Director/Owner	17.6%		
Owner	10.1%		
Assistant Director/Assistant Principal	4.8%		
Principal	2.1%		
N	188		

Panel C: *Center Characteristics*	Overall	Standard Deviation	% Missing
Staffing			
Number of Assistant Teachers	5.0	4.17	9.0%
Number of Lead Teachers	6.8	4.3	2.7%
Enrolled Children Ages 0-5	42.7		3.7%
N	188		

Table 2. Teachers, Leaders, and Race Match by Race

	Overall	Black	Hispanic	White
Racial/Ethnic Distribution [n (%)]				
Teacher Sample	1011 (100%)	272 (27%)	140 (14%)	599 (59%)
Leader Sample	188 (100%)	52 (28%)	11 (6%)	125 (66%)
% of teachers with race match	66%	42%	36%	84%

Table 3. Views of Leadership, Satisfaction, and Turnover Outcomes by Teacher Race

	Overall	Black		Hispanic		White
Panel A	Mean	Mean	t-test significance (relative to White)	Mean	t-test significance (relative to White)	Mean
Views of Leadership (% agreed / strongly agreed)						
Satisfaction with interactions with leader	84.5%	86.4%		81.3%		84.4%
Feel respected by leader	81.1%	78.0%		82.7%		82.1%
Supported by leader in managing behavior	78.9%	74.8%	*	74.4%	+	81.8%
Trust leader	76.4%	72.3%	+	78.4%		77.8%
Leader communicates clear vision	76.7%	75.4%		76.1%		77.5%
Comfortable raising issues with leader	78.3%	75.9%		73.9%	+	80.4%
Leader committed to antiracist learning environment	89.8%	83.4%	**	87.2%	*	93.3%
Leader supports professional development	82.3%	76.9%	*	82.1%		84.8%
Panel B						
Job Satisfaction Overall job satisfaction (% somewhat or very satisfied)	89.6%	89.9%		89.5%		89.5%
Enjoy current job (% agreed /strongly agreed)	84.1%	84.3%		85.7%		83.7%
Turnover						
Turnover (% that left site over a six-month period)	15.9%	17.7%		11.5%		16.1%

** p<0.01, * p<0.05, + p<0.1.

Note: Characteristics are self-reported and come from the teacher survey. Satisfaction items (overall job satisfaction and satisfaction with interaction with leaders) were asked on a four-point scale (not at all satisfied to very satisfied), and views of leadership items were asked on five-point scale (strongly disagree to strongly agree).

Table 4. Views of Leadership by Teacher-Leader Racial/Ethnic Match

	Overall			Black			Hispanic			White		
Outcome	Without match	Match	t-test significance	Without match	Match	t-test significance	Without match	Match	t-test significance	Without match	Match	t-test significance
Views of Leadership (% agreed / strongly agreed)												
Satisfaction with interactions with leader	83.5%	84.9%		85.2%	87.2%		77.6%	87.8%		86.0%	84.1%	
Feel respected by leader	79.2%	81.9%		76.3%	79.8%		81.0%	85.7%		82.1%	82.0%	
Supported by leader in managing behavior	74.7%	80.9%	*	69.3%	82.0%	*	73.8%	75.5%		84.2%	81.3%	
Trust leader	74.5%	77.0%		68.4%	76.4%		80.0%	75.5%		79.2%	77.4%	
Leader communicates clear vision	75.4%	77.3%		74.3%	76.4%		72.9%	81.6%		79.2%	77.1%	
Comfortable raising issues with leader	74.5%	79.9%		71.1%	81.1%		69.4%	81.6%		84.4%	79.5%	
Leader committed to antiracist learning environment	85.2%	91.8%	**	77.3%	90.1%	**	86.9%	87.8%		95.8%	92.6%	
Leader supports professional development	79.5%	83.6%		75.0%	79.1%		80.0%	85.7%		86.3%	84.4%	
N	346	665		159	113		89	51		98	501	

** p<0.01, * p<0.05

Note: Characteristics are self-reported and come from the teacher survey. Satisfaction items (overall job satisfaction and satisfaction with interaction with leaders) were asked on a four-point scale (not at all satisfied to very satisfied), and views of leadership items were asked on five-point scale (strongly disagree to strongly agree).

Table 5. Regressions of Views of Leadership on Race/Ethnicity and Racial/Ethnic Match

VARIABLES	(1)	(2)	(3)	(4)	(5)
Match (any race)	0.0158	-0.0435	0.110	0.0170	0.135
	(0.0816)	(0.220)	(0.0909)	(0.221)	(0.112)
Black	-0.131	-0.204			
	(0.103)	(0.214)			
Hispanic	-0.167	-0.176			
	(0.118)	(0.195)			
Black Leader	0.0702	-0.0102			
	(0.0868)	(0.194)			
Hispanic Leader	-0.106	-0.0830			
	(0.180)	(0.198)			
Match x Black		0.191			
		(0.386)			
Match x Hispanic		-0.0513 (0.391)			
Male	0.152	0.152	0.121	-0.122	0.227
	(0.179)	(0.178)	(0.248)	(0.902)	(0.273)
Lead Teacher	-0.250	-0.251	-0.116	-0.231	-0.0357
	(0.149)	(0.151)	(0.155)	(0.460)	(0.178)
Bachelors degree or higher	0.0615	0.0570	-0.0110	0.0761	-0.0226
	(0.0774)	(0.0776)	(0.0904)	(0.265)	(0.105)
Associates degree	-0.0859	-0.0853	-0.0493	0.120	-0.0544
	(0.0972)	(0.0976)	(0.107)	(0.240)	(0.130)
Child Development Associates	0.0713	0.0702	0.102	0.455	0.0372
	(0.0923)	(0.0933)	(0.100)	(0.239)	(0.124)
Hourly Wage	0.0378**	0.0380**	0.0398*	-0.0363	0.0642**
	(0.0129)	(0.0129)	(0.0170)	(0.0490)	(0.0206)
Age	-0.000245	-0.000333	-0.00155	0.00232	-0.00558
	(0.00310)	(0.00312)	(0.00313)	(0.00754)	(0.00367)
Experience in ECE	-0.00465	-0.00447	-0.00387	-0.0211	-0.00125
	(0.00626)	(0.00631)	(0.00613)	(0.0148)	(0.00705)
Leaders experience in ECE	-0.00384	-0.00375			
	(0.00446)	(0.00443)			
Enrollment Children 0-5	-0.00353	-0.00366			
	(0.00229)	(0.00229)			
Number of Assistant Teachers	0.0203	0.0211*			
	(0.0105)	(0.0104)			
Number of Lead Teachers	-0.00826	-0.00765			
	(0.00854)	(0.00854)			
Leader's highest degree = Associates	0.0228	0.0260			
	(0.128)	(0.128)			
Leader's highest degree = Bachelors	0.00799	0.00251			
	(0.135)	(0.135)			
Leader's highest degree = Masters or higher	0.0131	0.0101			
	(0.135)	(0.136)			
Constant	-0.0241	0.0290	-0.428	0.633	-0.664*
	(0.336)	(0.397)	(0.250)	(0.698)	(0.294)
Observations	935	935	935	191	651
R-squared	0.079	0.079	0.327	0.364	0.353
Center Fixed Effects?	No	No	Yes	Yes	Yes
Centers Sample	Full	Full	Full	Black Leader	White Leader

Robust standard errors in parentheses
Significance levels: ** p<0.01, * p<0.05

Note: The dependent variable is a standardized factor score for holding a positive view of the center leader. The race/ethnicity variables
(Black and Hispanic) refer to the teacher's race. The regression models include the following center-level covariates: leader's race/ethnicity, leader's years of experience, leader's degree attainment, number of children ages 0-5 enrolled at the center, total number of lead teachers employed by the center, total number of assistant teachers, total number of lead teachers, community in the Preschool Development Grant (PDG), and corresponding missing indicators for each variable. The regression models also include the following individual-level covariates: whether the teacher is a lead teacher, whether the teacher teaches primarily infants and toddlers, degree attainment, whether a teacher has a Child Development Associates, their reported annual salary, their experience in early childhood education, their age, and corresponding missing indicators for each variable. The coefficients for community indicators and missing indicators are not shown above. Sample sizes vary due to missing outcome data for the 1011 teachers in the overall sample.

Table 6. Job Satisfaction and Turnover by Teacher-Leader Racial/Ethnic Match

	Overall			Black			Hispanic			White	
Outcome	Without match	t-test significance	Match	Without match	t-test significance	Match	Without match	t-test significance	Match	Without match	Match
Panel A											
Job Satisfaction											
Overall job satisfaction (% somewhat or very satisfied)	90.7%		89.0%	89.5%		89.8%	89.4%		89.6%	93.8%	88.7%
Enjoy current job (% agreed /strongly agreed)	83.8%		84.0%	81.7%		87.2%	84.7%		87.5%	86.3%	83.0%
Panel B											
Turnover											
Turnover (% that left site over a six-month period)	15.5%		16.4%	17.2%		18.9%	14.5%		6.3%	13.5%	16.8%
N	346		665	159		113	89		51	98	501

** p<0.01, * p<0.05

Note: Characteristics are self-reported and come from the teacher survey. Satisfaction items (overall job satisfaction and satisfaction with interaction with leaders) were asked on a four-point scale (not at all satisfied to very satisfied), and views of leadership items were asked on five-point scale (strongly disagree to strongly agree).

Table 7. Regressions of Job Satisfaction on Race/Ethnicity and Racial/Ethnic Match

VARIABLES	(1)	(2)	(3)	(4)	(5)
Match (any race)	-0.00769	-0.0354	0.0501	0.238	0.0311
	(0.0768)	(0.242)	(0.0917)	(0.210)	(0.115)
Black	0.00314	-0.0538			
	(0.0967)	(0.238)			
Hispanic	-0.216	-0.178			
	(0.123)	(0.226)			
Black Leader	0.0888	0.0168			
	(0.0962)	(0.223)			
Hispanic Leader	0.153	0.244			
	(0.187)	(0.235)			
Match x Black		0.182			
		(0.437)			
Match x Hispanic		-0.212			
		(0.433)			
Male	-0.0840	-0.0805	-0.0558	-0.129	-0.139
	(0.159)	(0.163)	(0.232)	(0.639)	(0.210)
Lead Teacher	-0.0656	-0.0644	-0.0296	0.695*	-0.0829
	(0.145)	(0.146)	(0.168)	(0.344)	(0.204)
Bachelors degree or higher	-0.0630	-0.0695	-0.0667	-0.0874	-0.0862
	(0.0832)	(0.0831)	(0.0906)	(0.256)	(0.106)
Associates degree	-0.154	-0.153	-0.116	-0.112	-0.113
	(0.0995)	(0.0998)	(0.111)	(0.222)	(0.140)
Child Development Associates	0.0911	0.0931	0.103	0.419	0.0747
	(0.0915)	(0.0925)	(0.0955)	(0.223)	(0.116)
Hourly Wage	0.0214	0.0213	0.0149	-0.0629	0.0286
	(0.0143)	(0.0143)	(0.0173)	(0.0462)	(0.0215)
Age	0.00577*	0.00563	0.00629*	0.00811	0.00303
	(0.00289)	(0.00291)	(0.00303)	(0.00677)	(0.00366)
Experience in ECE	0.00235	0.00251	0.00320	-0.0179	0.00755
	(0.00523)	(0.00524)	(0.00525)	(0.0125)	(0.00632)
Leaders experience in ECE	-0.00165	-0.00145			
	(0.00459)	(0.00453)			
Enrollment Children 0-5	-0.00419	-0.00443			
	(0.00229)	(0.00227)			
Number of Assistant Teachers	0.0142	0.0157			
	(0.0116)	(0.0114)			
Number of Lead Teachers	-0.00220	-0.00104			
	(0.00984)	(0.00979)			
Leader's highest degree = Associates	0.0519	0.0605			
	(0.123)	(0.123)			
Leader's highest degree = Bachelors	0.0591	0.0502			
	(0.136)	(0.137)			
Leader's highest degree = Masters or higher	0.0377	0.0349			
	(0.137)	(0.137)			
Constant	-0.174	-0.147	-0.501	-0.155	-0.528
	(0.321)	(0.394)	(0.260)	(0.574)	(0.330)
Observations	952	952	952	190	672
R-squared	0.079	0.080	0.315	0.412	0.322
Center Fixed Effects?	No	No	Yes	Yes	Yes
Centers Sample	Full	Full	Full	Black Leader	White Leader

Robust standard errors in parentheses

Significance levels: ** p<0.01, * p<0.05

Note: The dependent variable is a standardized factor score for teacher job satisfaction. The race/ethnicity variables (Black and Hispanic) refer to the teacher's race. The regression models include the following center-level covariates: leader's race/ethnicity, leader's years of experience, leader's degree attainment, number of children ages 0-5 enrolled at the center, total number of lead teachers employed by the center, total number of assistant teachers, total number of lead teachers, community in the Preschool Development Grant (PDG), and corresponding missing indicators for each variable. The regression models also include the following individual-level covariates: whether the teacher is a lead teacher, whether the teacher teaches primarily infants and toddlers, degree attainment, whether a teacher has a Child Development Associates, their reported annual salary, their experience in early childhood education, their age, and corresponding missing indicators for each variable. The coefficients for community indicators and missing indicators are not shown above. Sample sizes vary due to missing outcome data for the 1011 teachers in the overall sample.

Table 8. Regressions of Turnover Intentions on Race/Ethnicity and Racial/Ethnic Match

VARIABLES	(1)	(2)	(3)	(4)	(5)
Match (any race)	0.0142	0.0662	-0.0175	0.00233	-0.0352
	(0.0263)	(0.0791)	(0.0299)	(0.0815)	(0.0343)
Black	0.0272	0.0806			
	(0.0339)	(0.0753)			
Hispanic	0.0823	0.115			
	(0.0432)	(0.0868)			
Black Leader	0.0244	0.0776			
	(0.0313)	(0.0792)			
Hispanic Leader	0.0408	0.0632			
	(0.0433)	(0.0832)			
Match x Black		-0.120			
		(0.150)			
Match x Hispanic		-0.0517 (0.152)			
Male	-0.00805	-0.00624	-0.0574	-0.244	0.0288
	(0.0688)	(0.0702)	(0.0891)	(0.165)	(0.132)
Lead Teacher	-0.0687	-0.0685	-0.131**	-0.434**	-0.102*
	(0.0362)	(0.0359)	(0.0451)	(0.153)	(0.0464)
Bachelors degree or higher	0.0280	0.0293	0.0190	0.0870	0.0139
	(0.0259)	(0.0261)	(0.0290)	(0.0886)	(0.0348)
Associates degree	0.0184	0.0183	0.0220	-0.0116	0.0352
	(0.0305)	(0.0303)	(0.0317)	(0.0972)	(0.0375)
Child Development Associates	-0.0635**	-0.0623**	-0.0689**	-0.122	-0.0481
	(0.0233)	(0.0228)	(0.0261)	(0.0780)	(0.0283)
Hourly Wage	-0.0103*	-0.0106*	-0.00136	0.00527	-0.00354
	(0.00408)	(0.00406)	(0.00494)	(0.0137)	(0.00674)
Age	-0.00261**	-0.00258**	-0.00218*	-0.00290	-0.00191
	(0.000952)	(0.000950)	(0.000999)	(0.00288)	(0.00113)
Experience in ECE	-0.000976	-0.00104	-0.00234	0.00423	-0.00405*
	(0.00151)	(0.00152)	(0.00169)	(0.00520)	(0.00197)
Leaders experience in ECE	0.00181	0.00182			
	(0.00127)	(0.00127)			
Enrollment Children 0-5	0.000242	0.000271			
	(0.000648)	(0.000639)			
Number of Assistant Teachers	-0.00562	-0.00575			
	(0.00313)	(0.00314)			
Number of Lead Teachers	-0.000614	-0.000724			
	(0.00201)	(0.00201)			
Leader's highest degree = Associates	0.0270	0.0269			
	(0.0385)	(0.0389)			
Leader's highest degree = Bachelors	0.000813	0.00234			
	(0.0375)	(0.0376)			
Leader's highest degree = Masters or higher	-0.00711	-0.00654			
	(0.0345)	(0.0341)			
Constant	0.340**	0.292*	0.363**	0.527**	0.382**
	(0.0913)	(0.122)	(0.0730)	(0.196)	(0.0933)
Observations	967	967	967	195	680
R-squared	0.087	0.088	0.284	0.420	0.271
Center Fixed Effects?	No	No	Yes	Yes	Yes
Centers Sample	Full	Full	Full	Black Leader	White Leader

Robust standard errors in parentheses

Significance levels: ** p<0.01, * p<0.05

Note: The dependent variable is a binary indicator equal to one if the teacher reported in Fall 2020 that they expected to still work at their center in May 2021. The race/ethnicity variables (Black and Hispanic) refer to the teacher's race. The regression models include the following center-level covariates: leader's race/ethnicity, leader's years of experience, leader's degree attainment, number of children ages 0-5 enrolled at the center, total number of lead teachers employed by the center, total number of assistant teachers, total number of lead teachers, community in the Preschool Development Grant (PDG), and corresponding missing indicators for each variable. The regression models also include the following individual-level covariates: whether the

teacher is a lead teacher, whether the teacher teaches primarily infants and toddlers, degree attainment, whether a teacher has a Child Development Associates, their reported annual salary, their experience in early childhood education, their age, and corresponding missing indicators for each variable. The coefficients for community indicators and missing indicators are not shown above. Sample sizes vary due to missing outcome data for the 1011 teachers in the overall sample.

Table 9. Regressions of Turnover on Race/Ethnicity and Racial/Ethnic Match

VARIABLES	(1)	(2)	(3)	(4)	(5)
Match (any race)	0.0112	0.0480	0.00111	-0.0526	-0.000414
	(0.0268)	(0.0792)	(0.0314)	(0.0989)	(0.0380)
Black	0.0292	0.0751			
	(0.0345)	(0.0756)			
Hispanic	0.0537	0.0575			
	(0.0450)	(0.0868)			
Black Leader	-0.0277	0.0222			
	(0.0418)	(0.0703)			
Hispanic Leader	-0.0182	-0.0337			
	(0.0461)	(0.0729)			
Match x Black		-0.119			
		(0.142)			
Match x Hispanic		0.0399			
		(0.140)			
Male	-0.0361	-0.0369	-0.0649	0.283	-0.154
	(0.0794)	(0.0793)	(0.0948)	(0.361)	(0.122)
Lead Teacher	0.0189	0.0174	-0.0290	0.0368	-0.0775
	(0.0455)	(0.0455)	(0.0514)	(0.198)	(0.0569)
Bachelors degree or higher	0.0596	0.0611	0.0408	0.0999	0.0285
	(0.0319)	(0.0320)	(0.0325)	(0.111)	(0.0377)
Associates degree	-0.0537	-0.0546	-0.0344	0.0167	-0.0584
	(0.0293)	(0.0296)	(0.0326)	(0.0852)	(0.0384)
Child Development Associates	-0.0426	-0.0424	-0.0443	-0.0699	-0.0452
	(0.0283)	(0.0282)	(0.0316)	(0.0993)	(0.0368)
Hourly Wage	-0.00291	-0.00296	0.000485	-0.00649	0.000212
	(0.00488)	(0.00489)	(0.00633)	(0.0175)	(0.00880)
Age	-0.00467**	-0.00459**	-0.00409**	-0.00186	-0.00486**
	(0.00109)	(0.00108)	(0.00105)	(0.00325)	(0.00122)
Experience in ECE	0.000168	4.76e-05	-0.000484	-0.000422	-0.000657
	(0.00175)	(0.00176)	(0.00178)	(0.00545)	(0.00216)
Leaders experience in ECE	-0.000516	-0.000562			
	(0.00152)	(0.00151)			
Enrollment Children 0-5	-0.000306	-0.000212			
	(0.000671)	(0.000678)			
Number of Assistant Teachers	-0.000733	-0.00137			
	(0.00301)	(0.00309)			
Number of Lead Teachers	-3.73e-05	-0.000472			
	(0.00177)	(0.00182)			
Leader's highest degree = Associates	-0.0506	-0.0526			
	(0.0458)	(0.0458)			
Leader's highest degree = Bachelors	-0.119*	-0.115*			
	(0.0472)	(0.0469)			
Leader's highest degree = Masters or higher	-0.116**	-0.113**			
	(0.0432)	(0.0429)			
Constant	0.369**	0.335*	0.324**	0.272	0.413**
	(0.122)	(0.143)	(0.0883)	(0.267)	(0.115)
Observations	965	965	965	200	674
R-squared	0.117	0.118	0.321	0.428	0.308
Center Fixed Effects?	No	No	Yes	Yes	Yes
Centers Sample	Full	Full	Full	Black Leader	White Leader

Robust standard errors in parentheses

Significance levels: ** p<0.01, * p<0.05

Note: The dependent variable is a binary indicator equal to one if the teacher left their center over a six-month period. The race/ethnicity variables (Black and Hispanic) refer to the teacher's race. The regression models include the following center-level covariates: leader's race/ethnicity, leader's years of experience, leader's degree attainment, number of children ages 0-5 enrolled at the center, total number of lead teachers employed by the center, total number of assistant teachers, total number of lead teachers, community in the Preschool Development Grant (PDG), and corresponding missing indicators for each variable. The

regression models also include the following individual-level covariates: whether the teacher is a lead teacher, whether the teacher teaches primarily infants and toddlers, degree attainment, whether a teacher has a Child Development Associates, their reported annual salary, their experience in early childhood education, their age, and corresponding missing indicators for each variable. The coefficients for community indicators and missing indicators are not shown above. Sample sizes vary due to missing outcome data for the 1011 teachers in the overall sample.

Appendix A: Supplemental Tables

Table A1: Outcome Distribution for Views of Leaders, Job Satisfaction, and Turnover

Variable (Response scale if applicable)	Item if applicable	Mean	SD	% Missing	ICC
Views of Leadership					
Think about your experiences with this site leader since August. Please indicate how much you agree with the following statements. (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree)	I feel respected by the site leader.	4.09	0.98	4.35	0.08
	The site leader supports teachers' efforts to manage challenging behavior.	4.04	0.98	3.86	0.11
	I trust the site leader to do what they say they will do.	3.98	1.04	4.15	0.11
	The site leader communicates a clear vision for this site.	4.01	1	4.35	0.12
	I feel comfortable raising issues and concerns that are important to me with the site leader.	3.99	1.07	3.56	0.11
	The site leader is committed to creating an anti-racist learning environment for the children we serve.	4.34	0.85	4.65	0.08
	The site leader supports the professional development of staff.	4.14	0.91	3.86	0.07
Job Satisfaction					
Think about your position at your site. How satisfied are you with each of the following? (Not at all satisfied, A little bit satisfied, Somewhat satisfied, Very satisfied)	The job overall	3.43	0.74	4.35	0.06
	Interactions with the site leaders	3.36	0.85	4.65	0.10
Please indicate how much you agree with the following statements. (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree)	I really enjoy my present job.	4.23	0.86	3.86	0.00
Turnover					
Teacher turnover after six months		0.16	0.37	4.55	0.09
How likely is it that you will continue working in early childhood education? Please rate how likely you find each of the following statements. (Not likely, A little bit likely, Moderately likely, Very likely)	I will be working at this site in May 2021.	3.54	0.76	4.35	0.05

CHAPTER 2

No Hitting: Effects of a Ban on Corporal Punishment for Students with Disabilities

Todd M. Hall*

University of Virginia

Acknowledgements: This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA-NSF Grants Program" from the National Science Foundation under NSF award NSF-DRL #1749275. Opinions reflect those of the author and do not necessarily reflect those AERA or NSF. Todd Hall's time for analysis and writing was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200005 to the University of Virginia. I would also like to thank Daphna Bassok, Brendan Bartanen, Beth Schueler, Vivian Wong, participants in the EdPoliyWorks Lab Series and participants at the Association for Public Policy and Management 46th Annual Meeting for helpful comments. The contents of this article are those of the authors and do not represent the official views of these organizations or institutions.

*Corresponding author. Contact: PO Box 400879, Charlottesville, VA 22904. tmh4xa@virginia.edu

Abstract: Fifteen states allow corporal punishment—paddling, spanking, and other physical discipline—in public schools, which reported nearly 100,000 incidents in 2017-2018. Students with disabilities (SWDs) disproportionately experience school corporal punishment, a controversial practice associated with injuries and adverse outcomes. Despite renewed calls to eliminate the practice, several Southern legislatures have rejected bills to ban corporal punishment in public schools as recently as March 2024 in Florida. Yet five Southern states have recently restricted corporal punishment for SWDs only. There has been almost no research on these restrictions. We have little evidence on compliance or unintended consequences like replacing physical punishment with exclusionary discipline, another punitive practice sometimes framed as a substitute for corporal punishment. This study fills this gap. Applying a quadruple difference method to the Civil Rights Data Collection, I estimate the effects of Louisiana's 2017 ban on corporal punishment for SWDs on corporal punishment, out-of-school suspension, and in-school-suspension. Using multiple placebo comparisons, school-by-year fixed effects, disability-status-by-year fixed effects, and disability-status-by-school fixed effects, I rule out a host of competing explanations for changes in discipline outcomes. I find that Louisiana's 2017 ban significantly reduced corporal punishment for SWDs, though compliance was imperfect. I find no evidence of an increase in suspensions, suggesting that policymakers need not choose between exclusionary and physical discipline. I conclude with policy implications, calling for monitoring and accountability to support the implementation of discipline reforms.

Fifteen states expressly allow corporal punishment—paddling, spanking, and other physical discipline—in public schools (Cardona, 2023), which reported nearly 100,000 incidents of corporal punishment in 2017-2018 (Keierleber, 2021). Students with disabilities (SWDs) were disproportionately subjected to the practice (Dhaliwal et al., 2024; Losen et al., 2019; MacSuga-Gage et al., 2021). Corporal punishment is associated with injuries (E. Gershoff et al., 2019; Heekes et al., 2022), lower social-emotional skills (Cuartas, 2022), anti-social behavior (E. Gershoff, 2010; E. T. Gershoff, 2017; E. T. Gershoff & Font, 2016), and lower academic achievement (E. Gershoff et al., 2019; Organda Portela & Pells, 2015). U.S. Secretaries of Education have urged states and districts to ban the practice, citing disparities and adverse effects (Cardona, 2023; King, Jr., 2016). Yet several Southern state legislatures have rejected bills to ban corporal punishment. Louisiana's state legislature rejected five school corporal punishment bans since 2010, with a 2022 bill failing 51 to 42 (Noakes, 2022). Likewise, in March 2024, Florida's Education and Employment Committee rejected House Bill 439, which would have prohibited corporal punishment for SWDs and required parental consent to use corporal punishment for all other students (House Bill 439 (2024) - The Florida Senate, 2024). Some educators defend the practice as a more efficient alternative to suspension that maximizes instructional time and has approval from parents (Kennedy et al., 2017; Mweru, 2010).

Notably, five Southern states, starting with Louisiana in 2017, have enacted restrictions on corporal punishment for SWDs while condoning it for other students (Cardona, 2023). Research has lagged behind these reforms, leaving substantial gaps in our understanding. First, there is almost no evidence on the effects of these exemptions on corporal punishment for SWDs, including how closely schools complied and whether there were unintended consequences. Researchers have struggled to identify whether state bans on corporal punishment

reduce corporal punishment apart from simultaneous changes in district or school policies, such as local policy changes, changes in district or school leadership, or adoption of restorative practices (Curran & Kitchin, 2018; Heekes et al., 2022).

Second, scant research has examined the relationship between suspension and corporal punishment reforms. Only one study has examined how suspension rates changed as U.S. school districts stopped using corporal punishment, and it inferred "de facto" bans on corporal punishment based on districts reporting zero corporal punishment rather than identifying explicit policy changes (Curran & Kitchin, 2018). It is important to examine effects of banning corporal punishment on both physical and exclusionary discipline because school administrators and district handbooks have framed paddling as an alternative to suspension (Kennedy et al., 2017; Louisiana House of Representatives, 2021; Mathewson, 2022a). Banning corporal punishment for SWDs might cause administrators to replace corporal punishment with suspension. This unintended substitution effect would likely offset some benefits of a corporal punishment ban because suspensions are also associated with a host of negative academic and long-term outcomes (Bacher-Hicks et al., 2019; Davison et al., 2022; Lacoe & Steinberg, 2019). In contrast, a corporal punishment ban might encourage educators to explore less punitive alternatives for SWDs or even improve student behavior by making students less agitated. The net effect on suspensions is theoretically ambiguous. Studies of discipline reforms in general have typically focused on suspensions, and relatively few have examined unintended consequences (Anderson et al., 2019; Khafaji-King, 2024; Sorensen et al., 2022).

This study is the first to evaluate both the first-order effects (compliance) of a ban on corporal punishment for SWDs and potential unintended consequences for exclusionary discipline outcomes. I fill this gap by applying quasi-experimental methods to federal data on

student discipline outcomes to estimate the impact of Louisiana's 2017 ban on corporal punishment for SWDs. I will address two research questions:

(RQ1) To what extent did schools comply with Louisiana's ban on corporal punishment for students with disabilities?

(RQ2) To what extent did the ban impact exclusionary discipline for students with disabilities, namely in-school-suspensions (ISS) and out-of-school suspensions (OSS)?

This study advances the causal literature on school corporal punishment and school discipline reform. The quadruple difference method implements a series of nested placebo comparisons: comparing the change in discipline outcomes over time for SWDs to the change for their peers without disabilities within the same schools, comparing that difference-in-differences across schools previously using corporal punishment vs. not, and comparing that triple difference across Louisiana vs. comparison states. This approach rules out a host of competing explanations such as changes in district policies or school leadership for why corporal punishment declined sharply for SWDs in Louisiana. Prior correlational research has suggested that school- or district policies and cultural shifts are more worthwhile because top-down bans were inconsequential (Curran & Kitchin, 2018; Heekes et al., 2022). More rigorous evidence on compliance with a state corporal punishment ban can help reformers prioritize efforts between top-down and grass-roots policy changes.

By examining different discipline practices—suspensions and corporal punishment—I contribute to emerging evidence in the broader discipline reform literature that considers tradeoffs rather than a single practice in isolation (Sorensen et al., 2022). Knowing if schools tend to replace corporal punishment with suspensions after corporal punishment bans can help policymakers proactively target educational resources. For example, Massachusetts recently

began requiring schools to provide opportunities for students to learn while suspended (Cleveland, 2023). A policy like the one in Massachusetts or a program like the Saturday detention expansion in Iberia Parish (*Iberia Parish School Board Meeting on September 21, 2022*, 2022) might appeal to principals who said they preferred paddling over suspension so that students being disciplined who often struggled academically would not miss as much instructional time (Kennedy et al., 2017).

I find that Louisiana's ban on corporal punishment for SWDs significantly decreased corporal punishment for SWDs in schools that reported any corporal punishment since 2009–2010 by 1.18 percentage points. This effect represents a 41% decline relative to the prior year corporal punishment rate. The significant decline in corporal punishment for SWDs is robust to alternative specifications using a triple difference and using pseudo-Poisson maximum likelihood estimation instead of OLS. Sensitivity analyses that assume varying magnitudes of bias suggest that the decline remains significant even if the ban slightly reduced reporting of corporal punishment for SWDs in Louisiana. I find no evidence that the ban increased suspensions, suggesting that policymakers and principals need not choose the lesser of two evils between physical and exclusionary discipline.

Background

Disproportionate Discipline of Students with Disabilities

SWDs disproportionately experience physical and exclusionary discipline (Cardona, 2023; Losen et al., 2019; Welsh & Little, 2018). In 2011–2012, SWDs were at least 50% more likely to receive corporal punishment than students without disabilities in 35% of Louisiana public schools that reported at least 10 corporal punishment incidents (E. T. Gershoff & Font, 2016). Disparities of this magnitude were similarly prevalent in Tennessee and Georgia, and

more prevalent in Alabama, Arkansas, and Mississippi (E. T. Gershoff & Font, 2016). Students with learning disabilities were more likely to be suspended or expelled than their peers without disabilities (Brobbey, 2018; Welsh & Little, 2018). Students with emotional/behavioral disorders and attention-deficit/hyperactivity disorder were even more likely to be suspended or expelled than students with learning disabilities (LD) (Achilles et al., 2007).

There are also racial inequities in discipline among SWDs. Black SWDs were more likely to receive corporal punishment (E. T. Gershoff & Font, 2016) and be suspended (Brobbey, 2018; Welsh & Little, 2018) than White SWDs. These disproportionalities reflect interactions between racism and ableism (Cruz, Firestone, et al., 2021; Cruz, Kulkarni, et al., 2021; MacSuga-Gage et al., 2021). Researchers have argued that discipline interventions often reduce overall suspension rates without reducing racial disparities because they target student behavior rather than educator biases that predict disparities (Cruz, Firestone, et al., 2021; Welsh & Little, 2018).

Corporal Punishment in U.S. and Louisiana Public Schools

The regulation of corporal punishment in U.S. public schools has historically been decentralized and racialized (Dhaliwal et al., 2024; Ward et al., 2021). In the 18th and 19th centuries, schools used corporal punishment routinely based on the legal doctrine of *in loco parentis*. Communities endorsed teachers' use of corporal punishment as a means of building moral character and maintaining order in public education (Dhaliwal et al., 2024). During the Antebellum period, early challenges to corporal punishment selectively opposed the practice for White students; articles in education journals argued that corporal punishment would degrade White students to a slavish mentality and was more appropriate for "'the negro plantation' than the republican schoolhouse" (Glenn, 1984, p. 57).

In 1977, the Supreme Court's landmark ruling in *Ingram v. Wright* declared that the Eighth Amendment prohibition on cruel and unusual punishment did not apply to corporal punishment in public schools, even if the corporal punishment was excessive (E. T. Gershoff & Font, 2016; *Ingraham v. Wright, 430 U.S. 651 (1977)*, 1977; Olsen, 1977). The Court noted that administrators could face civil and criminal liability for excessive corporal punishment based on common law and cited how some states had common law tests of reasonableness. Markers of reasonableness included notifying parents, only allowing the principal to administer corporal punishment, and requiring an adult witness. By deferring to existing legal recourse, the Court left the responsibility to define what constituted reasonable corporal punishment to states and districts (*Ingraham v. Wright, 430 U.S. 651 (1977)*, 1977; Olsen, 1977).

Until recent years, the state of Louisiana defined corporal punishment but deferred the regulation to districts. Louisiana's definition of school corporal punishment mirrors the definitions in other states (E. T. Gershoff & Font, 2016): "using physical force to discipline a student, with or without an object. Corporal punishment includes hitting, paddling, striking, spanking, slapping, or any other physical force that causes pain or physical discomfort." Louisiana State Legislature (Act 266 - House Bill No. 79, 2017). Until Louisiana banned corporal punishment for SWDs in 2017, the state law simply stated that each school board that exercises its discretion to allow corporal punishment must also adopt rules and regulations that it deems necessary to implement corporal punishment (Act 266 - House Bill No. 79, 2017).

These regulations are promulgated in district policy manuals, and at least 19 of Louisiana's 69 school districts permitted corporal punishment as of 2022 (Sentell, 2022). The Livingston Parish manual stipulated, for example, that corporal punishment can only be administered by the school principal, assistant principal, or their designated representative in the

principal's office or a designated location with a second school employee present as a witness (Livingston Parish School Board, 2017). The district manual required schools to keep documentation of the punishment with the students' name, the time, and details of the behavior, signed by the witness and available to the child's parent upon request (Livingston Parish School Board, 2017). According to its 2013–2014 handbook, the Claiborne Parish School Board issued schools a "standard paddle" and set a limit of four strikes on the posterior (Claiborne Parish School Board, 2023).

Few studies have described the firsthand perspectives of students receiving and educators administering corporal punishment. In one study of student perspectives, young adults who reported receiving school corporal recalled that either a principal or assistant principal (41%) or a teacher (39%) administered the punishment, most often using wooden paddles (52%), hands (21%), or rulers (15%) (E. Gershoff et al., 2019). Administrators in Florida schools that allowed corporal punishment believed it deterred disruptive behavior and developed character while maximizing class time. They frequently described OSS as "a vacation" that deprived students who often struggle academically of instruction (Kennedy et al., 2017). The research on administrators' responses to corporal punishment bans is limited to international studies (Lwo & Yuan, 2011; Mweru, 2010). Educators in Kenya reported using corporal punishment after a ban because they considered it effective and approved by parents (Mweru, 2010).

Compliance with Corporal Punishment Bans

The few studies on compliance with U.S. state and international corporal punishment bans suggest that they were mostly unrelated to rates of corporal punishment (Curran & Kitchin, 2018; Heekes et al., 2022). However, U.S. state bans were historically moot. The districts within almost every state that implemented a ban from 1980 to 2006 had phased out corporal

punishment almost entirely before the state ban was enacted (Curran & Kitchin, 2018). Curran and Kicthin (2018) acknowledge that their analysis was not causal but suggest that state bans were often enacted too late to make a difference. Concluding that state and national bans were insufficient, researchers have recommended that reformers instead focus on district-level policies and local cultural change (Curran & Kitchin, 2018; Heekes et al., 2022). These conclusions may not apply to more recent bans in U.S. states whose districts were still using corporal punishment and for current policy debates in Southern states still condoning the practice (Keierleber, 2024).

Still, corporal punishment bans might experience implementation failures like those that researchers have identified with suspension bans. Anderson (2018) found that three years after Arkansas banned OSS for truancy, 47% of the schools in the state that reported any truancy still reported issuing at least one OSS for truancy, and 9% of the schools used OSS for *all* of their truancy incidents (Anderson, 2018). Anderson (2018) hypothesizes that a lack of communication and accountability played a role. Despite non-compliance, the Arkansas Department of Education sent its first reminder to districts about the ban on OSS for truancy more than three years after the ban (Anderson, 2018). In Louisiana, the 2017 ban on corporal punishment for SWDs did not include any provisions for monitoring, accountability, or enforcement.

Corporal Punishment and Exclusionary Discipline

The risk that schools replace corporal punishment with exclusionary discipline reflects a broader concern about unintended consequences after discipline reform. Researchers studying restrictions on OSS for specific infractions such as defiance and disruption have found that reductions in the prohibited category of OSS were offset by increases in other categories of OSS, resulting in no significant decrease in total OSS (Craigie, 2022; Lacoe & Steinberg, 2018; Wang, 2022). Unintended consequences can extend beyond changes across different categories of

suspensions. Khafaji-King (2024) found that New York City's ban on suspensions for disorderly behavior increased disability classifications, primarily for specific learning disability, emotional disturbance, and intellectual disability. Khafaji-King (2024) hypothesizes that some of the alternatives recommended in the reform such as referrals and behavioral assessments may have increased students' contact with teams that classify disabilities.

Substitution between corporal punishment and suspensions may occur when principals respond to office discipline referrals. In in-depth interviews, Florida principals and assistant principals described various exchange rates between corporal punishment and suspension: one believed that one "lick" with the paddle could replace three days of ISS while another believed that one lick could replace one day of ISS or OSS. Many described punishment as a necessary deterrent, and nearly all 27 administrators interviewed saw effective consequences as those that students feared (Kennedy et al., 2017). Media coverage of school corporal punishment echoes this punitive framing of alternatives: students and parents report administrators giving them a choice between paddling or suspension (Keierleber, 2024; Mathewson, 2022a, 2022b). School administrators with a punitive approach to social control (Black, 1993) may see suspension as the logical alternative when corporal punishment is banned.

It is plausible, however, that banning corporal punishment could reduce exclusionary discipline by changing administrator or student behavior. For example, after Arkansas banned on OSS for truancy, some school administrators began taking away course credit; adding extra school days; sending students home for part of the day; and giving morning, after-school, and Saturday detentions (Anderson, 2018). The extra days and Saturday detentions were not exclusionary and perhaps increased instructional time. Banning corporal punishment might also avoid aggravating student behavior, and thereby alleviate other discipline issues. Adults ages 18-

23 who received corporal punishment in school reported feeling embarrassed, angry and, to a lesser extent, vengeful (E. Gershoff et al., 2019). Corporal punishment at home is associated with worse long-term behavior (E. Gershoff, 2010; E. T. Gershoff, 2002). To the extent that school corporal punishment erodes trust between students and adults and exacerbates student behavior, eliminating the practice may reduce suspensions.

Data and Methods

Data

I use data from the federal Civil Rights Data Collection (CRDC), a representative survey of schools covering 85% of the nation's students in 2009–2010 and a biannual census of U.S. public schools for school years 2011–2012, 2013–2014, 2015–2016, and 2017–2018. The CRDC required schools to report both enrollment and the number of students who received corporal punishment, out-of-school suspension (OSS), and in-school-suspension (ISS) in a given year for each combination of race/ethnicity, gender, and disability status (e.g., Black male students with a disability served under IDEA).[1]

I restrict the sample to Louisiana and the four other states that imposed similar restrictions on corporal punishment for students with disabilities one to five years after Louisiana: Kentucky, Mississippi, Oklahoma, and Tennessee (Cardona, 2023). Mississippi, Oklahoma, and Kentucky banned corporal for SWDs with no exceptions in 2019, 2020, and 2022 respectively (Keierleber, 2021; Kentucky Department of Education, 2022; National Center on Safe Supportive Learning Environments, 2024; Oklahoma State Department of Education, 2022; Title 704 Chapter 7 Regulation 170, 2022, p. 704). Tennessee banned corporal punishment for SWDs except with parents' affirmative written consent in May 2018 (Tennessee HB2330 | 2017–2018 | 110th General Assembly, 2018). I chose these four comparison states for two

reasons. First, states that restricted corporal punishment for SWDs selected into a similar treatment as Louisiana and might be more similar on unobserved factors such as leanings of the state legislature and Board of Education than states that did not implement comparable policies. Second, as I will show, Kentucky, Mississippi, Oklahoma, and Tennessee pooled together followed a similar trend in corporal punishment rates for SWDs as Louisiana prior to 2017. This helps establish a counterfactual for what the discipline may have looked like for SWDs in Louisiana in the absence of Louisiana's ban. The restrictions on corporal punishment in these comparison states could not be studied due to a pause in data collection for the CRDC during the COVID-19 pandemic, but they provide a natural comparison for Louisiana.

I further restrict the sample to schools that appeared in the CRDC for all five years. This restriction avoids bias from changes in the composition of schools across the panel. The resulting analytic sample consists of 4,795 schools and 47,950 school-by-year-by-disability status observations. This fixed sample includes about 80% of Louisiana schools and about 65% of the schools in the comparison states that ever appeared in the CRDC from 2009-10 to 2017-18.

Table 1 below describes the schools in my sample by state and year. Since my analysis will make comparisons between students with and without disabilities, Table 1 separately presents demographic characteristics for students with and without disabilities within states. With a few exceptions, the within-state race/ethnicity breakdowns of students with and without disabilities were similar and mostly stable across years. In all states, the percentage of SWDs who were White decreased and the percentage of SWDs who were Hispanic increased, with a large percent change given the low baseline percentages. Across states and in both years, male students accounted for about two-thirds of the students with disabilities.

[TABLE 1 HERE]

The percentage of students classified as having a disability increased in both Louisiana and comparison states, with a slightly larger increase in Louisiana. This is important to note because a decline in the share of SWDs in Louisiana might raise concerns that schools were identifying fewer SWDs so that they could legally subject more students to corporal punishment. For this type of unintended consequence to have occurred, schools would have had to simultaneously classify fewer students as having disabilities and replace them with newly classified students from a similar background since both the share of SWDs increased and their demographics remained stable. Thus, changes in identification in response to the ban seem unlikely.

Method

I apply a quadruple difference method similar to Craigie (2022) to attribute changes in corporal punishment and suspensions for SWDs in Louisiana to the state's ban on corporal punishment for this subgroup. The quadruple difference extends the basic difference-in-differences approach by introducing a third and fourth difference. Each difference introduces a placebo comparison group that can absorb sources of bias but should not be affected by Louisiana's 2017 ban (Olden & Møen, 2022). The following paragraphs unpack the quadruple difference by explaining each placebo comparison and the sources of bias mitigated by each of the four differences. I then articulate the assumptions under which the quadruple difference identifies the average treatment-on-treated (ATT) effect of Louisiana's ban on corporal punishment for students with disabilities. Finally, I explain how I estimate the quadruple difference.

First and Second Differences. The first difference is the change in discipline rates over time. This corresponds to a simple pre-post comparison of how discipline rates change for SWDs

in Louisiana pre- and post-2017. This pre-post comparison provides a naïve estimate of the effect of Louisiana's 2017 ban, but introducing additional differences allows me to rule out confounding factors to attain more robust estimate. For example, we know that school corporal punishment rates declined over time nationwide between 2013–2014 and 2017–2018 (U.S. Department of Education, Office for Civil Rights, 2023). Likewise, nationally representative surveys indicate that parents' use of spanking at home and their agreement that children sometimes need spanking have declined between the late 1980s and the late 2010s, albeit more slowly in the South (Finkelhor et al., 2019; Hines et al., 2022; Ryan et al., 2016). The naïve pre-post difference could overstate the effectiveness of Louisiana's ban by mistakenly attributing a secular downward trend in corporal punishment to the ban.

To avoid this confounding, the second difference compares the gap in the discipline rates between SWDs and students without disabilities within the same school. Taking the change in discipline rates for SWDs over time (the first difference) and subtracting the change in the discipline rate for students without disabilities attending the same school during the same year (the second difference) yields a standard difference-in-differences. The intuition behind this difference-in-differences at the core of the quadruple difference is to compare changes over time in discipline outcomes for SWDs, whom the ban targeted, to a placebo trend for students without disabilities within the same schools. Students without disabilities attending the same school during the same year should be subject to many of the same local conditions (e.g., principals, teachers, school climate, and district and school-level discipline policies), but they should not be affected by the state's ban on corporal punishment for SWDs. Using students without disabilities as a placebo comparison group allows me to include school-by-year fixed effects that control for

school-by-year shocks such as a new principal or assistant principal who does not endorse corporal punishment or a decline in parents' endorsement of corporal punishment.

Triple Difference. The third difference makes an additional comparison across states. The triple difference computes the difference-in-differences described above for Louisiana only and subtracts from it the analogous difference-in-differences for the pooled comparison states: Kentucky, Oklahoma, Mississippi, and Tennessee. This triple difference provides a placebo test because the difference in outcomes over time between SWDs and students without disabilities in other states should not be affected by Louisiana's ban (Cunningham, 2021). This third difference enhances the robustness of the difference-in-difference accounts for disability-status-by-year shocks that occur across states. For example, the triple difference absorbs any common effect of 2014 federal guidance cracking down on discipline disparities under the Obama administration and its reversal under the Trump administration (Kamenetz, 2018). The placebo group of SWDs also removes potential bias from disability-status-specific reporting anomalies or confusions about the CRDC.

Quadruple Difference. The fourth difference makes an additional comparison across schools likely to be *more affected*, defined as those that reported corporal punishment at any point in the pre-ban period, and schools likely to be *less affected*, which I define as those that had not previously reported any corporal punishment. The logic is that students with disabilities attending schools that had ever reported corporal punishment prior to 2017 would be *more affected* by a ban on corporal punishment for students with disabilities than students with disabilities in schools that had not previously reported corporal punishment. Schools that did not report any corporal punishment in the pre-period should not have had to change much or at all to comply with the ban. Students in these *less affected* schools are marked as untreated regardless

of disability status, alongside students without disabilities (the comparison group in the second difference) and students outside Louisiana (the comparison group in the third difference). This approach to defining treatment status is common in studying suspension reforms, where researchers have defined the treatment group as students who were *more affected* (*vs. less affected*) because they attended *more affected* schools—those with higher (vs. lower) pre-reform suspension rates (Cleveland, 2023; Craig & Martin, 2023; Craigie, 2022). The quadruple difference adds a placebo group of SWDs in *less affected* schools and accounts for potential bias that comes from state-and-yappear-specific shocks affecting all SWDs. For example, the quadruple difference would "difference out" common effects of statewide funding for special education or technical assistance on positive behavior interventions.

Assumption for Causal Inference: Parallel Trends. The parallel trends assumption for the quadruple difference is as follows. In the absence of Louisiana's 2017 ban, the change over time in the gap in discipline rates between students with and without disabilities within *more affected* Louisiana schools relative to *less affected* Louisiana schools would be the same as the change over time in the gap in discipline rates between students with and without disabilities within *more affected* non-Louisiana schools relative to *less affected* non-Louisiana schools. In other words, Louisiana's 2017 ban is the factor that uniquely differentiates the within-school change in discipline rates over time for students covered by the three-way interaction of (1) SWDs (2) in *more affected* schools (3) in Louisiana from the within-school change associated with any of these characteristics separately or any two-way interaction of these characteristics. The quadruple difference does not require a separate parallel trends assumption to hold for the difference-in-differences or the triple difference that are nested within it (Cunningham, 2021; Olden & Møen, 2022).

Assumption for Causal Inference: No Anticipation Effects. Since the data are biannual, anticipation effects can only bias the results if they appeared not one but two school years prior to Louisiana's 2017 ban. If anticipation effects were possible two years out, it is plausible that anticipation effects would have also been present for Mississippi and Tennessee, which restricted corporal punishment for SWDs in 2018 and 2019 respectively. To the extent that the anticipation effects were similar across states, differencing across states helps mitigate this potential bias.

Estimation

The quadruple difference estimates the gap in discipline rates for students with and without disabilities within schools over time in Louisiana vs. comparison states in *more affected* vs. *less affected* schools. I use a two-way fixed effects (TWFE) estimator with school-by-year fixed effects, school-by-disability status, and disability status-by-year fixed effects in an event study specification. Recent literature highlights bias from inappropriate comparisons involving previously treated units in TWFE with staggered treatment timing and dynamic treatment effects, but this analysis involves neither of those conditions. The TWFE estimator is interpretable when all treated units get treated at the same time (Roth et al., 2023), including this analysis with only one treated cohort and one post-treatment period: SWDs in Louisiana in 2017–2018.

For both corporal punishment and exclusionary discipline, I estimate the quadruple difference using Equation 1 below:

$$(1)\ Y_{ijst} = \beta_0 + \sum_{k=-a}^{b} \delta_k D.L.More_{ijst} + \sum_{k=-a}^{b} \gamma_k D.More_{ijt} + \sum_{k=-a}^{b} \theta_k More.L_{jst} + \sum_{k=-a}^{b} \psi_k D.L_{ist} + \sum_{k=-a}^{b} \rho_k More_{jt} + \sum_{k=-a}^{b} \omega_k D_{it} + \sum_{k=-a}^{b} \tau_k L_{st} + X_{ijt} + \lambda_{jt} + \eta_{ij} + \alpha_{it} + \varepsilon_{ijst}$$

$$(2) \qquad Y_{ijt} = \beta_0 + \sum_{\substack{k=-6 \\ k\ !=\ -2}}^{k=2} LA_{ijt}\tau_k + X_{ijt} + \alpha_j + \eta_t + \varepsilon_{ijt}$$

Y is the outcome for disability group *i* (with or without disabilities) in school *j* at time *t*. The outcome variables are the percentage of students receiving corporal punishment, the percentage of students receiving ISS, and the percentage receiving OSS. K represents two-year intervals relative to the ban in 2017–2018, and a to b indicate years in relative time before and after the ban. By convention, I use 2015–2016 as the omitted year since it is the period prior to treatment. *D* is an indicator representing the students with disabilities group, *L* is an indicator for Louisiana, and *More* is an indicator for *more affected*, i.e., an indicator equal to 1 if the school reported any corporal punishment between 2009–2010 and 2015–2016 and zero otherwise. I implement an event study specification where terms that include any indicator *D*, *L*, or *More* are interacted with event time for each year. The interaction between *D*, *L*, and *More* yields four coefficients for the treatment effect: three placebos from the pre-period and one post-treatment coefficient. The coefficient for 2017–2018, δ_{2018}, represents how much larger or smaller, relative to 2015–2016, the gap in discipline rates was between students with and without disabilities in Louisiana vs. comparison states in *more affected* vs. *less affected* schools. I interpret δ_{2018} causally as the ATT effect of Louisiana's 2017 ban on corporal punishment for students with disabilities. X is a vector of covariates for the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each disability status-by-school-by-year observation. I include these demographic covariates because these groups may be more likely to receive corporal punishment. λ is a school-by-year fixed effect, η is a disability-status-by-school fixed effect, and α is a disability-status-by-year fixed effect. Standard errors were clustered at the district-by-disability status level.

I tested for sensitivity to several alternative specifications. This included clustering standard errors at the level of treatment using the residual bootstrap with heteroskedasticity correction (Ferman & Pinto, 2019) for inference in difference-in-differences with few clusters. The residual bootstrap method accounts for having a single treated cluster: Louisiana SWDs in *more affected* schools. I also estimated robust confidence sets (Rambachan & Roth, 2023) to account for selective underreporting of corporal punishment, used the Poisson Pseudo-Maximum Likelihood (PPML) estimator (Correia et al., 2020) to account for having count data, and estimated a triple difference to account for spillovers to SWDs in *less affected* Louisiana schools. These methods and results are detailed in Section IX, "Robustness Checks." As a preview, the results from these sensitivity analyses remain similar to the results discussed below for the main specification.

Results

Results Overview

First, I show how corporal punishment rates changed during the panel for the treated group—students with disabilities in *more affected* Louisiana schools—and for each placebo comparison group. I use the trends in corporal punishment to illustrate the intuition behind the quadruple difference. Then I present the main regression results and event study figures.

Figure 1 below shows trends in rates of corporal punishment by disability status and treatment status (more vs. less affected), and state (Louisiana vs. pooled comparison states. Each marker represents the average year-by-disability status corporal punishment rate for schools in Louisiana, Kentucky, Mississippi, Oklahoma, and Tennessee, weighted by the enrollment in the school-by-disability-status cell. Recall that *more affected* schools are those that ever reported corporal punishment from 2010 to 2016. There were 379 *more affected* schools in Louisiana.

More affected schools are represented by solid squares (Louisiana) or triangles (comparison states). *Less affected* schools are represented by hollow squares or triangles; their corporal punishment rates remain exactly zero until 2016 by construction.

[FIGURE 1 HERE]

In 2009–2010, SWDs in *more affected* schools were slightly less likely to receive corporal punishment than their peers. This is unexpected since SWDs typically receive corporal punishment at disparate rates (Dhaliwal et al., 2024; Dillon, 2009; E. T. Gershoff & Font, 2016). This anomaly might reflect reporting confusions. The U.S. Department of Education data documentation for the 2009–2010 CRDC cautions that some districts may have mistakenly reported the number of incidents rather than the number of students disciplined (Office for Civil Rights, n.d.). If few students without disabilities received corporal punishment but had many incidents each while many SWDs had few incidents each, then counting incidents instead of students could falsely suggest similar corporal punishment rates. It could also falsely imply a sharp decline in the corporate punishment rate for students without disabilities in 2011–2012. Child abuse also increased during the Great Recession (Santaularia et al., 2022), and schools may have used corporal punishment more indiscriminately during this period. Whatever the explanation, this anomaly was not limited to Louisiana and occurred in the other states that later restricted corporal punishment for SWDs—Kentucky, Oklahoma, Mississippi, and Tennessee.

This helps illustrate the value of introducing the quadruple difference to "difference out" a pattern that was common across states. Corporal punishment rates by disability status within *more affected* schools across states follow a similar trend until a notable drop in 2017–2018 in corporal punishment for students with disabilities in *more affected* schools in Louisiana. A similar drop did not occur for students with disabilities in *more affected* schools in the

comparison states that did not pass a ban within the window of time under study here. In both Louisiana and the other states, more than 98% of the *less affected* schools continued to report zero incidents of corporal punishment for any students in 2017–2018. Therefore, when corporal punishment is the outcome, using the quadruple difference instead of the triple difference among *more affected* schools is akin to subtracting zero from both sides of the equation. Indeed, Appendix E Table E.1 Column 2 shows that the estimates for the effect on corporal punishment from the quadruple difference and from the triple difference among *more affected* schools are almost identical. Although the quadruple difference may not help satisfy the parallel trends assumption for corporal punishment, adding the fourth difference of *more affected* vs. *less affected* schools does help establish parallel pretends for the suspension outcomes.

Table 2 below summarizes the key findings, which I discuss by outcome in the following paragraphs. The coefficient on the interaction for 2017–2018 in Column 1 indicates that Louisiana's 2017 ban significantly reduced corporal punishment. The coefficients in the same row for columns 2 and 3 suggest that the ban did not significantly affect suspensions.

[TABLE 2 HERE]

Corporal Punishment Results

Figure 4 below presents event study coefficients for corporal punishment using OLS regressions. The omitted year 2015–2016 represents the ratio of Louisiana's corporal punishment rate for students with and without disabilities in *more affected* vs. *less affected* schools centered on the analogous ratio in other states. For OLS, this means that the ratio for 2015–2016 is centered at zero. The event study coefficients represent the differential change in corporal punishment rates associated with the interaction term representing students with disabilities in *more affected* schools in Louisiana in each year relative to the same interaction in the omitted

year 2015–2016. Dots represent the point estimates, and the bars show 95% confidence intervals. The estimates from before 2015–2016 provide evidence of parallel trends in the pre-period because the coefficients are not significantly different from zero.

[FIGURE 2 HERE]

The coefficient for 2017–2018 estimates the ATT effect of Louisiana's 2017 ban on corporal punishment for SWDs. Figure 4 suggests that Louisiana's ban on corporal punishment for SWDs significantly reduced corporal punishment for this subgroup. In 2017–2018, there is a statistically significant decrease of 1.19 percentage points in the share of SWDs in *more affected* Louisiana schools who were subjected to corporal punishment. This effect is small in absolute terms but large in relative terms. Since the average 2015–2016 corporal punishment rate for SWDs in *more affected* Louisiana schools was 2.86, the effect of 1.18 percentage points represents a decline of about 41%.

Suspension Results

Figure 5 below shows the OLS results for the percentage of students who received at least one OSS. Figure 5 provides evidence of parallel pre-trends, as coefficients in the pre-period are not significant at the 5% level. Louisiana's corporal punishment ban did not seem to increase OSS for SWDs.

[FIGURE 3 HERE]

Figure 6 below presents the OLS results for the percentage of students who received ISS. Figure 6 provides evidence of parallel pre-trends given that the event study coefficients from 2009–2010 to 2015-16 are not statistically different from zero. The coefficient for 2017–2018 is not significant either, suggesting that students with disabilities in *more affected* Louisiana schools were not more likely to receive ISS than their peers in 2017–2018 relative to the

disproportionality they experienced in the previous year. Appendix A replicates this null finding using the triple difference specification.

[FIGURE 4 HERE]

Robustness Checks

Accounting for a Small Number of Clusters

I reproduce my main results while clustering at the level of treatment and adjusting the p-values to account for having a small number of clusters: one treated cluster of Louisiana SWDs in *more affected* schools and 19 comparison clusters. With too few clusters and only one treated cluster, the commonly used cluster robust variance estimator (CRVE) can underestimate the standard errors (MacKinnon et al., 2023). Having one treated cluster made the design-based recommendation to cluster standard errors at the level of treatment inappropriate (Abadie et al., 2023). Instead, Roth et al. (2023) recommend alternatives that learn the distribution of the errors for few treated clusters from a larger number of control clusters. Among these, I use Ferman and Pinto's (2019) residual bootstrap, which aggregates the residuals from my main specification to one residual per cluster and then estimates the conditional variance of that residual using bootstrapped regressions. Their approach is robust to heteroskedasticity arising from differences in cluster sizes even with only one treated cluster (Ferman & Pinto, 2019; MacKinnon et al., 2023).

One wrinkle for implementing Ferman and Pinto's (2019) method while also using high-dimensional fixed effects is that the residual bootstrap can underestimate the variance of the errors as the ratio of predictors to observations moves away from zero and grows toward one (Karoui & Purdom, 2018). My main specification has 9,490 disability status-by-year fixed effects, 23,231 school-by-year fixed effects, 10 disability status-by-year fixed effects, and fewer

than 47,000 observations. This yields a of high ratio of predictors to observations, approximately 0.8, which can make residual bootstrap p-values prone to over-rejecting the null hypothesis (Karoui & Purdom, 2018). To avoid this issue, I apply the residual bootstrap with heteroskedasticity correction using both the main specification and specifications that introduce each set of fixed effects separately, which reduces the ratio of predictors to observations. Those ratios are approximately 0.5 using only school-by-year fixed effects, 0.4 using only school-by-disability status fixed effects, and 0 using only disability status-by-year fixed effects. The latter specification eliminates any risk of over-rejection from having many predictors relative to observations.

Appendix A Table A.1 and A.2 show that the residual bootstrap with heteroskedasticity correction does not substantively change my main findings. Whether I include all three sets of fixed effects or enter them in separate regressions, the results indicate that the ban significantly decreased the percentage of SWDs subjected to corporal punishment without significantly affecting the percentage subjected to suspensions. The main findings are robust to inference methods that account for having only one treated cluster.

Accounting for Selective Underreporting

To the extent that schools underreport corporal punishment similarly across states and over time, the quadruple difference accounts for underreporting. However, if Louisiana's ban itself dissuaded schools within the state from reporting corporal punishment for SWDs in 2017–2018, the decline in actual corporal punishment may be confounded with a negative effect on reporting. The only large-scale data on corporal punishment were reported by schools, so there is no source of statewide data that allows for a direct test of underreporting.

To address this data limitation, I conduct three different sensitivity analyses, summarized in the following subsections. Two sensitivity analyses each estimate a "switch point" for how large the bias from selective underreporting would have to be to nullify the significant decline in corporal punishment for SWDs. The first, Conditional Robustness of Inference to Replacement (Frank et al., 2013, 2023), quantifies a "switch point" in terms of sample replacement—intuitively, the proportion of schools in the data that would have to be replaced with schools that experience no treatment effect. The second, robust confidence sets for difference-in-differences (Rambachan & Roth, 2023), quantifies a switch point in magnitude relative to pretrend coefficients from the quadruple difference. My third sensitivity analysis uses process tracing (Collier, 2011) to examine whether Louisiana districts that took longer to align their policy manuals with the state's 2017 prohibition on corporal punishment for SWDs were the ones that continued to report corporal punishment in 2017–2018. This pattern would be consistent with schools reporting the corporal punishment for SWDs that occurred in 2017–2018 before administrators knew about the ban, as opposed to simply reporting zero because of the ban.

As a preview, the three sensitivity analyses lend some support to the inference that Louisiana's 2017 ban caused a real decline in corporal punishment for SWDs. Using different approaches, the first two sensitivity analyses identify a qualitatively similar switch point. The decline in corporal punishment for SWDs appears robust to replacing roughly a quarter of the schools in my data with schools for which the ban would have zero effect and to a reporting bias of modest relative magnitude. However, the effect could be nullified if selective underreporting was widespread. From the process tracing, I find that multiple districts that aligned their policy manuals in the middle of the school year saw large declines in the corporal punishment rate for SWDs from 2015–2016 to 2017–2018 without dropping to zero. Though not definitive, this

pattern may lend some credibility to the reporting process, as if schools reported the corporal punishment they used on SWDs prior to learning about the ban—rather than reporting zero.

Robustness to Sample Replacement. The first sensitivity analysis uses Conditional Robustness of Inference to Replacement (CRIR) to estimate how many schools would have to be replaced with schools that have a treatment effect of zero before the ATT effect on corporal punishment becomes insignificant. CRIR extends conventional Robustness of Inference to Replacement (RIR) (Frank et al., 2013, 2023) to interaction effects (Xu et al., 2024), such as the interaction of Louisiana x *more affected* x SWD x 2018 that yields ATT estimate. CRIR assumes a true interaction effect of zero conditional on the main effect, so that the estimated ATT effect represents bias due to differences in untreated potential outcomes when the interaction term equals 0 and when it equals 1. CRIR then supposes that we can replace some proportion of observed units with counterfactual units that do not introduce any bias because the expected value of their untreated potential outcomes is the same regardless of whether their four-way interaction term equals 0 or 1. Conventional RIR identifies the "switch point" to nullify a treatment effect when the proportion of replaced units exceeds one minus the estimated effect as a t-statistic divided by the threshold for statistical significance, often the standard error x the t-critical value (Frank et al., 2013, 2023). Because interaction effects represent a change in slope, CRIR involves an intermediate step of converting the t-statistics into correlation coefficients. Xu et al. (2024, work in progress) show that interaction effects cease to be statistically significant when the proportion of replaced units exceeds one minus the estimated interaction effect as a correlation coefficient divided by the threshold for statistical significance (e.g., the effect associated with $p = 0.05$) as a correlation coefficient (Xu et al., 2024).

Using CRIR and the konfound package in Stata (Xu et al., 2019), I estimate that about 23.5 percent of the schools in my sample would need to be replaced with schools for which there is no interaction effect to nullify the significant effect of Louisiana's ban on the percentage of SWDs subjected to corporal punishment. If selective underreporting were more widespread, then that could nullify the significant reduction in corporal punishment.

Robustness to Expanded Confidence Intervals. I locate a different "switch point" using Rambachan and Roth's (2023) method for constructing robust confidence intervals under specified violations of parallel trends in difference-in-differences. The estimator implemented by their HonestDiD package essentially widens the confidence interval around the ATT estimate to account for the upper bound on the largest pre-trend coefficient and repeats this procedure with biases of different relative magnitudes **M** times the largest event study coefficient. The intuition for using relative magnitudes is that the worst pre-trend coefficient offers a benchmark for describing the size of possible biases. The largest pre-trend coefficient in my main specification for corporal punishment is an insignificant 0.64 ($p = 0.28$, upper bound = 1.82), so I test robustness of the ATT estimate to the upper bound of possible bias as varying proportions of that pre-trend coefficient. The resulting set of confidence intervals describe how large a negative bias from selective underreporting would have to be to overturn the significant reduction in corporal punishment for SWDs in Louisiana.

The results of the sensitivity analysis indicate that the ATT is robust to the upper bound on a bias up to about half the size of the worst pre-trend coefficient. This is shown in Appendix B Figure B1. The upper bound of the robust confidence interval crosses zero when the multiplier **M** = 0.5. Moreover, since the pre-trend coefficients were insignificantly positive, it is likely that the post-period difference in trends would have also been insignificantly positive in the absence

of Louisiana's ban (in the absence of treatment). Corporal punishment rates typically declined more slowly for SWDs in Louisiana relative to their peers within schools than corporal punishment rates for SWDs in comparison states relative to their peers within schools. A negative bias from selective reporting would likely need to counteract a weakly positive bias before overturning the significant negative effect of the ban. Together, my sensitivity analysis and the direction of the pre-trend coefficients suggest that the -1.18-percentage point effect of Louisiana's ban on corporal punishment is robust to a modest amount of selective underreporting. However, more severe or widespread underreporting could render the effect on corporal punishment indistinguishable from zero.

Patterns in Corporal Punishment for SWDs by Louisiana District Policy Implementation. If Louisiana's 2017 ban reduced corporal punishment as opposed to only changing reporting, then slow diffusion of the policy should be reflected in gradual rather than abrupt declines in corporal punishment. We know that ban did entirely stifle reporting because Louisiana schools across 18 districts reported a total of 77 incidents of corporal punishment of SWDs in 2017–2018. One would hypothesize that the schools still reporting corporal punishment might be in districts that went several months into 2017–2018 before including the ban in their policy manuals; if the reporting is credible, such schools should report any corporal punishment they used on SWDs before they became aware of the law. I assess whether this pattern occurs in two steps described in Appendix C. First, I infer when 53 out of 63 districts in my sample updated their corporal punishment policy by documenting the inclusion of verbatim language from the state ban and the corresponding timestamps (Appendix C Figure C2). This does not perfectly measure when school administrators became aware of the law, but these changes reflect school board motions and may be the best proxy for awareness of the ban available statewide.

Second, I plot the corporal punishment rates across Louisiana at the district level for 2015–2016 and 2017–2018 ordered by when the district updated their policy (Appendix C Figure C3).

The descriptive patterns are mostly consistent with a gradual reduction in corporal punishment for SWDs as districts incorporated the state policy throughout 2017-2018. As shown in Appendix C, Figure C3, many districts updated their policy manuals during the 2017–2018 school year as opposed to immediately in August 2017. More than half (14 out of 27) of the districts that updated their policy manuals between September 2017 and July 2018 had schools report at least once instance of CP of SWDs in 2017–2018. These 14 districts saw declines in corporal punishment rates for SWDs from 2015–2016 to 2017–2018, but they did not drop to zero. Schools in these districts may have reported corporal punishment that occurred prior to their awareness of the state ban, lending some credibility to their reporting. Meanwhile, three districts with corporal punishment rates above 1 percent (roughly the state average) in 2015–2016 reported zero CP of SWDs in 2017–2018 despite taking a year or longer to update their manual. The drops to zero for these few districts may suggest some selective underreporting, although school administrators may have learned about the ban prior to their district updating their policy manual.

Accounting for Count Data

I use Poisson pseudo-maximum likelihood (PPML) fixed effect regression (Correia et al., 2020) to estimate the quadruple difference as an alternative to OLS because my outcomes take only non-negative values and have many zeros. All my outcomes are counts of students disciplined per 100 students with many zeros, and OLS regression may yield the wrong standard errors because the errors are not normally distributed. Wooldridge (2023) and Correia (2019)

respectively recommend PPML for nonnegative outcomes with many zeros as an alternative to linear models, which ignore the fact that the conditional mean cannot be negative, and log-linear models, which cannot handle zeros and may be inconsistent when there is heteroscedasticity (Correia et al., 2020). This approach follows Wang's (2022) difference-in-difference study of a suspension ban. PPML is robust in that it only requires correct specification of the outcome as an exponential conditional mean for consistent estimation, not a Poisson distribution, and it can handle over- or underdispersion (Cameron & Trivedi, 2005; Wooldridge, 2010). I use PPML instead of the negative binomial because the latter will be inconsistent if there is underdispersion (Wooldridge, 2010).

Appendix D Table D1 shows the PPML regression results. For corporal punishment (Column 1), the PPML regression yields results similar to the OLS results. Here, a coefficient overlapping with one provides evidence of parallel trends in the pre-period. The smallest possible incident rate ratio is zero. An incidence rate ratio of one implies that the multiplier on our interaction of interest—students with disabilities in *more affected* schools in Louisiana—does not increase or decrease the rate of corporal punishment any more or less than it did in 2015–2016, the omitted year. Although the coefficient 2009–2010 differed from one, the coefficients overlapping with one in 2011–2012 and 2013–2014 provide evidence of parallel trends in later pre-treatment periods. The coefficient for 2017–2018 implies a statistically significant decrease in the rate of corporal punishment because the multiplier on the interaction term representing students with disabilities in *more affected* schools in Louisiana was only 25% of the size it was in 2015–2016. Therefore, both the OLS and PPML results suggest that Louisiana's 2017 ban significantly reduced corporal punishment for students with disabilities.

Table C1 Columns 2 and 3 show PPML results that suggest significant declines in OSS and ISS. However, parallel trends may not hold for 2009–2010 in this functional form where the outcome is expressed as an exponential conditional mean. Whereas the 2017–2018 PPML coefficient for corporal punishment is much larger than the 2009–2010 PPML coefficient, the magnitude of the coefficients for 2009–2010 and 2017–2018 are similar in the PPML results for OSS and ISS. The PPML results for OSS and ISS should be interpreted with caution given the significant pre-trend difference of a similar magnitude in 2009–2010. That said, the direction of the coefficients for OSS and ISS are consistently less than 1 for the PPML results. This suggests that on average schools did not replace corporal punishment with suspensions.

Accounting for Spillovers

Using SWDs in *less affected* schools as a placebo comparison group in the quadruple difference accounts for statewide factors that affected all SWDs in Louisiana, but it raises the risk of a spillover that could violate the Stable Unit Treatment Value Assumption (SUTVA). SUTVA implies that there are no spillover effects from treated to untreated units. The quadruple difference defines treated units as SWDs in Louisiana schools that reported at least one incident of corporal punishment for any student from 2009–2010 to 2015–2016. Schools that never reported corporal punishment during the pre-period but would have otherwise adopted the practice in 2017–2018 may have been dissuaded by the ban from using corporal punishment against SWDs that year.

I use the triple difference as a robustness check to address this concern about spillovers. The triple difference defines the treated group as all SWDs in Louisiana, regardless of whether their schools previously reported corporal punishment. This eliminates the possibility of spillovers to the *less affected* schools in Louisiana.

Appendix Table D1 provides the results of this robustness check. I find substantively similar results as those in my main specification for corporal punishment, though the coefficient is smaller because it now represents the ATT effect for all SWDs in Louisiana, not just SWDs in *more affected* Louisiana schools. This includes schools that had not previously reported corporal punishment. With the sample restricted to *more affected* schools (Table D1 Column 2), the coefficient is nearly identical to the one from the quadruple difference. In the full sample, the triple difference results suggest no significant effect on out-of-school suspension (Column 3) or in-school suspension (Column 5). Despite similar results, I prefer the quadruple difference because it accounts for additional sources of bias and provides stronger evidence of parallel pre-trends.

SUTVA violations could arise from spillovers to students without disabilities, but these are unlikely to yield biased conclusions about the effects of the ban. If the ban caused Louisiana schools to reduce corporal punishment for students without disabilities, this implies a weaker treatment contrast. That would lead to underestimating rather than overestimating the impact of the ban. It is also unlikely that the ban caused Louisiana school administrators to hit children without disabilities *more*. Corporal punishment has been declining for students with and without disabilities in Louisiana and the comparison states (U.S. Department of Education, Office for Civil Rights, 2023), and qualitative research suggests that principals see corporal punishment as tough love or an unpleasant duty (Kennedy et al., 2017). It seems unlikely, then, that principals would compensate for less corporal punishment of SWDs by subjecting more students without disabilities to corporal punishment.

Discussion & Policy Implications

I find that a state ban on corporal punishment for SWDs significantly reduced the percentage of these students subjected to corporal punishment. I rule out a host of competing explanations and contemporaneous policy changes—other than Louisiana's 2017 ban—for why the discipline rates of SWDs in *more affected* Louisiana schools uniquely declined in 2017–2018. This finding is not driven by changes in school leadership or district policy, each state's particular approach to educating SWDs, or federal guidance from the Department of Education. It is also robust to a modest amount of bias from selective underreporting. My results that Louisiana's ban significantly reduced corporal punishment contrasts with prior research suggesting that state bans on corporal punishment may be inconsequential (Curran & Kitchin, 2018; Heekes et al., 2022).

Although I find a significant reduction corporal punishment, I also find some non-compliance. This affirms prior research that even in 2017–2018, SWDs accounted for 6.7% of the corporal punishment reported in Louisiana (Dhaliwal et al., 2024). It is plausible that some SWDs in Louisiana were illegally subjected to corporal punishment because school administrators were not aware of the 2017 ban. Records of changes to Louisiana district policy manuals since 2017 suggests that school boards did not update their discipline policies immediately; they gradually updated their policies to add verbatim language from the state's ban on corporal punishment for SWDs (Claiborne Parish School Board, 2023; Livingston Parish School Board, 2017). For example, school board meeting minutes from West Carroll Parish show that the board updated their corporal punishment policy to match the state ban on November 1, 2021—more than four years late (*West Carroll Parish School Board Meeting*, 2021). Future research on discipline reform could use more formal process tracing to track the path from legislation to district policy change and school-level implementation.

The lag between state legislation and local adoption highlights implications for policy and practice. State discipline bans can reduce prohibited practices, but implementation with fidelity requires follow-up, monitoring, and accountability. Advocates, interest groups, and community organizers seeking to eliminate corporal punishment do not have mutually-exclusive options of advocating for top-down policy or grass roots change. Rather, they can pursue a two-pronged approach to support state legislation and then ensure that school boards are aware of and update their local policies and practices in a timely manner. Aligning local policies with state laws and letting school boards know when the district risks being in violation of state law may be a low-hanging fruit in enforcing discipline reforms.

My results indicate that the significant reduction in corporal punishment was not undermined by an increase in students subjected to in-school- or out-of-school suspensions. If anything, some of my results suggest that the ban may have caused slight reductions in out-of-school-suspensions. This bodes well for education reformers who wish to see a shift away from punitive discipline policies altogether. It is also good news for educators who fear that eliminating corporal punishment will force them to embrace suspensions, even if they do not see suspensions as effective. Hard-won bans on corporal punishment may accomplish their first-order goal without increasing exclusionary discipline.

Since exclusionary discipline did not replace corporal punishment after Louisiana's ban for SWDs, my findings motivate more granular data collection and in-depth qualitative research on the alternatives that administrators used instead. For example, in 2021—years after the ban for SWDs I analyze in this paper—the Superintendent of Louisiana's Iberia Parish directed principals to stop using corporal punishment for all students. To support this guidance, his administration added behavioral interventionists to two schools and expanded Saturday detention

district-wide, before/after school detention, parent conferences, and school-based social workers (*Iberia Parish School Board Meeting on September 21, 2022*, 2022). Similarly, Anderson (2018) interviewed a few educators in districts that reported the largest number of "Other Consequences" after Arkansas banned OSS for truancy and found that some schools implemented loss of course credit, extra school days, and morning, afterschool, and Saturday detentions. Large-scale data collection may overlook these punishments or lump them into an "other" category. Qualitative and mixed methods research could illuminate how administrators, principals, and teachers adapt and what changes students experience in daily school life in the aftermath of corporal punishment bans.

Endnotes

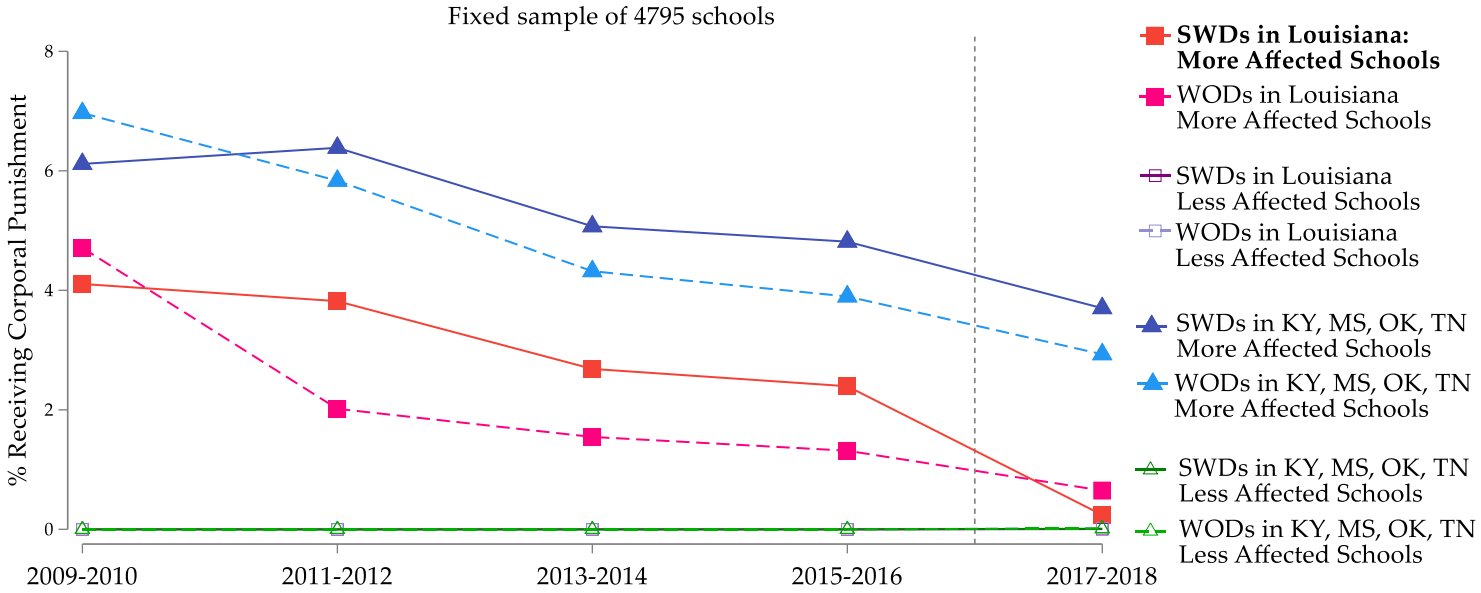
1. Enrollment for students with disabilities served under both IDEA and Section 504 are disaggregated by race/ethnicity. However, only students with disabilities served under IDEA have their discipline outcomes disaggregated by race. Discipline outcomes for students with disabilities served under Section 504 are not available separately by race/ethnicity. When calculating the percentage of Black, Hispanic, or White students with disabilities who received corporal punishment, I use only the corresponding number of students served under IDEA who received corporal punishment divided by the corresponding enrollment of students served under IDEA.

Table 1: Sample Descriptives

	Louisiana					KY, MS, OK, TN				
	2010	2012	2014	2016	2018	2010	2012	2014	2016	2018
Percent Receiving Corporal Punishment										
All students with disabilities (SWDs)	1.63	1.56	1.05	0.95	0.09	2.23	2.78	1.91	1.93	1.33
Students without disabilities (WODs)	1.82	0.82	0.58	0.52	0.28	2.43	2.14	1.60	1.51	1.08
Percent Receiving Suspensions										
SWDs receiving in-school-suspension	11.62	15.71	13.76	16.04	14.01	9.65	12.13	9.73	10.63	9.76
Students WODs receiving in-school-suspension	11.59	9.93	9.37	9.17	8.05	8.94	8.67	7.38	7.61	7.55
SWDs receiving out-of-school suspensions	11.23	17.21	14.85	17.80	14.46	7.33	11.49	9.36	11.11	9.08
Students WODs receiving out-of-school suspensions	9.86	8.75	7.91	8.26	7.59	6.18	6.35	6.17	6.00	5.53
Percent of SWD enrollment										
Black	44.22	46.19	45.93	46.26	45.88	20.03	20.60	20.00	20.02	20.38
Hispanic	0.84	1.57	2.05	2.65	4.06	3.58	3.88	5.11	6.08	7.98
White	48.33	48.28	47.84	46.55	45.83	65.00	66.86	65.37	63.78	62.61
Female	33.31	32.78	33.35	33.84	34.43	33.33	33.70	33.89	33.86	34.22
Male	66.69	67.22	66.65	66.16	65.57	66.67	66.30	66.11	66.14	65.78
Percent of Students WOD enrollment										
Black	43.70	43.47	43.34	43.09	42.97	20.11	19.64	19.54	19.31	19.18
Hispanic	3.45	4.25	5.21	6.24	6.88	6.29	7.22	8.61	9.50	10.03
White	50.04	48.40	47.16	45.89	44.90	66.78	65.45	63.55	62.19	61.09
Female	51.32	50.97	51.13	51.21	51.38	50.69	50.69	50.87	50.76	50.86
Male	45.43	45.21	45.16	44.77	44.79	45.06	45.30	45.40	45.30	44.92
Percent of overall school enrollment										
Students with disabilities	15.23	14.70	15.67	16.92	17.73	15.07	14.97	15.32	15.52	16.81
Total enrollment	535.50	545.95	553.78	546.71	539.04	546.77	546.54	547.68	551.65	541.78
Number of schools	1081	1081	1081	1081	1081	3714	3714	3714	3714	3714
Number of districts	63	63	63	63	63	501	501	501	501	501

Notes. This table shows descriptive statistics for fixed sample of 4,795 schools in Louisiana, Kentucky, Mississippi, and Tennessee that appear in the Civil Rights Data Collection from 2009–2010 to 2017–2018.

Figure 1. School Corporal Punishment Rates by Disability Status and State



Notes. This figure presents average corporal punishment rates separately for students with and without disabilities, for Louisiana and comparison states, and for *more affected* and *less affected schools*. *More affected* schools are those that reported any corporal punishment between 2009–2010 and 2015–2016. Conversely, by definition, the corporal punishment rates for *less affected* schools in all states remain at zero during the pre-period. The points are state-by-disability status-by-year averages weighted by each school's cell enrollment, the number of SWDs or students without disabilities (WOD) enrolled in each school that year.

Table 2: Effects of Louisiana's Ban on Corporal Punishment from a Quadruple Difference

VARIABLES	(1) % students receiving corporal punishment	(2) % students receiving out-of-school suspension	(3) % students receiving in-school suspension
SWD x LA x More Affected x 2009–2010	0.0917	-2.399	1.615
	(0.654)	(2.703)	(2.699)
SWD x LA x More Affected x 2011–2012	0.643	-0.460	1.094
	(0.598)	(1.702)	(2.124)
SWD x LA x More Affected x 2013–2014	0.233	0.206	1.026
	(0.432)	(1.967)	(2.037)

SWD x LA x More Affected x 2017–2018	-1.188**	-1.972	-0.901
	(0.463)	(1.656)	(1.593)
Constant	0.232	2.718**	5.967**
	(0.275)	(0.652)	(0.760)
Observations	46,462	46,036	46,092
R-squared	0.946	0.934	0.951

Notes. Coefficients are event study estimates of the impact of Louisiana's ban on corporal punishment for students with disabilities (SWDs) from a quadruple difference. Observations are school-by-year-by-disability status cells. In addition to the four-way interaction that yields the ATT estimate, all regressions include all other combinations of the indicators for SWDs, Louisiana, more affected schools, and event time. The pre-ban year 2015–2016 is omitted as a reference period. All regressions also control for school-by-year, school-by-disability status, and disability status-by year fixed effects as well as the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each cell. Standard errors are clustered at the district-by-disability-status level. Robust standard errors are shown in parentheses (*** $p<0.01$, ** $p<0.05$, $p<0.1$).

Figure 2: TWFE Event Study Regression Results

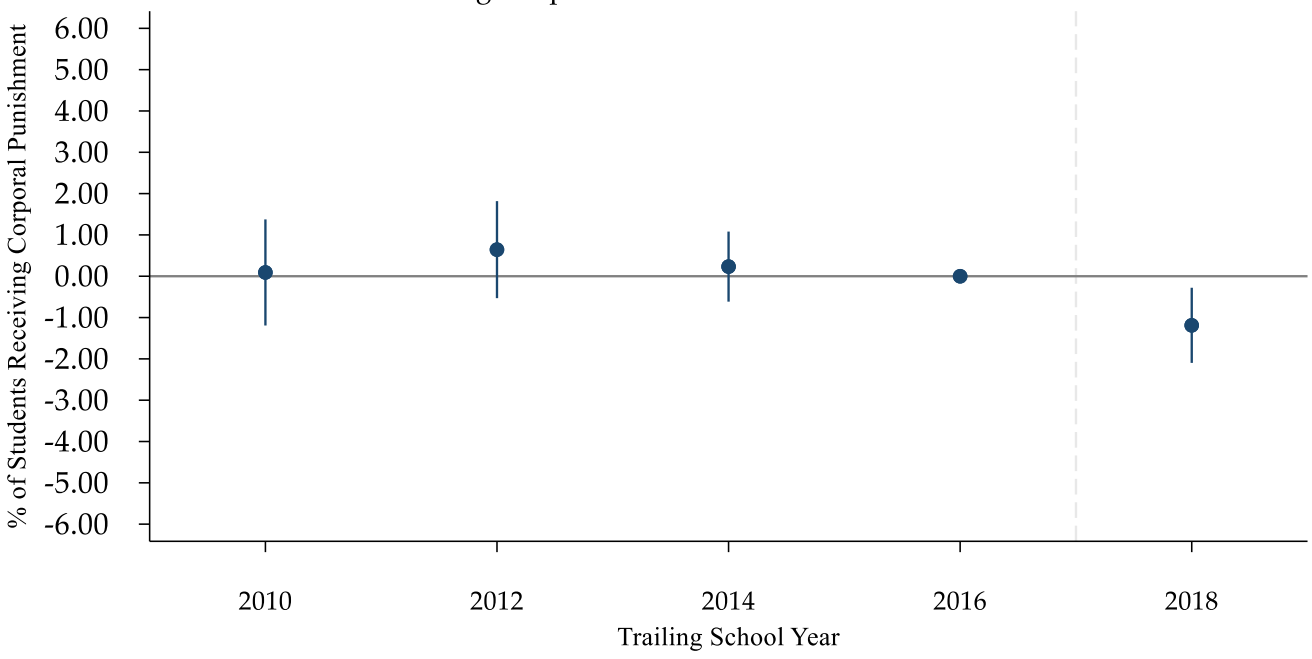


Figure 3: TWFE Event Study Regression Results

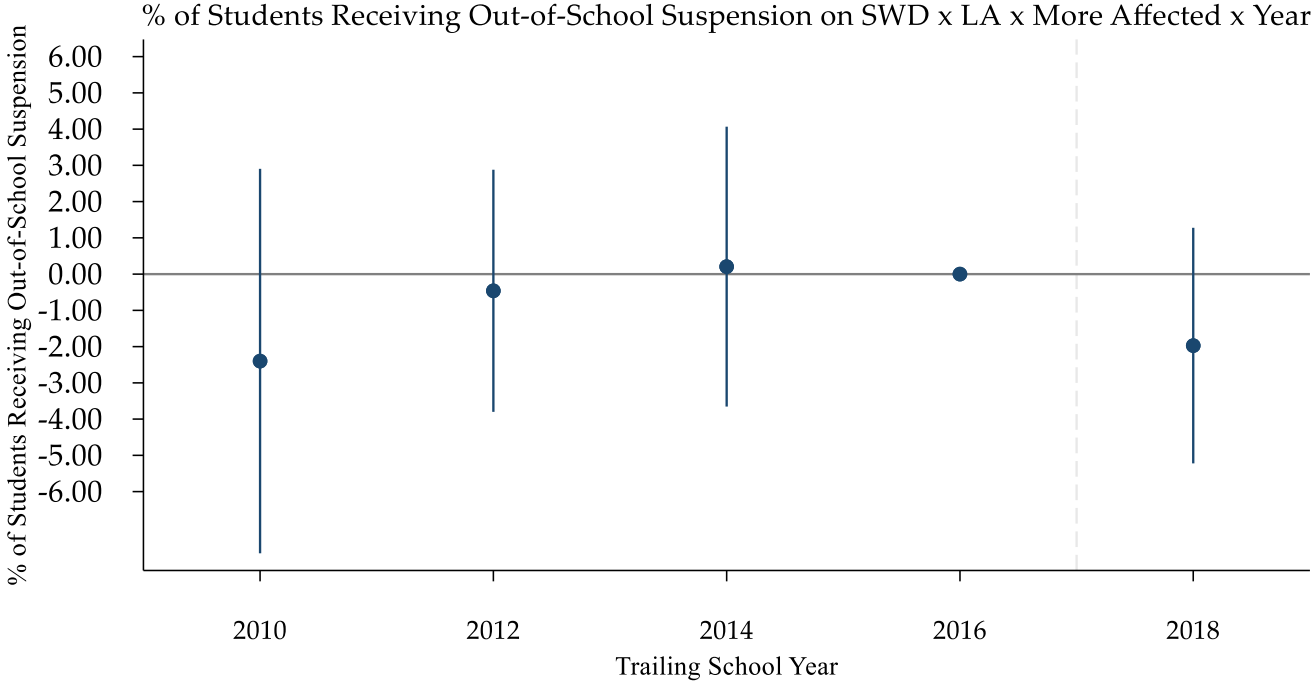


Figure 4: TWFE Event Study Regression Results

Notes for Figure 2. The coefficients plotted here are quadruple difference event study estimates of the impact of Louisiana's ban on corporal punishment for students with disabilities (SWDs) on the percentage of students receiving corporal punishment. Observations are school-by-year-by-disability status cells. In addition to the four-way interaction that yields the ATT estimate shown in these plots, all regressions include all other combinations of the indicators for SWDs, Louisiana, more affected schools, and event time. The pre-ban year 2015–2016 is omitted as a reference period. All regressions also control for school-by-year, school-by-disability status, and disability status-by year fixed effects as well as the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each cell. Standard errors are clustered at the district-by-disability-status level. Robust standard errors are shown in parentheses (*** $p<0.01$, ** $p<0.05$, $p<0.1$).

Notes for Figure 3. The coefficients plotted here are quadruple difference event study estimates of the impact of Louisiana's ban on corporal punishment for students with disabilities (SWDs) on the percentage of students receiving out-of-school suspensions. Observations are school-by-year-by-disability status cells. In addition to the four-way interaction that yields the ATT estimate shown in these plots, all regressions include all other combinations of the indicators for SWDs, Louisiana, more affected schools, and event time. The pre-ban year 2015–2016 is omitted as a reference period. All regressions also control for school-by-year, school-by-disability status, and disability status-by year fixed effects as well as the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each cell. Standard errors are clustered at the district-by-disability-status level. Robust standard errors are shown in parentheses (*** $p<0.01$, ** $p<0.05$, $p<0.1$).

Notes for Figure 4. The coefficients plotted here are quadruple difference event study estimates of the impact of Louisiana's ban on corporal punishment for students with disabilities (SWDs) on the percentage of students receiving in-school suspension. Observations are school-by-year-by-disability status cells. In addition to the four-way interaction that yields the ATT estimate shown in these plots, all regressions include all other combinations of the indicators for SWDs, Louisiana, more affected schools, and event time. The pre-ban year 2015–2016 is omitted as a reference period. All regressions also control for school-by-year, school-by-disability status, and disability status-by year fixed effects as well as the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each cell. Standard errors are clustered at the district-by-disability-status level. Robust standard errors are shown in parentheses (*** $p<0.01$, ** $p<0.05$, $p<0.1$).

Appendices

Appendix A. Robustness Checks for Inference with One Treated Cluster

Table A1. Quadruple Difference Results for Corporal Punishment using a Residual Bootstrap with Heteroskedasticity Correction for Inference with a Single Treated Cluster

	% receiving corporal punishment (CP)			
	(1)	(2)	(3)	(4)
SWD x LA x More Affected x 2017–2018	-1.441	-1.197	-1.480	-1.158
	(0.242)	(0.461)	(0.367)	(1.751)
Ferman and Pinto p-value with heteroskedasticity correction	0***	0.04***	0***	0.041***
School x year fixed effects	Yes	Yes		
School x disability status fixed effects	Yes		Yes	
Disability status x year fixed effects	Yes			Yes
Ratio of predictors (fixed effects plus regressors) to observations	0.81	0.50	0.40	0.00
Observations	46,462	46,508	47,203	47,224
R-squared	0.945	0.927	0.634	0.044

Notes. Coefficients are TWFE quadruple difference estimates of the impact of Louisiana's ban on corporal punishment for SWDs on the rate of corporal punishment. Following Ferman and Pinto (2019), these come from a specification that treats all pre-periods as a single pre-policy period rather than from the event study used in the main specification. Each coefficient estimates the ATT effect of the ban relative to the entire pre-policy period rather than to 2015–2016. Heteroskedasticity robust standard errors are clustered at the state-by-disability status-by-*more affected* level and shown in parentheses, but they may be underestimated due to the small number of clusters. Instead, the hypothesis testing is based on p-values estimated using Ferman and Pinto's (2019) residual bootstrap with heteroskedasticity correction to allow for one treated cluster (SWDs in *more affected* Louisiana schools) and 19 comparison clusters. School-by-year, school-by-disability status, and disability status-by year fixed effects are included in Column 1 to mirror the main specification. Then each set of fixed effects is included separately to avoid over-rejection that can occur in a residual bootstrap when the number of predictors approaches the number of observations (Karoui & Purdom, 2018).

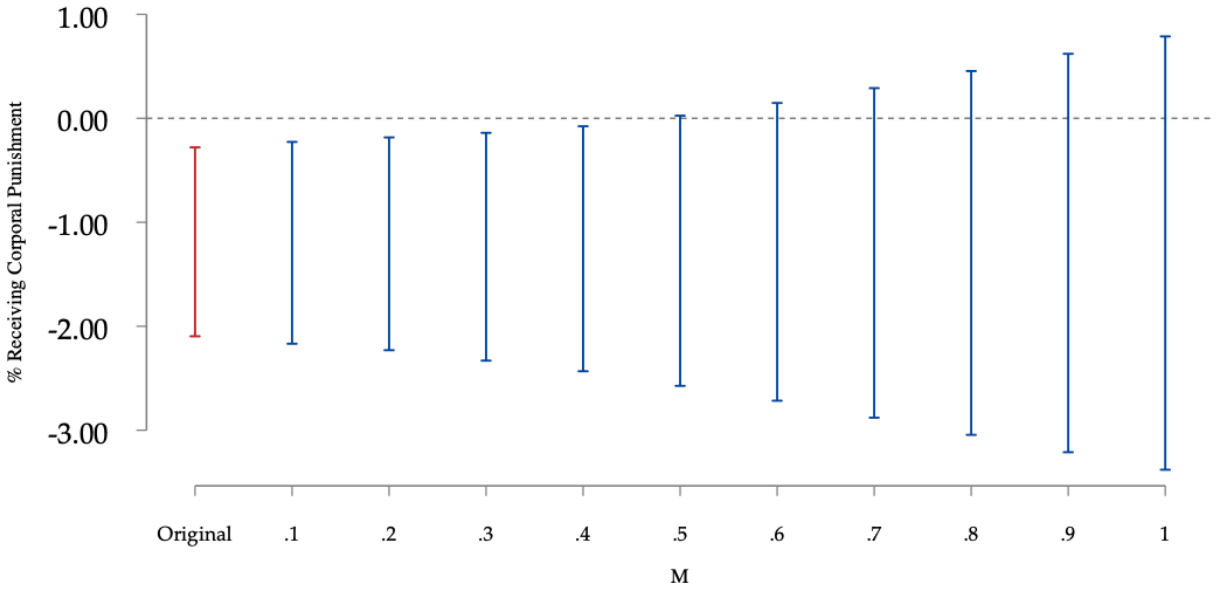
Table A2. Quadruple Difference Results for Suspension using a Residual Bootstrap with Heteroskedasticity Correction for Inference with a Single Treated Cluster

	% receiving out-of-school suspension (OSS)				% receiving in-school suspension (ISS)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SWD x LA x More Affected x 2017–2018	-1.311	-1.332	-2.154	-1.994	-1.756	-2.114	-2.186	-2.627
	(0.606)	(0.986)	(0.807)	(0.767)	(0.504)	(0.502)	(1.117)	(1.075)
Ferman and Pinto p-value with heteroskedasticity correction	0.194	0.189	0.125	0.126	0.125	0.068*	0.068*	0.068*
School x year fixed effects	Yes	Yes			Yes	Yes		
School x disability status fixed effects	Yes		Yes		Yes		Yes	
Disability status x year fixed effects	Yes			Yes	Yes			Yes
Ratio of predictors (fixed effects plus regressors) to observations	0.81	0.50	0.41	0.00	0.81	0.50	0.41	0.00
Observations	46,036	46,092	46,802	46,828	46,092	46,144	46,836	46,860
R-squared	0.933	0.890	0.729	0.178	0.951	0.926	0.687	0.064

Notes. Coefficients are TWFE quadruple difference estimates of the impact of Louisiana's ban on corporal punishment for SWDs on the rate of out-of-school and in-school suspensions. Following Ferman and Pinto (2019), these come from a specification that treats all pre-periods as a single pre-policy period rather than from the event study used in the main specification. Each coefficient estimates the ATT effect of the ban relative to the entire pre-policy period rather than to 2015–2016. Heteroskedasticity robust standard errors are clustered at the state-by-disability status-by-*more affected* level and shown in parentheses, but they may be underestimated due to the small number of clusters. Instead, the hypothesis testing is based on p-values estimated using Ferman and Pinto's (2019) residual bootstrap with heteroskedasticity correction to allow for one treated cluster (SWDs in *more affected* Louisiana schools) and 19 comparison clusters. School-by-year, school-by-disability status, and disability status-by year fixed effects are included in Columns 1 and 5 to mirror the main specification. Then each set of fixed effects is included separately to avoid over-rejection that can occur in a residual bootstrap when the number of predictors approaches the number of observations (Karoui & Purdom, 2018).

Appendix B. Sensitivity Analysis to Account for Selective Underreporting

Figure B1. Sensitivity Analysis: Robust Confidence Intervals under Violations of the Parallel Trends Assumption



Notes. The red bar shows the original confidence interval from my main specification for the estimated effect of Louisiana's 2017 ban on corporal punishment for SWDs on the rate of corporal punishment. This confidence interval corresponds to the coefficient of interest from the quadruple difference-in-differences event study specification. The blue confidence intervals correspond to a sensitivity analysis for how large bias from selective underreporting would have to be to erase the significant reduction in corporal punishment. These confidence intervals are constructed using the method proposed by Ramabachan and Roth (2023) and their `honestdid` command to accommodate the upper bound on possible biases of magnitude M relative to the worst pre-trend event study coefficient. The worst pre-trend coefficient was a positive 0.65. M = 1 implies a bias that was equally as large. The confidence intervals crossing zero at 0.5 implies that selective underreporting would have to have been half that magnitude in the opposite (negative) direction for the reduction in corporal punishment to be indistinguishable from zero.

Appendix C: Patterns in CP for SWDs by District Policy Implementation

Between February and July 2024, I gathered information on when each of the 63 Louisiana school district in my fixed sample updated their District Policy Manuals to incorporate the state's August 2017 ban on CP for SWDs. Of these, 36 districts used a third-party online system called Computer Assisted Policy Service (CAPS). Process tracing was easiest with CAPS because the manuals followed a standard structure with timestamps in the same place, but I found similar information for other districts using either a service used by districts called BoardDocs or from PDF files linked from district websites. I was able to assign a date when 53 out of 63 districts in my sample incorporated the state ban into their policy manual.

For simplicity, I describe my process referencing CAPS. Figure C1 below shows an example with screenshots from one district. CAPS displays a timestamp for when the last update occurred anywhere in the online policy manual, allowing me to confirm that the manuals were maintained well after August 2017 and often as recently as 2023 or 2024. Importantly, each district's CP policy has its own page in the CAPS online manual, each page displays the dates of each revision, and most districts edited their CP policy only once every few years. Equally important, all districts that prohibited CP for SWDs copied verbatim language from the state law:

> No form of corporal punishment shall be administered to a student with an exceptionality as defined in La. Rev. Stat. Ann. §17:1942 or to a student who has been determined to be eligible for services under *Section 504 of the Rehabilitation Act of 1973* and has an *Individual Accommodation Plan*." (Act 266 - House Bill No. 79, 2017, p. 2).

Since changes to the CP policies are few and far between, the verbatim language added after the state ban clearly implicates state law as opposed to a district's homegrown policy. Figure C2 provides a formal decision tree for how I estimated how long it took each district to incorporate the state ban on CP for SWDs into their policy manual.

Figure C1. Example Screenshots of the CP Policy Change Log from a District Policy Manual

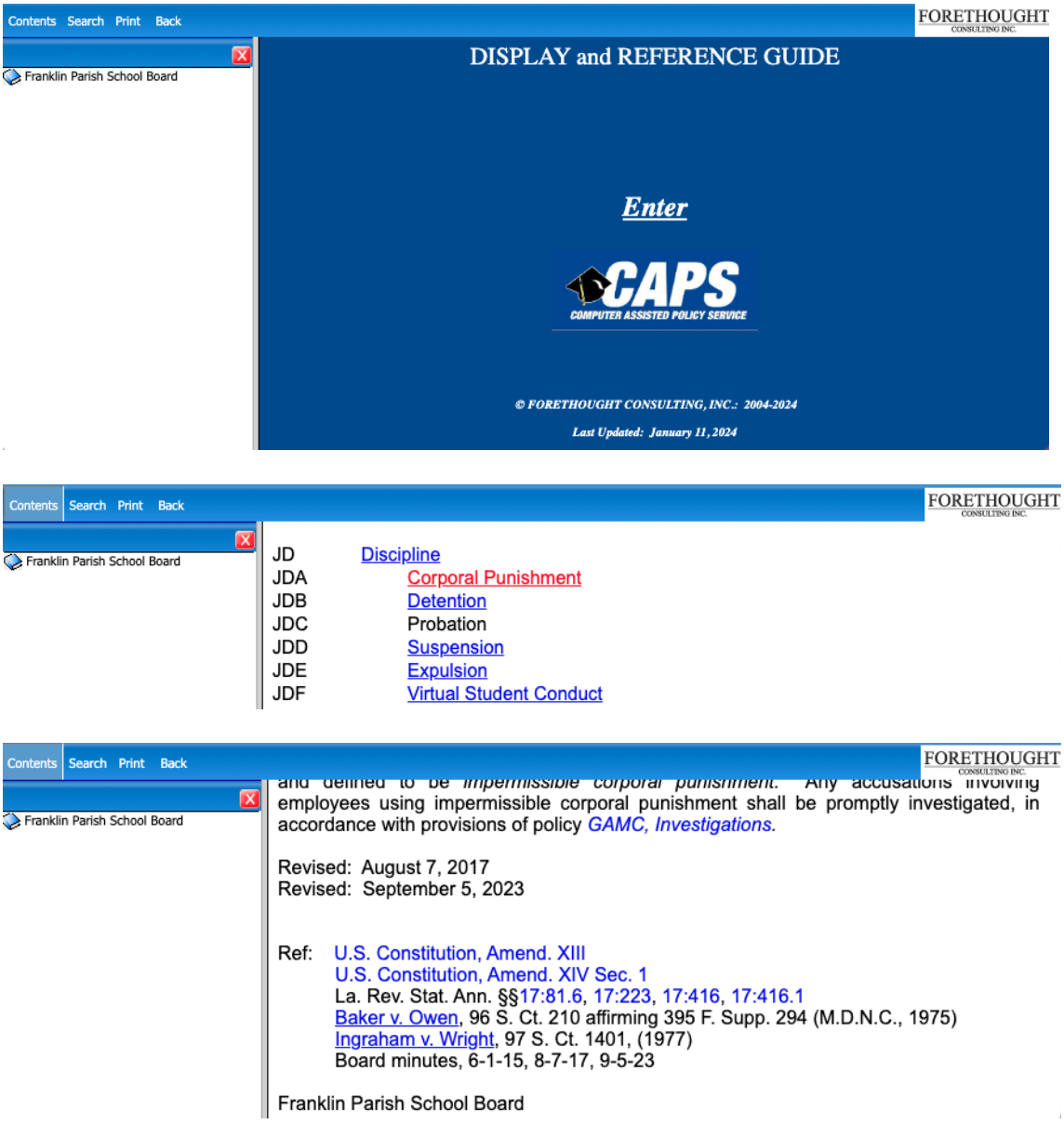


Figure C2. Decision Tree for When Districts Incorporated Louisiana's Ban on Corporal Punishment (CP) for Students with Disabilities (SWDs)

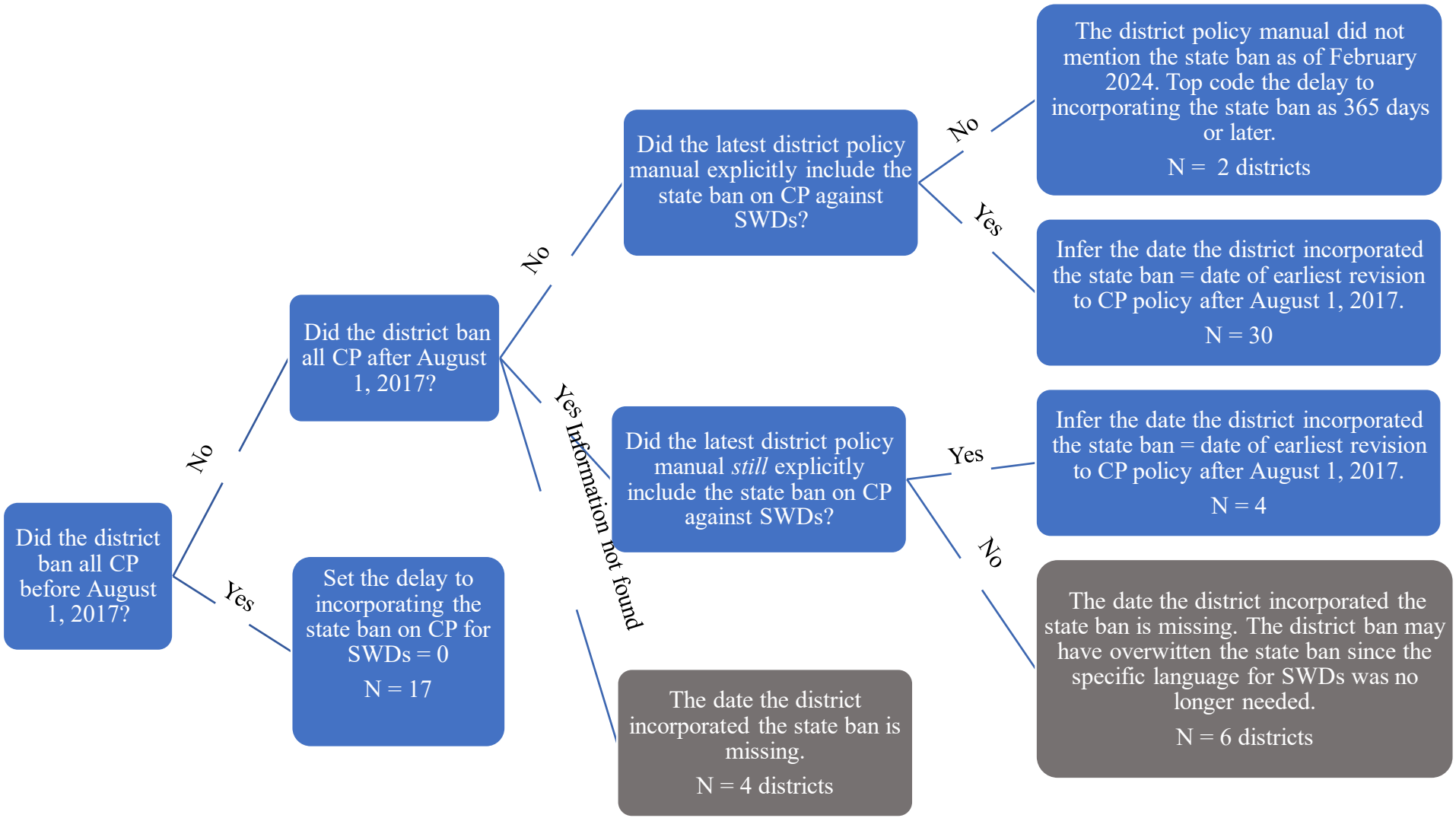
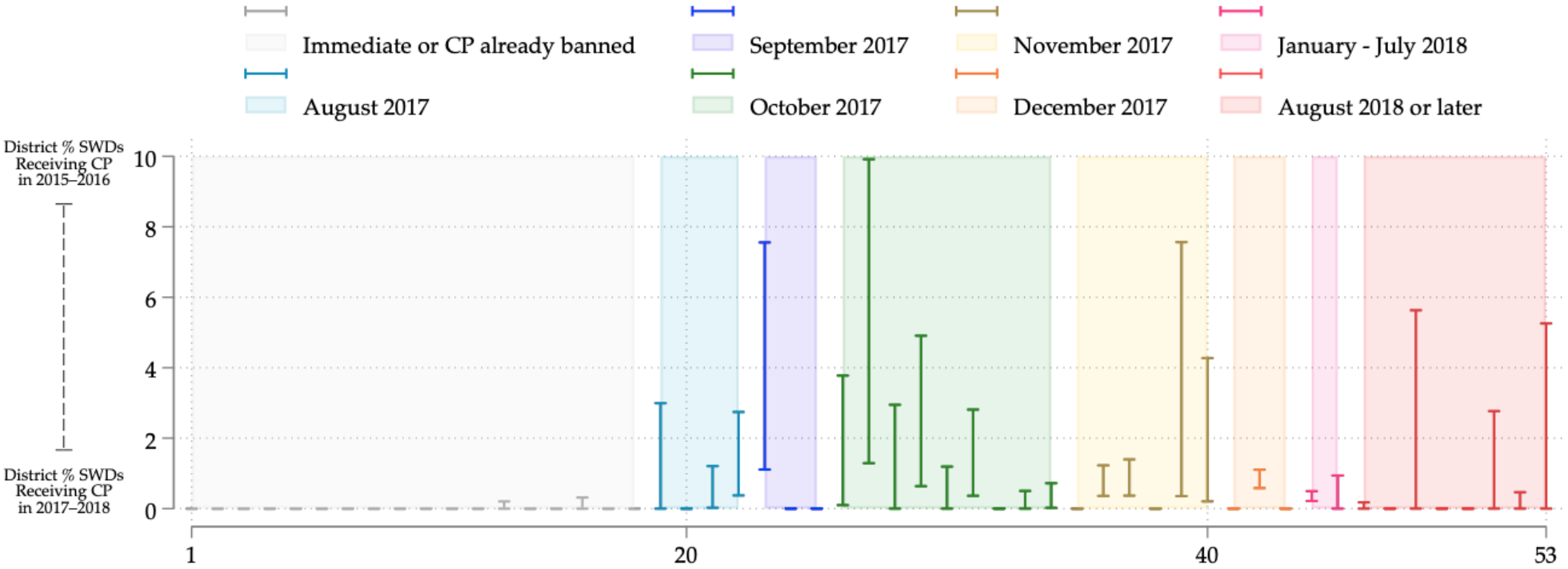


Figure C3. Decline in District-Level Corporal Punishment (CP) Rates for Students with Disabilities (SWDs) in Louisiana from 2015–2016 (top bar) to 2017–2018 (bottom bar) by Date of Revisions to the District's CP Policy

Shading and colors reflect timestamps when districts revised their policy and likely incorporated the ban on CP for SWDs.



Rank of Louisiana districts from first to last to incorporate the state ban on CP for SWDs into their policy manuals.
For readability, an arbitrary tiebreaker spreads out districts with the same rank (e.g., all the districts that had no delay).

Appendix D. Accounting for Count Data with Poisson Psuedo-Maximum Likelihood Estimation

Table D1. Effects of Louisiana's Ban on Corporal Punishment (CP) from a Quadruple Difference using Poisson Psuedo-Maximum Likelihood Estimation

	(1)	(2)	(3)
	% students receiving CP	% students receiving OSS	% students receiving ISS
VARIABLES			
Model	PPML	PPML	PPML
SWD x LA x More Affected x 2009–2010	0.770***	0.751**	0.833
	(0.0758)	(0.102)	(0.118)
SWD x LA x More Affected x 2011–2012	1.122	0.928	0.997
	(0.106)	(0.0768)	(0.0848)
SWD x LA x More Affected x 2013–2014	1.047	0.966	0.987
	(0.0838)	(0.0871)	(0.0879)
SWD x LA x More Affected x 2017–2018	0.211***	0.884**	0.845**
	(0.0415)	(0.0546)	(0.0674)
	6.473***	8.455***	13.70***
Constant	(0.837)	(0.455)	(0.884)
	0.770***	0.751**	0.833
		(0.102)	
Observations	8,574	38,512	31,244
R-squared			

Notes. Coefficients are event study estimates of the impact of Louisiana's ban on corporal punishment for students with disabilities (SWDs) from a quadruple difference using TWFE with Poisson Pseudo-Maximum Likelihood Estimation (PPML). This was estimated using the ppmlhdfe Stata command (Correia et al., 2020). Observations are school-by-year-by-disability status cells. In addition to the four-way interaction that yields the ATT estimate, all regressions include all other combinations of the indicators for SWDs, Louisiana, more affected schools, and event time. The pre-ban year 2015–2016 is omitted as a reference period. All regressions also control for school-by-year, school-by-disability status, and disability status-by year fixed effects as well as the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each cell. Standard errors are clustered at the district-by-disability-status level. Robust standard errors are shown in parentheses (*** $p<0.01$, ** $p<0.05$, $p<0.1$).

Appendix E. Accounting for Spillovers using a Triple Difference

Table E1. Effects of Louisiana's Ban on Corporal Punishment from a Triple Difference

VARIABLES	(1) % students receiving CP	(2) % students receiving CP	(3) % students receiving OSS	(4) % students receiving OSS	(5) % students receiving ISS	(6) % students receiving ISS
SWD x LA x 2009–2010	-0.0643	-0.107	-4.632***	-6.337***	-5.001***	-4.101**
	(0.242)	(0.675)	(1.639)	(2.295)	(1.710)	(1.917)
SWD x LA x 2011–2012	0.228	0.666	-1.031	-1.320	-1.828	-1.024
	(0.190)	(0.600)	(1.082)	(1.430)	(1.508)	(0.996)
SWD x LA x 2013–2014	0.0741	0.241	-1.000	-0.947	-1.852	-1.235
	(0.131)	(0.435)	(1.312)	(1.124)	(1.323)	(0.983)
SWD x LA x 2017–2018	-0.377**	-1.197**	-0.903	-2.324**	0.203	-0.484
	(0.175)	(0.465)	(1.089)	(1.072)	(0.920)	(0.981)
Samples restricted to *more affected* schools	No	Yes	No	Yes	No	Yes
Constant	1.282***	4.195***	7.387***	7.073***	10.30***	8.937***
	(0.00318)	(0.0101)	(0.0243)	(0.0339)	(0.0270)	(0.0272)
Observations	47,060	16,328	46,598	16,284	46,650	16,304
R-squared	0.945	0.934	0.933	0.930	0.951	0.942

Notes. Coefficients are event study estimates of the impact of Louisiana's ban on corporal punishment for students with disabilities (SWDs) from a triple difference using OLS regression. Observations are school-by-year-by-disability status cells. The three-way interaction between Louisiana, SWDs, and 2017–2018 yields the ATT estimate, and all regressions include all other combinations of the indicators for SWDs, Louisiana, and event time. The pre-ban year 2015–2016 is omitted as a reference period. All regressions also control for school-by-year, school-by-disability status, and disability status-by year fixed effects as well as the percentage of Black students, the percentage of Hispanic students, and the percentage of male students in each cell. Standard errors are clustered at the district-by-disability-status level. Robust standard errors are shown in parentheses (*** $p<0.01$, ** $p<0.05$, $p<0.1$).

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When Should You Adjust Standard Errors for Clustering?*. *The Quarterly Journal of Economics*, *138*(1), 1–35. https://doi.org/10.1093/qje/qjac038

Achilles, G. M., Mclaughlin, M. J., & Croninger, R. G. (2007). Sociocultural Correlates of Disciplinary Exclusion Among Students With Emotional, Behavioral, and Learning Disabilities in the SEELS National Dataset. *Journal of Emotional and Behavioral Disorders*, *15*(1), 33–45. https://doi.org/10.1177/10634266070150010401

Act 266 - House Bill No. 79, Act 266, Louisiana House of Representatives 2017 Regular Session, HB79 (2017). http://www.legis.la.gov/Legis/ViewDocument.aspx?d=1052035

Anderson, K. P. (2018). Inequitable Compliance: Implementation Failure of a Statewide Student Discipline Reform. *Peabody Journal of Education*, *93*(2), 244–263. https://doi.org/10.1080/0161956X.2018.1435052

Anderson, K. P., Egalite, A. J., & Mills, J. N. (2019). Discipline Reform: The Impact of a Statewide Ban on Suspensions for Truancy. *Journal of Education for Students Placed at Risk (JESPAR)*, *24*(1), 68–91. https://doi.org/10.1080/10824669.2018.1537794

Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). *The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime* (Working Paper 26257; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w26257

Black, D. (1993). SOCIAL CONTROL AS A DEPENDENT VARIABLE**Reprinted with permission (and minor revision) from pages 1–36 in Toward a General Theory of Social Control, Volume 1: Fundamentals, edited by Donald Black. Orlando: Academic Press, 1984.I thank M. R Baumgartner, Marc Clinton, Mark Cooney, John Griffiths, Calvin

Morrill, and Trevor Nagel for commenting on an earlier draft of this chapter. *The Social Structure of Right and Wrong*, 1–26.

Brobbey, G. (2018). Punishing the Vulnerable: Exploring Suspension Rates for Students With Learning Disabilities. *Intervention in School and Clinic*, *53*(4), 216–219. https://doi.org/10.1177/1053451217712953

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Cardona, M. (2023, March 24). *March 24, 2023—Letter from Secretary Cardona calling for an end to corporal punishment in schools* [Letters (Correspondence)]. US Department of Education (ED). https://www2.ed.gov/policy/gen/guid/secletter/230324.html

Claiborne Parish School Board. (2023). *2023-24 Claiborne Parish Student Handbook.pdf*. https://resources.finalsite.net/images/v1699455061/claibornepsborg/tkeqmpyf2lopngfc0bpp/2023-24ParishStudentHandbook.pdf

Cleveland. (2023). *Rethinking Discipline: The Effects of State Discipline Reform Laws on Students*. https://scholar.harvard.edu/sites/scholar.harvard.edu/files/chcleveland/files/cleveland_2023_rethinking_discipline_the_impacts_of_state_discipline_reform.pdf

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, *44*(4), 823–830. https://doi.org/10.1017/S1049096511001429

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, *20*(1), 95–115. https://doi.org/10.1177/1536867X20909691

Craig, A., & Martin, D. C. (2023). Discipline Reform, School Culture, and Student Achievement. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4346539

Craigie, T.-A. (2022). Do School Suspension Reforms Work? Evidence From Rhode Island. *Educational Evaluation and Policy Analysis*, *44*(4), 667–688. https://doi.org/10.3102/01623737221090264

Cruz, R. A., Firestone, A. R., & Rodl, J. E. (2021). Disproportionality Reduction in Exclusionary School Discipline: A Best-Evidence Synthesis. *Review of Educational Research*, *91*(3), 397–431. https://doi.org/10.3102/0034654321995255

Cruz, R. A., Kulkarni, S. S., & Firestone, A. R. (2021). A QuantCrit Analysis of Context, Discipline, Special Education, and Disproportionality. *AERA Open*, *7*, 23328584211041354. https://doi.org/10.1177/23328584211041354

Cuartas, J. (2022). The effect of spanking on early social-emotional skills. *Child Development*, *93*(1), 180–193. https://doi.org/10.1111/cdev.13646

Cunningham, S. (2021). *Causal Inference The Mixtape—9 Difference-in-Differences*. Yale University Press. https://mixtape.scunning.com/09-difference_in_differences

Curran, F. C., & Kitchin, J. (2018). Estimating the Relationship between Corporal Punishment Use and School Suspensions: Longitudinal Evidence from the Civil Rights Data Collection. *Peabody Journal of Education*, *93*(2), 139–160. https://doi.org/10.1080/0161956X.2018.1435036

Davison, M., Penner, A. M., Penner, E. K., Pharris-Ciurej, N., Porter, S. R., Rose, E. K., Shem-Tov, Y., & Yoo, P. (2022). School Discipline and Racial Disparities in Early Adulthood. *Educational Researcher*, *51*(3), 231–234. https://doi.org/10.3102/0013189X211061732

Dhaliwal, T. K., Graham, J., Chiang, Y.-C., & Johnson, A. S. (2024). Spare the Rod, Spoil the Child? A Critical Discourse Analysis of State Corporal Punishment Policies and Practices. *Educational Evaluation and Policy Analysis*, 01623737231213040. https://doi.org/10.3102/01623737231213040

Dillon, S. (2009, August 10). Disabled Students Are Spanked More. *The New York Times*. https://www.nytimes.com/2009/08/11/education/11punish.html

Ferman, B., & Pinto, C. (2019). Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity. *The Review of Economics and Statistics*, *101*(3), 452–467. https://doi.org/10.1162/rest_a_00759

Finkelhor, D., Turner, H., Wormuth, B. K., Vanderminden, J., & Hamby, S. (2019). Corporal Punishment: Current Rates from a National Survey. *Journal of Child and Family Studies*, *28*(7), 1991–1997. https://doi.org/10.1007/s10826-019-01426-4

Frank, K. A., Lin, Q., Xu, R., Maroulis, S., & Mueller, A. (2023). Quantifying the robustness of causal inferences: Sensitivity analysis for pragmatic social science. *Social Science Research*, *110*, 102815. https://doi.org/10.1016/j.ssresearch.2022.102815

Frank, K. A., Maroulis, S. J., Duong, M. Q., & Kelcey, B. M. (2013). What Would It Take to Change an Inference? Using Rubin's Causal Model to Interpret the Robustness of Causal Inferences. *Educational Evaluation and Policy Analysis*, *35*(4), 437–460. https://doi.org/10.3102/0162373713493129

Gershoff, E. (2010). More Harm Than Good: A Summary of Scientific Research on the Intended and Unintended Effects of Corporal Punishment on Children. *Law and Contemporary Problems*, *73*(2), 31–56.

Gershoff, E., Sattler, K. M. P., & Holden, G. W. (2019). School corporal punishment and its associations with achievement and adjustment. *Journal of Applied Developmental Psychology*, *63*, 1–8. https://doi.org/10.1016/j.appdev.2019.05.004

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, *128*(4), 539–579. https://doi.org/10.1037/0033-2909.128.4.539

Gershoff, E. T. (2017). School corporal punishment in global perspective: Prevalence, outcomes, and efforts at intervention. *Psychology, Health & Medicine*, *22*(sup1), 224–239. https://doi.org/10.1080/13548506.2016.1271955

Gershoff, E. T., & Font, S. A. (2016). Corporal Punishment in U.S. Public Schools: Prevalence, Disparities in Use, and Status in State and Federal Policy. *Social Policy Report*, *30*, 1.

Glenn, M. C. (1984). *Campaigns Against Corporal Punishment: Prisoners, Sailors, Women, and Children in Antebellum America* (Law Stacks, By Request). State University of New York Press.

Heekes, S.-L., Kruger, C. B., Lester, S. N., & Ward, C. L. (2022). A Systematic Review of Corporal Punishment in Schools: Global Prevalence and Correlates. *Trauma, Violence, & Abuse*, *23*(1), 52–72. https://doi.org/10.1177/1524838020925787

Hines, C. T., Kalil, A., & Ryan, R. M. (2022). Differences in Parents' Attitudes Toward Spanking Across Socioeconomic Status and Region, 1986–2016. *Social Indicators Research*, *160*(1), 133–158. https://doi.org/10.1007/s11205-021-02803-7

House Bill 439 (2024) - The Florida Senate, HB 439, Florida House of Representatives 2024 (2024). https://www.flsenate.gov/Session/Bill/2024/439

Iberia Parish School Board Meeting on September 21, 2022, Regular Session (2022). https://gabbart-caps.s3.amazonaws.com/iberiaCAPS/Minutes/2022-09-21.htm

Ingraham v. Wright, 430 U.S. 651 (1977), No. 75-6527 (U.S. Supreme Court April 19, 1977). https://supreme.justia.com/cases/federal/us/430/651/

Kamenetz, A. (2018, December 18). DeVos To Rescind Obama-Era Guidance On School Discipline. *NPR*. https://www.npr.org/2018/12/18/675556455/devos-to-rescind-obama-era-guidance-on-school-discipline

Karoui, N. E., & Purdom, E. (2018). Can We Trust the Bootstrap in High-dimensions? The Case of Linear Models. *Journal of Machine Learning Research*, *19*(5), 1–66.

Keierleber, M. (2021, May 19). Kids Keep Getting Hit, Even in Schools Where It's Banned. *The74Million*. https://www.the74million.org/article/kids-keep-getting-hit-at-school-even-where-corporal-punishment-is-banned/

Keierleber, M. (2024, February 15). *Florida Students Seize on 'Parental Rights' to Stop Educators From Hitting Kids*. https://www.the74million.org/article/florida-students-seize-on-parental-rights-to-stop-educators-from-hitting-kids/

Kennedy, B. L., Murphy, A. S., & Jordan, A. (2017). Title I Middle School Administrators' Beliefs and Choices about Using Corporal Punishment and Exclusionary Discipline. *American Journal of Education*, *123*(2), 243–280. https://doi.org/10.1086/689929

Kentucky Department of Education. (2022). *Guidance on 704 KAR 7:170 Corporal Punishment*.

Khafaji-King, J. A. (2024). *Disability as Discipline? Effects of the New York City Suspension Ban on Identification of Students with Disabilities*. https://jk7317.github.io/jakwebsite/disabilitydiscipline_01052024.pdf

King, Jr., J. B. (2016, November 22). *Letter to Governors and Chief School Officers Re: Corporal Punishment*. https://www2.ed.gov/documents/press-releases/11212016-corporal-punishment.pdf

Lacoe, J., & Steinberg, M. P. (2018). Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes. *Peabody Journal of Education*, *93*(2), 207–227. https://doi.org/10.1080/0161956X.2018.1435047

Lacoe, J., & Steinberg, M. P. (2019). Do Suspensions Affect Student Outcomes? *Educational Evaluation and Policy Analysis*, *41*(1), 34–62. https://doi.org/10.3102/0162373718794897

Livingston Parish School Board. (2017, August 17). *Livingston Parish School Board—Corporal Punishment*. https://cdnsm5-ss4.sharpschool.com/UserFiles/Servers/Server_546/File/Policy/Policy%202017/JDA-17.pdf

Losen, D., Whitaker, A., Kizzire, J., Savitsky, Z., & Dunn, K. (2019). *The Striking Outlier: The Persistent, Painful, and Problematic Practice of Corporal Punishment in Schools*. Southern Poverty Law Center. https://www.splcenter.org/20190611/striking-outlier-persistent-painful-and-problematic-practice-corporal-punishment-schools

Louisiana House of Representatives (Director). (2021, April 28). *LA House On-Demand Video: April 28, 2021 Louisiana House Education Committee* [Video recording]. https://house.louisiana.gov/H_Video/VideoArchivePlayer?v=house/2021/apr/0428_21_ED

Lwo, L., & Yuan, Y.-S. (2011). Teachers' Perceptions and Concerns on the Banning of Corporal Punishment and Its Alternative Disciplines. *Education and Urban Society*, *43*(2), 137–164. https://doi.org/10.1177/0013124510380232

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *232*(2), 272–299. https://doi.org/10.1016/j.jeconom.2022.04.001

MacSuga-Gage, A. S., Gage, N. A., Katsiyannis, A., Hirsch, S. E., & Kisner, H. (2021). Disproportionate Corporal Punishment of Students With Disabilities and Black and Hispanic Students. *Journal of Disability Policy Studies*, *32*(3), 212–223. https://doi.org/10.1177/1044207320949960

Mathewson, T. G. (2022a, June 6). *'State-sanctioned violence:' Inside one of the thousands of schools that still paddles students*. The Hechinger Report. http://hechingerreport.org/state-sanctioned-violence-inside-one-of-the-thousands-of-schools-that-still-paddles-students/

Mathewson, T. G. (2022b, September 1). *Missouri district brings back corporal punishment—At the urging of parents, it says*. The Hechinger Report. http://hechingerreport.org/missouri-district-brings-back-corporal-punishment-at-the-urging-of-parents-it-says/

Mweru, M. (2010). Why are Kenyan teachers still using corporal punishment eight years after a ban on corporal punishment? *Child Abuse Review*, *19*(4), 248–258. https://doi.org/10.1002/car.1121

National Center on Safe Supportive Learning Environments. (2024). *Oklahoma School Discipline Laws & Regulations: Corporal Punishment*.

https://safesupportivelearning.ed.gov/discipline-compendium?state=Oklahoma&sub_category=Corporal%20Punishment

Noakes, A. (2022, May 5). *La. House rejects HB649 to ban corporal punishment in state's public schools*. KALB 5. https://www.kalb.com/2022/05/06/la-house-rejects-hb649-ban-corporal-punishment-states-public-schools/

Office for Civil Rights. (n.d.). *Civil Rights Data Collection: State and National Estimation Data Notes*. Retrieved June 24, 2023, from https://ocrdata.ed.gov/assets/downloads/EstimationDataNotes.docx

Oklahoma State Department of Education. (2022, September). *Rubric for Determining Student Eligibility for OAAP September 2022.pdf*. https://sde.ok.gov/sites/default/files/Rubric%20for%20Determining%20Student%20Eligibility%20for%20OAAP%20September%202022.pdf

Olden, A., & Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*, *25*(3), 531–553. https://doi.org/10.1093/ectj/utac010

Olsen, J. C. (1977). Constitutional Law—Eighth Amendment—Physical Punishment in Public Schools (Ingraham v. Wright). *Marquette Law Review*, *61*.

Organda Portela, M. J., & Pells, K. (2015). *Corporal Punishment in Schools: Longitudinal Evidence from Ethiopia, India Peru and Viet Nam* (Innocenti Discussion Paper No. 2015-02; Young Lives: An International Study of Childhood Poverty). UNICEF Office of Research. https://www.unicef-irc.org/publications/pdf/CORPORAL%20PUNISHMENTfinal.pdf

Rambachan, A., & Roth, J. (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies*, *90*(5), 2555–2591. https://doi.org/10.1093/restud/rdad018

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature☆. *Journal of Econometrics*.

Ryan, R. M., Kalil, A., Ziol-Guest, K. M., & Padilla, C. (2016). Socioeconomic Gaps in Parents' Discipline Strategies From 1988 to 2011. *Pediatrics*, *138*(6), e20160720. https://doi.org/10.1542/peds.2016-0720

Santaularia, N. J., Osypuk, T. L., Ramirez, M. R., & Mason, S. M. (2022). Violence in the Great Recession. *American Journal of Epidemiology*, *191*(11), 1847–1855. https://doi.org/10.1093/aje/kwac114

Sentell, W. (2022, May 25). Louisiana House approves ban on spanking in public schools; parents could authorize punishment. *The Advocate*. https://www.theadvocate.com/baton_rouge/news/politics/legislature/louisiana-house-approves-ban-on-spanking-in-public-schools-parents-could-authorize-punishment/article_9a3aa5ec-dc69-11ec-9778-cb394677bea6.html

Sorensen, L. C., Bushway, S. D., & Gifford, E. J. (2022). Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes. *Education Finance and Policy*, *17*(2), 255–284. https://doi.org/10.1162/edfp_a_00341

Tennessee HB2330 | 2017-2018 | 110th General Assembly, 110, 900 (2018). https://legiscan.com/TN/text/HB2330/id/1802167

Title 704 Chapter 7 Regulation 170, 704 KAR 7:170 (2022). https://apps.legislature.ky.gov/law/kar/titles/704/007/170/

U.S. Department of Education, Office for Civil Rights. (2023). *Corporal Punishment in U.S. Public Schools* (Focus on Student Discipline: Part 4 of 5). https://www2.ed.gov/about/offices/list/ocr/docs/corporal-punishment-part-4.pdf

Wang, R. (2022). The Impact of Suspension Reforms on Discipline Outcomes: Evidence From California High Schools. *AERA Open*, *8*, 23328584211068067. https://doi.org/10.1177/23328584211068067

Ward, G., Petersen, N., Kupchik, A., & Pratt, J. (2021). Historic Lynching and Corporal Punishment in Contemporary Southern Schools. *Social Problems*, *68*(1), 41–62. https://doi.org/10.1093/socpro/spz044

Welsh, R. O., & Little, S. (2018). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research*, *88*(5), 752–794. https://doi.org/10.3102/0034654318791582

West Carroll Parish School Board Meeting, West Carroll Parish School Board (2021). https://www.wcpsb.com/WestCarrollCAPS/WestCarrollCAPS.htm

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data, second edition*. MIT Press.

Xu, R., Frank, K. A., Lin, Q., & Maroulis, S. J. (2024). *Quantifying the Robustness of Inferences to Replacement Data: Applications to Main Effects and Moderators*. https://www.dropbox.com/scl/fi/rpufarnepxdcyohrovwfe/aerj-conditional-RIR-full-text-10-1-24.docx?dl=0&e=1&rlkey=f4ljl0be7bzsi9czx1m3hk0il

Xu, R., Frank, K. A., Maroulis, S. J., & Rosenberg, J. M. (2019). konfound: Command to quantify robustness of causal inferences. *The Stata Journal*, *19*(3), 523–550. https://doi.org/10.1177/1536867X19874223

CHAPTER 3

How Racial Disparities Shape Black Parents' School Preferences: Evidence from a Survey Experiment

Todd Hall
University of Virginia
tmh4xa@virginia.edu

Chantal Hailey
University of Texas at Austin
chantal.a.hailey@utexas.edu

Jeremy Prim
jkprim@ucdavis.edu

Janeria Easley
janeria.easley@emory.edu

Acknowledgements: This research was generously supported by the Summer Institute for Computational Social Science (SICSS)-Howard/Mathematica, the Social Science Research Council and the Russell Sage Foundation. These materials are based on work supported by a research grant from SICSS sponsored by Howard University and Mathematica in 2021 (sicss.io/2021/howard-mathematica) organized by Naniette H. Coleman (SICSS), and co-sponsored by Calvin Hadley (Howard University), & Akira Bell (Mathematica). Todd Hall's time was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200005 to the University of Virginia. The contents of this article are those of the authors and do not represent the official views of these organizations or institutions.

Public school choice is increasingly common and accessible to nearly half of Black parents. In 2000-2001, fewer than one in three of the 112 largest school districts allowed parents to opt into a traditional (non-charter) public school outside their zoned neighborhood school within the school district. By 2015-2016, 56% of these large districts offered choice among traditional public schools within the district (Whitehurst, 2017). Nationally in 2019, nearly half of Black parents (46%) indicated that public school choices were available to them, either through multiple options for traditional public schools or charter schools, compared to 38.3 percent of White parents (National Center for Education Statistics, 2021c).

Despite Black parents' greater access than non-Black parents to public school choice, we know relatively little about how they choose schools. A large literature has documented White parents' aversion to schools with majority Black student enrollment using both observational studies and survey experiments that hold all other school characteristics equal (Billingham & Hunt, 2016; Goldring & Hausman, 1999; Hailey, 2022a; Holmes Erickson, 2017; Houston & Henig, 2021, 2022; Schneider & Buckley, 2002). We have learned much about how race factors into school choice for White parents, but large-scale studies rarely examine the school preferences of Black parents.

Early proponents argued that school choice policies would especially benefit marginalized communities (Chubb & Moe, 1990; Friedman & Friedman, 1980; Schneider et al., 1998). Other scholars have emphasized how structural constraints such as residential segregation still limit Black families' access to high achievement schools even in contexts with formal choice policies such as charter schools (Eisenlohr et al., 2023) and even for middle-class Black families with financial resources (Cooper, 2005; Lareau et al., 2021; Saporito & Lareau, 1999; Simms & Talbert, 2019; Ukanwa et al., 2022). It could be that Black families lack schools nearby with

higher achievement than the school assigned to them based on their attendance zone. Alternatively, Black families may have access to higher-achievement schools but perceive them as worse on other dimensions. For example, Eisenlohr et al. (2023) found that after charter school reforms intended to promote racial equity, Black students attended charter schools that were only marginally higher-achievement and less segregated than their neighborhood public schools. Without hearing from Black families directly, we cannot distinguish the effect of barriers to accessing higher-achievement schools from concerns that such schools of choice would be marginalizing for their Black children (Eisenlohr et al., 2023).

These concerns can be understood as expectations about school racial climate: "perceptions of interracial interactions and the socialization around race and culture in a school" Byrd (2017, p. 1). If concerns about racial inequities deter Black parents from choosing higher-achievement schools, then promoting equitable enrollment in high-quality schools requires policymakers to improve school racial climates and address educational disparities in tandem with alleviating structural access barriers.

In qualitative studies, Black parents engage in constant "racialized risk assessment", feeling torn between majority-White schools with higher test scores and racial disparities versus lower achievement schools where they worry less about their Black children being singled out (Butler & Quarles, 2024; Lareau et al., 2021; Posey-Maddox et al., 2021, p. 46). A quantitative study of school choice rankings in one district found that Black parents' preference for majority-Black schools made it harder to enact their preference for schools with high test scores, which on average had lower shares of students of color (Hastings et al., 2009). Yet the evidence on how Black parents navigate tradeoffs between diversity, potentially stigmatizing environments, and school quality is limited and mixed. More recent survey experiments that randomly varied

characteristics in school profiles found that the share of Black enrollment did not significantly affect Black parents' preferences (Hailey, 2022a; Mellon & Siegler, 2023). Another suggested that Black parents were more willing than White parents to both endure longer commutes and give up a same-race majority for a higher ranked school (Ukanwa et al., 2022). The common practice in survey experiments of varying racial composition may be insufficient for understanding how Black parents' documented desire to protect their children from negative racial climates (Cooper, 2005; Hailey, 2022a; Herelle, 2022; Lareau et al., 2021; Posey-Maddox et al., 2021) affects their school choices.

This survey experiment goes beyond racial composition and manipulates test score and suspension gaps between Black students and their non-Black peers. We randomly assign respondents in a large, national survey of Black parents (N = 1,677) to school profile vignettes including either, both, or neither of these opportunity gaps to assess how test score and suspension disparities affect Black parents' hypothetical school choice. We designed the vignette so that one focal school has higher overall achievement and lower overall suspension rates than an anchor school and should be preferred by a wide margin. In the control condition, all students at the focal school have better outcomes than those at the anchor school and there are no disparities between Black students and the school averages. In the treatment conditions, all student groups have better outcomes than the anchor school *except for Black students*, who experience gaps in either test scores, suspensions, or both. This setup allows us to answer our primary, pre-registered research question: to what extent do test score gaps and suspension gaps between Black students and their peers deter Black parents from choosing schools with higher overall test scores and lower overall suspension rates? Since Black-White achievement gaps are larger in more affluent districts and metro areas (Reardon et al., 2019), the treatment contrasts in

our experiment might mirror the tradeoffs of middle or upper middle class Black families with means to choose schools by moving to a more well-resourced district with wider racial gaps. It is unclear how well the treatment contrasts reflect options between schools within districts with formal choice policies since the evidence on how school-level disparities relate to overall school achievement is limited. We revisit the generalizability of our experiment in the discussion.

We also addressed two underpowered, exploratory questions: Do these racial disparities affect Black parents' expectations of school racial climate and belonging? Do the effects on choice and expectations vary by parent and child gender? Analyzing heterogeneity by parent gender can help address the lack of research on Black fathers' involvement in and preferences for school choice (Posey-Maddox, 2017; Reynolds et al., 2015; Wallace, 2017). Whether suspension and test score gaps will be more salient when Black parents consider their male or female children is also theoretically ambiguous. Some small-scale qualitative studies have found Black parents especially concerned about their sons' academic and disciplinary experiences, including how they might respond to bullying and deficit-oriented teacher perceptions (Cooper, 2005; Posey-Maddox, 2017). Media narratives and advocacy efforts have likewise often focused on the mistreatment of Black boys, but scholars have pointed out that Black girls have been criminalized and disproportionately subjected to exclusionary discipline as well (Morris, 2016).

Our results suggest that racial disparities in student outcomes likely dissuades some Black families from high achievement schools, but the majority would still choose a school with some disparities if it had overall higher achievement. In our experiment, racial disparities in both test scores and suspensions significantly reduced Black parents' desire for higher-achievement schools by about 0.12 standard deviations. We found that having a suspension gap between Black and non-Black students reduced the likelihood of choosing the otherwise higher-

achievement school by about 6.5 percentage points. However, Black parents were still more likely to choose the higher-achievement school with both a test score gap and a suspension gap than the anchor school with lower achievement and neither gap. Thus, racial disparities are likely insufficient to explain why Black families remain in lower-achievement schools in districts with charter schools and open enrollment, even though schools may improve their appeal Black families by closing racial gaps in student outcomes. Our results suggest that this is not an all or nothing proposition. Black parents desired the focal school more when the test score gap was eliminated even though the suspension gap remained and vice versa. Our underpowered, exploratory analyses yielded mostly null results but raised areas for future research. Overall, our findings add to the motivation for closing racial gaps and underscore the need for more engagement with Black voices to understand the opportunities and limits of school choice as a tool for racial equity.

Background: Situating and Motivating our Experiment

Our review of the literature first situates our experiment in what we know and do not yet know about how Black parents choose schools and then motivates our experiment to address these gaps. The first section describes how we draw on two theories of school choice to accommodate both rational and emotional decision-making in the school choice process. The second section explains why we manipulate test score gaps and suspension gaps. The third section motivates our use of a factorial experiment, which allows us to include separate treatment arms with test score gaps alone, suspension gaps alone, and the interaction of both gaps.

Rational and Positioned Choice: Combining Theoretical Frames

Our survey experiment straddles competing theoretical frameworks for school choice. The rational choice model applies a market logic that frames parents as utility maximizing

consumers whose decisions can be modeled as a function of specific school attributes (Eisenlohr et al., 2023; Mellon & Siegler, 2023; Ukanwa et al., 2022). Several survey experiments have employed this frame (Mellon & Siegler, 2023; Ukanwa et al., 2022), which aligns with our fundamental approach of analyzing how parents respond to information. However, our choice of racial outcome gaps as treatments is grounded in Cooper's (2005) theoretical framing of positioned choice, which accommodates the structural constraints that Black parents navigate in their decision-making.

The rational choice theory corresponds to a market-based theory of change wherein giving parents more options beyond their neighborhood public school will create competition that improves public education (Chubb & Moe, 1990; Friedman & Friedman, 1980). School choice advocates argue that parents will choose the highest quality schools as they gather information about school quality and choose to satisfy preferences as much as possible subject to constraints such as transportation (Chubb & Moe, 1990). The argument follows that competition should reward higher quality schools and force lower quality schools to close (Haderlein, 2022), while potentially reducing school segregation by decoupling school enrollment from residential segregation (Eisenlohr et al., 2023).

Some scholars have cautioned, however, that choice may fail to increase equitable access to schools or even intensify de facto racial segregation. This can occur when families rule schools out based on racial demographics rather than achievementance (Billingham & Hunt, 2016; Glazerman & Dotter, 2017; Goldring & Hausman, 1999; Hailey, 2022a; Holmes Erickson, 2017; Houston & Henig, 2021, 2023; Schneider & Buckley, 2002). Choice may also contribute to segregation rather than improvement if schools resist enrolling students they perceive as more

difficult, different racial groups have divergent preferences, or information gaps persist across racial groups (Eisenlohr et al., 2023; Mellon & Siegler, 2023; Ukanwa et al., 2022).

Qualitative and mixed method studies of Black parents' school choice processes challenge the utility maximizing framing of parents as objective, rational consumers of information (Eisenlohr et al., 2023; Herelle, 2022; Lewis-McCoy, 2016; Posey-Maddox et al., 2021; Waitoller & Super, 2017). Cooper (2005) argued that parents make positioned choices wherein race, class, and gender are inseparable from a highly subjective, emotional, and value-laden school choice process. Herrelle (2022, p. 10) quotes a Black mother and engineer who said her process for choosing a school "wasn't the most scientific." Instead, she visited the school and intuited that the teachers, administrators, and even the lunch people cared about her child. Scholars document that Black parents pay attention to how schools teach and treat Black students when considering schools (Lewis-McCoy, 2016; Posey-Maddox et al. 2021).

Black parents may be acting both rationally *and* emotionally by avoiding schools that have disparate student outcomes (even if they may have high test scores). For example, a Black parent may estimate after observing racial disparities in test scores or discipline at a school that the expected value of attending that school is lower for their Black child than for non-Black children (Hailey 2025; Lewis-McCoy, 2016). Positioned choice suggests that this information would be more salient on average for Black parents than for White parents. From either the rational or positional perspective, prior survey experiments have overlooked racial disparities as either a source of information or a trigger for an intuitive response.

Why Manipulate Suspension and Test Score Disparities

Black parents prioritize academic achievement when choosing schools (Chin, 2022; Hanson et al., 2020; Schneider et al., 1998; Teske et al., 2006; Ukanwa et al., 2022; Waitoller &

Super, 2017). Conditional on neighborhood income and prior test scores, Black, Latine, and multi-racial families' school choice rankings showed stronger preferences for high test scores than White and Asian families (Hastings et al., 2009). A recent survey experiment also found that Black parents had stronger preferences for higher school achievementance ratings than White parents (Ukanwa et al., 2022). Using unified enrollment rankings from a large urban school district, Chin (2022) found that Black parents' top choice of middle school scored higher than average on a state quality rating based on test score levels and growth. Black parents' top choice middle school was also farther from home on average than either Hispanic or White parents' top choice, suggesting a willingness among Black parents to travel (Chin, 2022; Rowley & McNeill, 2021). In New York, Abdulkadiroğlu et al. (2020) found that Black parents, like Hispanic and White parents, preferred schools with higher temst score levels, though no group prioritized schools with harder-to-observe test score value-added. Black parents interviewed about high school choice in Chicago desired schools with access to resources like tutoring and reading programs that prepared students for college (Pattillo, 2015). Some researchers have argued that Black parents' marginalized social position inclines them to emphasize academic achievementance to promote intergenerational mobility through college and career opportunities (Schneider et al., 1998; Ukanwa et al., 2022).

Black parents valued school discipline when it promoted a safe learning environment but worried about excessive discipline (Golann et al., 2019; Hanson et al., 2020; Pattillo, 2015; Schneider et al., 1998; Sempelles et al., 2024; Waitoller et al., 2019). A nationally-representative survey in 2023 found that 84% of Black parents rated "safety (including student discipline)" as very important compared to 69% of White parents (Sempelles et al., 2024). In qualitative research, Black parents approved of strict discipline to the extent that it allowed students to focus

on academics and develop self-discipline (Golann et al., 2019; Pattillo, 2015; Waitoller & Super, 2017). For example, Black mothers in two studies praised their kids' charter schools for issuing detentions for missed homework and making students complete the assignments they missed (Golann et al., 2019; Waitoller & Super, 2017). In contrast, Black charter school parents felt that punishing small infractions like humming and limiting when students could raise their hands stifled students' ability to develop self-discipline and self-advocacy (Golann et al., 2019).

Black parents also worried that discrimination, stereotypes, and bias would reduce teacher expectations and limit academic opportunities for Black children. Black parents (Bañales et al., 2020) and adults (Alesina et al., 2024) attributed Black students doing less well in school than White students more to structural factors such as school resources than to individual factors such as personal decisions. Within predominantly White, high-achieving schools, Black parents voiced concerns about their children being over-identified for special education, under-identified for gifted classes, and overlooked in the classroom (Posey-Maddox et al., 2021). Both Black and White parents cited teacher and curricular quality as reasons for choosing predominantly-Black magnet schools (Rowley & McNeill, 2021). However, the Black parents also cited negative prior experiences at schools with lower shares of Black enrollment, noting that their students faced lower teacher expectations (Rowley & McNeill, 2021).

Black parents voiced strong concerns about excessive discipline of their own and other Black children, especially when administrators seemed unfair and resistant to parental engagement. From 34 interviews, Butler and Quarles (2024) found that Black mothers worried about their children being suspended repeatedly despite their efforts to engage with the school. In another study, a Black mother switched to a private school because her son's public school kindergarten teacher automatically blamed him when he got into an argument with a little girl;

the teacher never spoke to the mother or her son (Herelle, 2022). Quantitative research substantiates concerns about Black children being perceived as more blameworthy. Using incident-level data from a large school district, Liu et al. (2022) found that when Black or Hispanic engaged in the same exact fight with White students, the Black and Hispanic students received harsher punishments. Since suspensions typically represent school administrator decisions (Gilliam & Reyes, 2018), racial discipline disparities might trigger concerns that Black students are not treated fairly (Herelle, 2022; Lareau et al., 2021; Posey-Maddox et al., 2021).

Negative Perceptions of School Racial Climate Might Deter Black Parents

School racial climate deals with perceptions of how a school context promotes equal or unequal status, positive or negative interactions, stereotypes, and various messages about race and culture (Byrd, 2017). Since racial climate refers to perceptions, two schools with the same racial composition can have different racial climates. Likewise, two students at the same school can reach different conclusions about racial climate based on both experience and observations. Byrd notes that a Black student who has never been punished at school may still perceive negative racial climate if they see other Black students being punished more frequently than their White peers. Conversely, a Black student who believes they are being discriminated against may still have a neutral or positive view of school racial climate if they feel that the school generally treats students fairly (Byrd, 2017). We focus in this study on the equal status component of racial climate, which pertains to perceptions of how students of different races are treated (Byrd, 2017).

Qualitative literature suggests that concerns about equal status may deter some Black parents from choosing schools that would otherwise meet their academic preferences. The explanation for why one Black mother in Butler and Quarles' (2024) study changed schools

typifies the concerns documented in other studies (Cooper, 2005; Herelle, 2022; Lareau et al., 2021). She explained:

> What I didn't like is the way the predominantly White staff yelled at these Black kids. Some of these Black children have parents in prison and you're rating them, labeling them, based on their behavior … I didn't like what that was modeling for my son. He had a great education. But I took him out anyway. (Butler & Quarles, 2024, p. 14).

This parent complained about the modeling or messaging that the punitive approach to behavior management sent to her Black son about other Black students. Likewise, strict discipline itself may not have been an issue. This parent remarked on what they perceived as a racial dynamic and the attachment of problem behavior to students' identities through rating and labeling. This concern about identifying Black students as persistently disruptive echoes evidence from a survey experiment. Researchers found that teachers responding to a vignette were more likely to view multiple infractions as a sign of a connected pattern when randomly assigned to a Black student rather than a White student (Okonofua & Eberhardt, 2015).

Researchers have documented widespread concerns about school racial climate among Black adults, but the implications for school choice remain unclear. A nationally-representative survey in 2022 found that Black adults were less trusting of public school teachers than non-Black adults overall and concerning issues of diversity (Phi Delta Kappan International, 2022). In a survey largely representative of the US urban population, about 60 percent of Black adults agreed that Black people often experience discrimination or are made to feel inferior because of their race at school (Alesina et al., n.d., 2024). We do not know how often such concerns about unequal status impact Black parents' school choices and are exacerbated by racial disparities. We fill this gap by first testing whether racial disparities in test scores and suspensions deter Black parents from higher-achieving schools, and then exploring a potential mechanism—more negative perceptions of racial climate.

The Current Study

The key contributions of this study are threefold. First, quantitative studies of school choice often focus on White parents and have rarely considered the preferences of parents of color at large scale. Asking Black parents about their preferences directly helps clarify what they desire in schools, allowing us to disentangle preferences from the constraints posed by transportation and residential segregation (Eisenlohr et al., 2023). Black parents may wonder if schools that are good for non-Black students are equally good for Black students. The limited evidence suggests that gains in overall achievement do not always imply gains for Black students. In statewide data from the MAP Growth assessment, schools with larger growth in reading outcomes for White students had smaller gaps in growth between Black and White students—but schools with larger growth in math for White students had larger gaps in growth rates between Black and White students (Soland, 2021).

Second, as the first to manipulate racial disparities in student outcomes, our survey experiment offers practical insight on whether schools might attract Black families by closing gaps. Prior research suggests that some Black parents face a high tradeoff between their desires for majority-Black schools and for higher-achieving schools (Hastings et al., 2009). Fixating on this tradeoff might imply that high-achievement schools without a critical mass of Black enrollment will inevitably struggle to attract more Black families. We test a more actionable hypothesis: perhaps high-achievement schools can attract Black families by demonstrating outcomes for Black students on par with the school's overall outcomes. We argue in the discussion that this hypothesis is useful to test even if Black parents care only about the levels of Black students' outcomes and not about racial gaps.

Our factorial design enhances practical utility by testing the main effects of test score gaps on Black parents' desire for schools regardless of the suspension gaps and vice versa. This matters because some practitioners and policymakers may find it easier to mitigate test score gaps than suspension gaps. Interventions such as high-dosage tutoring (Fryer & Howard-Noveck, 2020; Kraft, 2015) and "No Excuses" urban charter schools have boosted test scores more for Black students than for White students (Angrist et al., 2013), showing potential to reduce test score gaps. With the exception of restorative practices (Adukia et al., 2024; Augustine et al., 2018), few interventions have been found to reduce Black-White suspension gaps (Welsh, 2023; Welsh & Little, 2018). Analyzing the main effects of suspension and test score gaps on parental preferences helps us to consider how schools might attract Black families by first closing the gap that is the lower-hanging fruit in their context. We also assess whether having both test score and suspension gaps cause an interaction effect. This potential interaction could affect many schools assuming that test score and suspension gaps are correlated at the school level like they are at the district level (Pearman et al., 2019).

Third, we consider Black fathers. Black fathers are rarely included in studies on school involvement, where the samples of parents often exclusively or predominantly consist of mothers (Posey-Maddox, 2017; Reynolds et al., 2015; Wallace, 2017). In qualitative research, Black fathers reported that school officials were noticeably surprised by their involvement; some explicitly remarked that Black fathers rarely show up, making comments that Black fathers felt could dissuade others from engaging (Allen, 2013; Reynolds et al., 2015). Although Black fathers may navigate gendered stereotypes about being absent, there is little evidence on whether Black fathers are sensitive to racial disparities when considering schools.

Method

Experimental Design

We recruited Black parents in the U.S. from the Centiment survey panel. Centiment maintains a panel by recruiting people from social media (Facebook, LinkedIn, etc.), conducting security screening, and collecting detailed demographic and consumption data to create a panelist profile. Centiment only invited the panelists who were eligible to take our survey based on their profiles, i.e., who had previously self-identified as Black and as a parent. If a panelist qualifies for a survey based on eligibility questions answered more than 30 days ago, Centiment asks the questions again before showing the survey for additional quality control (Reuther, n.d.). Similar to other panels (Cobanoglu et al., 2021; Ford, 2017; Smith et al., 2016), Centiment monitors responses across surveys to ensure that respondents identify themselves consistently. Beyond Centiment targeting panelists based on our eligibility criteria, the beginning of our survey asked respondents to mark all the races and ethnicities that apply to them.[1] Respondents could select any combination of races or ethnicities, but, unbeknownst to them, those who did not include "Black or African American" among their selections were barred from completing the survey. We also asked respondents how many children under age 18 lived with them. We restrict our sample to parents living with children under 18, excluding those who skipped this question or reported zero children at home.

We pre-registered our experiment on the Open Science Framework (https://osf.io/4zegs) and collected two waves of data: one in June 2023 (N = 776 Black parents) and one in February 2024 (N = 901 Black parents). Respondents in both waves encountered the same vignette:

> Please imagine that your oldest child is about to enter middle school for the first time this fall. You have two public middle schools to choose from: Walker and Prim.

[1] We concealed our restriction to Black parents to avoid priming respondents to focus more on race and to prevent non-Black respondents from pretending to be Black. If we allowed non-Black respondents to complete almost the entire survey before revealing that they were not eligible, it would only be ethical to pay them. Putting demographic questions first allowed us to screen out non-Black respondents before they completed the survey.

> You visit your state's official school directory and see the school profiles for Walker and Prim on the next page. Please consider the two school profiles. Then you will answer questions about your preferences.

The next page showed respondents two school profiles, one for a focal school that we manipulated based on random assignment and one for an anchor school that never varied. As in prior survey experiments (Mellon & Siegler, 2023; Ukanwa et al., 2022), we modeled our profiles after state- district-run school search websites that are frequented by parents (Virginia Department of Education, 2022). Prim Middle School is the focal school whose profile was randomly assigned to a control or one of three treatment conditions shown in Table 1 below. The profile for Prim when assigned to the "both gaps" condition is shown in Figure 1 below.

Table 1: Randomly Assigned Conditions

		Test score gap between Black and non-Black students at Prim Middle School?	
		No	Yes
Suspension gap between Black and non-Black students at Prim Middle School?	No	Control (No gaps) [-1, -1]	Treatment 1 (Test Score Gap Only) [-1, 1]
	Yes	Treatment 2 (Suspension Gap Only) [1, -1]	Treatment 3 (Both Gaps) [1, 1]

The Walker Middle School profile is an anchor with all details held constant regardless of random assignment (Appendix A). This anchor allows all respondents to picture the same alternative to Prim, the focal school. Consider Walker part of the instructions, not a control group or an arm of the experiment because it never varies across respondents.

For ease of comparison, we show the information used in the profiles for Prim and Walker before showing example profiles. Table 2 shows the information for Prim Middle School, the focal school, and for Walker Middle School, the fixed anchor school. Information in Panel A

did not vary by randomly assigned treatment status. Panel A shows that Walker and Prim always had the same student-teacher ratio, distance from home, and accreditation status. Prim Middle School had a slightly higher share of teachers with three or more years of experience, and a lower share of Black students, though both schools are racially and ethnically diverse. Prim had a smaller share than Walker of economically disadvantaged students, students with disabilities, and English Learners. Our intention was to differentiate Walker and Prim just enough to make the options seem more realistic and to avoid the priming the respondents out by having student outcomes be the only things that differed across profiles.

Regardless of the randomly assigned condition, Prim scored higher than Walker in overall test score proficiency by 10 percentage points in English (75% vs. 65%) and 9 percentage points in math (75% vs. 66%). In all conditions, Prim also had lower overall suspension rates (4% vs. 7%). At Walker, Black students always had the same proficiency rates as the school average, 65% in English and 66% in math, and a 6% suspension rate (Appendix A).

Table 2: Summary of School Profile Information

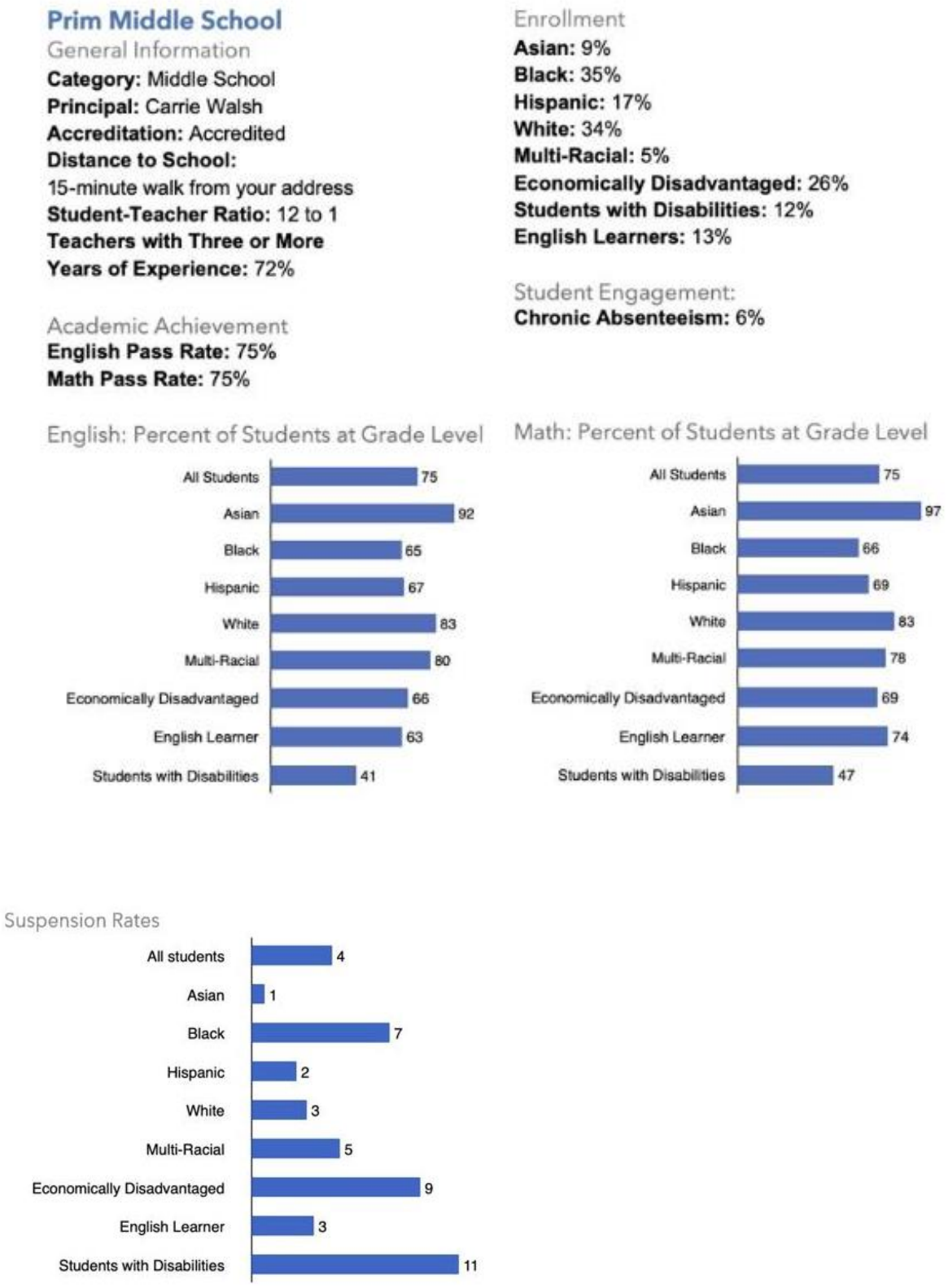
Panel A: Information Held Constant (Not Manipulated with Treatment Status)

Name	Prim Middle School	Walker Middle School
Category	Middle School	Middle School
Principal	Carrie Walsh	John Thomas
Accreditation	Accredited	Accredited
Distance to School	15-minute walk from your address	15-minute walk from your address
Student-Teacher Ratio	12 to 1	12 to 1
Teachers with Three or More Years of Experience	72%	68%
Enrollment	Asian: 9% Black: 35% Hispanic: 17% White: 34% Multi-Racial: 5% Economically Disadvantaged: 26% Students with Disabilities: 12% English Learners: 13%	Asian: 5% Black: 42% Hispanic: 20% White: 29% Multi-Racial: 4% Economically Disadvantaged: 32% Students with Disabilities: 13% English Learners: 16%
Student Engagement	Chronic Absenteeism: 6%	Chronic Absenteeism: 8%
English Pass Rate	75%	65%
Math Pass Rate	75%	66%
Overall Suspension Rate	4%	7%

Panel B: Randomized Levels for Prim Middle School (Bold Indicates Difference from the Control)

		Control (No gaps)	Treatment 1 (Test Score Gap Only)	Treatment 2 (Suspension Gap Only)	Treatment 3 (Both Gaps)
English: Percent of Students at Grade Level	Overall	75%	75%	75%	75%
	Asian	93%	**92%**	93%	**92%**
	Black	74%	**65%**	74%	**65%**
	Hispanic	71%	**67%**	71%	**67%**
	White	74%	**83%**	74%	**83%**
	Multi-Racial	82%	**80%**	82%	**80%**
	Economically Disadvantaged	66%	66%	66%	66%
	English Learner	63%	63%	63%	63%
	Students with Disabilities	41%	41%	41%	41%
Math: Percent of Students at Grade Level	Overall	75%	75%	75%	75%
	Asian	95%	**97%**	95%	**97%**
	Black	74%	**66%**	74%	**66%**
	Hispanic	76%	**69%**	76%	**69%**
	White	74%	**83%**	74%	**83%**
	Multi-Racial	85%	**78%**	85%	**78%**
	Economically Disadvantaged	69%	69%	69%	69%
	English Learner	74%	74%%	74%	74%%
	Students with Disabilities	47%	47%	47%	47%
Suspension Rates	Overall	4%	4%	4%	4%
	Asian	1%	1%	1%	1%
	Black	3%	3%	**7%**	**7%**
	Hispanic	5%	5%	**2%**	**2%**
	White	5%	5%	3%	3%
	Multi-Racial	5%	5%	5%	5%
	Economically Disadvantaged	9%	9%	9%	9%
	English Learner	5%	5%	**3%**	3%
	Students with Disabilities	9%	9%	**11%**	**11%**

Figure 1: Prim Middle School Profile



Panel B shows information that varied by treatment status. In the control condition, Prim had no gaps between Black students and their peers in either test scores or suspension rates. In the treatment conditions, the Prim profile had either (1) a gap between Black students and their peers in math and reading test score proficiency rates, (2) a gap between Black students and their peers in suspension rates, or (3) gaps in both test scores and suspension rates. Although recent work suggests that displaying test score growth instead of status may prompt White parents to consider more racially diverse schools, we used test score proficiency for simplicity and because test scores remain ubiquitous on state school performance profiles whereas growth data are more often supplemental (Houston & Henig, 2023).

After viewing the profiles, respondents were asked about their desire to enroll in Prim Middle School, their perceptions about the school's racial climate, and expected belonging at the school. The items are described in the "Measures" section below and in Appendix B. One page asked about desire to enroll and choice in both waves one and two. In wave two only, respondents encountered a second page of questions about school racial climate and expected student belonging. Parents in wave two could not go back to change their answers after moving on, so these questions could not influence parents' responses about desire to enroll and choice. A final page in wave two asked parents which factors influenced their responses on the previous pages. They could mark all that apply for eight different factors, where "math test scores", "reading test scores", "suspension rates", and "gaps in student outcomes" were included as separate options. Parents did not have the option to go back and change prior answers. This precludes any priming from having "gaps in student outcomes" as an option.

The experiment was the same in the two survey waves with some exceptions. First, the second wave asked respondents how they would describe the gender of their oldest child, whom

the vignette referred to. This question was added to address an exploratory research question about whether effects of test score and suspension gaps on parent preferences differed by child gender. Second, the first wave asked parents whether they would like to view school suspension rates. We included this step because test scores often appear front and center on school search websites, but discipline rates can require additional clicks. The test score gap did not significantly affect the likelihood of viewing the school suspension rates and the likelihood of viewing suspension rates did not differ significantly across treatment arms (see Appendix E). The second wave removed this step and simply showed parents the school suspension rates. This eliminates the possibility that asking parents if they want to see suspension rates might prime them to focus more on the suspensions than they otherwise would. Third, the second wave corrected a small typo (a floating number "15") in the school profile with the test-score gap treatment (See Appendix C). The typo did not cover any information being experimentally manipulated, nor did parents in cognitive interviews mention it. Fourth, the second wave added a series of questions about parent expectations of school racial climate and student belonging. Finally, the last question in the second wave asked parents to mark all factors from a predefined list that influenced their answers about the schools. For items collected in both waves, we pool the sample across waves with fixed effects for wave to maximize statistical power but report regression analyses separately by wave in Appendix E.

Sample

Table 3 below describes the full sample pooled across waves. Most of the parents identified as only Black (91.6 percent), though some were multiracial. Approximately half our respondents identified as male. About two-thirds were between the ages of 45 and 64, and one-third were married. Slightly more than one in five completed a BA degree or higher while about

30 percent reported high school as their highest degree. Nearly half resided in urban areas and more than a third lived in the South. Appendix F compares our sample to the available national estimates for Black parents in the U.S. or, if unavailable, Black adults and Black public-school students in the U.S. Our respondents were representative of U.S. Black parents in marital status and some levels of education (National Center for Education Statistics, 2023), but they were less concentrated in the South than the population of U.S. Black adults (Moslimani et al., 2024). Although this may limit external validity, our results may still be informative. The effects of information experiments in online samples were strongly correlated with the effects of the same information interventions in nationally-representative samples (Coppock, 2019).

Table 3: Sample Descriptives (Pooled Across Waves)

Race/ethnicity	
Black only	91.6%
Black multiracial/ethnic	8.4%
Male	50.1%
Age	
Under 25	8.8%
25-34	31.9%
35-44	36.6%
45-54	16.5%
55+	6.2%
Marital/cohabitation Status	
Married	33.2%
Cohabiting	20.3%
Kids at home	1.88
Highest level of education	
BA degree or higher	22.7%
Associates degree	11.1%
Some college	25.9%
High school	29.9%
Less than High school	5.5%
K-12 teaching experience	
Previously taught K-12	10.4%
Teaches K-12 now	3.8%

Locale		
	Rural	27.7%
	Suburban	25.0%
	Town	0.3%
	Urban	47.0%
Region		
	Northeast	20.4%
	Midwest	25.4%
	South	38.2%
	West	15.0%
N		1677

Note: Not all categories sum to 100 because some category levels are not shown. For example, divorced is not shown for marital status and trade school is not shown for education.

Covariate Balance

Overall, randomization was fairly successful in producing covariate balance. The balance table below shows results from regressing the effect-coded treatment variables on baseline covariates in the pooled sample. The F-test statistics and p-values for joint significance in Columns 5 and 6 assess whether the treatment assignment predicts each covariate. In the pooled sample across waves, two of the 25 F-statistics (8%) were statistically significant at the 0.05 level. This is slightly more than the 5% expected by chance and is driven by differences across experimental conditions in the distribution of age and locale, specifically whether the respondent lived in a suburb. In wave one, only two of the 25 F-tests (8%) were statistically significant at the 0.05 level, but there were three significant F-statistics in wave two (12%). Appendix D Table 1 shows standardized mean differences (effect sizes) between baseline characteristics in the treatment and control conditions for the pooled sample. Most of the standardized differences are close to zero, suggesting that the baseline differences across treatment conditions were small. Moreover, the significant F-statistic in one age bands is less concerning because the age distribution as a whole seems balanced. It is unclear whether the slightly higher share of

respondents who encountered a suspension gap that were suburban might introduce a positive or negative bias. The higher suburban share could inflate the presumably negative effect of the suspension gap if they are more used to high-achievement schools and thus are less impressed by Prim for reasons unrelated to the suspension gaps. For the same reason, however, the higher suburban share could reduce negative the effect of the suspension gap if these respondents are unwilling to consider Walker, the alternative, due to its lower-achievement. Whatever the potential bias, the slight imbalances in the pooled sample support the inclusion of covariates in our regressions, consistent with our pre-registration plan.

Table 4. Covariate Balance in the Pooled Sample

	(1)	Contrasts by Treatment Status (2)	(3)	(4)	(5)	(6)	(7)
	Grand Mean	Test Score Gap vs. No Test Score Gap	Suspension Gap vs. No Suspension Gap	Interaction Effect	F-statistic	p-value (all = grand mean)	Obs
Male	0.501***	0.00946	0.00275	-0.000538	0.219	0.883	1675
	(0.0122)	(0.0122)	(0.0122)	(0.0122)			
Black Multiracial/ethnic	0.0842***	-0.000353	-0.0115*	-0.00880	1.593	0.189	1677
	(0.00678)	(0.00678)	(0.00677)	(0.00677)			
Married	0.332***	0.00297	0.0146	0.00226	0.578	0.629	1677
	(0.0115)	(0.0115)	(0.0115)	(0.0115)			
Cohabiting	0.203***	0.00451	-0.00589	0.000381	0.191	0.903	1677
	(0.00983)	(0.00983)	(0.00983)	(0.00985)			
Less than High School	0.0554***	-0.0101*	0.00184	0.00240	1.242	0.293	1677
	(0.00559)	(0.00559)	(0.00558)	(0.00559)			
High School	0.299***	0.00833	-0.00938	0.00444	0.479	0.697	1677
	(0.0112)	(0.0112)	(0.0112)	(0.0112)			
Associate's Degree	0.111***	0.000170	0.00352	-0.00347	0.140	0.936	1677
	(0.00768)	(0.00768)	(0.00767)	(0.00769)			
Some College	0.259***	-0.00248	-0.00187	-0.00672	0.160	0.923	1677
	(0.0107)	(0.0107)	(0.0107)	(0.0107)			
BA degree or higher	0.227***	0.00506	0.00929	0.00715	0.501	0.682	1677
	(0.0102)	(0.0102)	(0.0102)	(0.0102)			
Previously taught K-12	0.104***	0.00973	-0.00382	0.0131*	1.882	0.131	1677
	(0.00744)	(0.00744)	(0.00744)	(0.00745)			
Teaches K-12 now	0.0375***	-0.00534	0.00173	0.00128	0.546	0.651	1677
	(0.00464)	(0.00464)	(0.00464)	(0.00463)			
Age Under 24	0.0882***	-0.00835	-0.0168**	0.00190	2.584	0.0518*	1677
	(0.00663)	(0.00663)	(0.00664)	(0.00664)			
Age 25-34	0.319***	0.000880	0.0153	-0.0106	0.900	0.440	1677
	(0.0114)	(0.0114)	(0.0114)	(0.0114)			
Age 35-44	0.365***	-0.0154	0.0131	0.00738	1.148	0.328	1677
	(0.0117)	(0.0117)	(0.0117)	(0.0117)			
Age 45-54x	0.165***	0.0216**	-0.0192**	-0.00469	3.336	0.0187**	1677
	(0.00897)	(0.00898)	(0.00897)	(0.00898)			
Age 55+	0.0620***	0.00133	0.00755	0.00599	0.876	0.453	1677
	(0.00585)	(0.00585)	(0.00586)	(0.00587)			
Suburban	0.250***	-0.0114	0.0384***	0.0132	5.698	0.000703***	1677
	(0.0105)	(0.0105)	(0.0105)	(0.0105)			
Rural	0.277***	0.00898	-0.0110	-0.0195*	1.545	0.201	1677
	(0.0109)	(0.0109)	(0.0109)	(0.0109)			
Town	0.00298**	0.00299**	0.000591	0.000604	1.674	0.171	1677
	(0.00133)	(0.00133)	(0.00133)	(0.00132)			
Urban	0.470***	-0.000604	-0.0280**	0.00569	1.841	0.138	1677
	(0.0122)	(0.0122)	(0.0122)	(0.0122)			
Northeast	0.215***	0.000102	0.0112	0.0136	1.031	0.378	1677
	(0.00998)	(0.00998)	(0.00998)	(0.00998)			
Midwest	0.254***	-0.0127	-0.00788	-0.0108	1.073	0.359	1677
	(0.0106)	(0.0106)	(0.0106)	(0.0106)			
South	0.150***	0.00571	-0.0117	-0.00243	0.740	0.528	1677
	(0.00872)	(0.00872)	(0.00872)	(0.00871)			

West	0.382***	0.00694	0.00837	-0.000411	0.298	0.827	1677
	(0.0117)	(0.0117)	(0.0117)	(0.0117)			
Kids under 18 at Home	1.880***	-0.0373	-0.00724	0.0302	1.093	0.351	1677
	(0.0280)	(0.0280)	(0.0280)	(0.0279)			

Robust standard errors in parentheses

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

We conduct a supplementary balance test based on randomization inference to address the shortcomings of counting the fraction of significant p-values. Kerwin et al. (Kerwin et al., 2024) explain that the fraction of p-values below 0.05 under the null hypothesis of no differences across experimental conditions is itself a random variable. The distribution has a mean of 5%, but the fraction of significant p-values in any one draw from that null distribution is very likely to exceed 5%. Using simulations, they show that assessing balance by counting significant p-values consistently over-rejects the null hypothesis. Kerwin et al. (2024) instead recommend a balanced test based on randomization inference. Following their procedure, we repeatedly randomize placebo treatments and regresses the placebo experimental conditions on the baseline covariates 1000 times using a multinomial logit, yielding 1,000 chi-square statistics for whether the covariates predict the experimental condition. This distribution of test statistics reflects a sharp null hypothesis of no effect of covariates on experimental condition for all observations. We then compute a p-value by noting how extreme the chi-square statistic is for the real data relative to the simulated sharp null distribution, e.g., whether it's above the 95th percentile. This process yields a p-value of 0.203, increasing our confidence that randomization was fairly successful.

Measures

Desire to Enroll. Our pre-registered, confirmatory outcome is a factor score based on a six-item scale for a construct we named desire to enroll. This score captures the parents' desire to enroll their child in Prim Middle School, the focal school we experimentally manipulated. The items that comprise the scale are listed below in Table 3 along with their factor loadings. We show the distribution of each item by experimental condition in the "Results" section. We developed the items iteratively with two sets of input: literature and piloting.

From the school choice literature, we developed a construct map for desire to enroll with multiple levels reflecting both choices and emotions. For example, middle-class Black parents expressed apprehension about avoiding low-achievement schools (Lareau et al., 2021). Some Black and White middle-class parents second-guessed their choice of an urban public school (Cucchiara, 2013). Low-income parents in a system of ranked school choice felt frustrated and disappointed when they registered late—often amid housing instability—and got assigned to leftover schools that they had not ranked (Fong & Faude, 2018). Meanwhile, some parents did not actively consider or compare schools when choice within the district was available but enrolled in their zoned school by default (Goldring & Hausman, 1999). We concluded that our items should together distinguish parents with a strong desire to attend or avoid a school from those who passively found it acceptable. While some parents with financial means chose where to live based on the public schools (Cuddy et al., 2020; Kimelberg, 2014; National Center for Education Statistics, 2021b), we did not want to conflate desire with financial means. Thus our items refrained from language in prior survey experiments that parents would have to pay tuition or move to avoid the hypothetical school (Billingham et al., 2020; Billingham & Hunt, 2016).

We piloted five items intended to allow range in desire to enroll—from aversion to acceptance to enthusiasm—with a convenience sample of 17 graduate students at a Southeastern University and in cognitive interviews with three Black parents. With the pilot feedback and item-total correlations, we retained “Avoid”, “Disappoint,” and “Enroll”; changed an item about ranking schools into an easier-to-understand item about choosing between schools; and added two items. In our final scale, two items represent a discrete choice (whether to enroll and whether to choose Prim over the alternative), three capture emotions (excitement, enthusiasm about the schools’ fit, and disappointment), and one considers desire to avoid a school.

We computed desire to enroll factor scores separately for each wave based on item analyses and our preregistration. Table 5 below shows the factor loadings. Note that we reverse coded the negatively worded items (Disappoint and Avoid) so that lower numbers (e.g., 1 "Strongly Disgree") represented a high desire to enroll, consistent with the other items. The eigenvalues suggested a two-factor solution, where Excited, Fit, Choose, and Enroll loaded more onto one factor while the negatively worded items were more correlated with each other. Still, the one-factor solution explained most of the variation. It had moderate factor loadings for the negatively worded items (0.661 and 0.678), high factor loadings for the other items (0.78 to 0.91), and strong internal consistency (Cronbach's alpha = 0.81 in wave one and 0.82 in wave two). Thus, we chose a single factor that was weighted more toward the positively worded items but used all six items as registered in our pre-analysis plan.

Table 5: Factor Loadings for the Desire to Enroll Scale

Name	Item	Factor Loadings Estimate (Std. Err)	
		Wave 1	Wave 2
Excited	I would be excited to enroll my child at Prim Middle School.	0.906 (0.007)	0.886 (0.008)
Fit	Prim Middle School would be a great fit for my child.	0.889 (0.009)	0.885 (0.008)
Choose	I would choose Prim Middle School over Walker Middle School.	0.780 (0.014)	0.826 (0.011)
Enroll	How likely would you be to enroll your child in Prim Middle School?	0.884 (0.008)	0.859 (0.009)
Disappoint	Enrolling my child in Prim Middle School would be disappointing.	0.647 (0.017)	0.661 (0.017)
Avoid	I would do everything realistically in my power to avoid Prim Middle School.	0.679 (0.017)	0.678 (0.017)
		N = 774	N = 901

For ease of interpretation, we also report results for the two items that represent discrete choices alongside our confirmatory factor score outcome. These were: "How likely would you be

to enroll your child in Prim Middle School?" and "I would choose Prim Middle School over Walker Middle School.", answered on a 4-point Likert scale from "Very Unlikely" to "Very Likely." We dichotomized both into 1 ("Very Likely" and "Somewhat Likely") and 0 ("Very Unlikely" and "Somewhat Unlikely"). We interpreted the former as a simpler, binary measure of whether the parent would enroll in Prim in general. We interpreted the latter as whether the parent would choose Prim over the anchor school, if those were the only two options, e.g., if moving, private school, or homeschool were not options.

Expectations of Racial Climate. We adapted the equal status subscale from the intergroup interactions domain of Byrd's (2017) measure of school racial climate. Byrd validated the original measure with 819 children ages 12-18 drawn from two independent national samples that were balanced on race/ethnicity to be 25% Asian, Black, Latine, and White. Researchers and school personnel reviewed the measure for face validity, and Byrd (2017) found high internal consistency for the equal status subscale (Cronbach's alpha = 0.86 in sample 1 and 0.87 in sample 2), as well as factor loadings of at least 0.8 across all items in both studies. We modified all three items in the equal status subscale to refer to the focal school in the profile. The resulting items were: "How likely is it that students of all races/ethnicities will be treated equally in Prim Middle School?", "How likely is it that the principal of Prim Middle School will treat students of all races/ethnicities fairly?", and "How likely is it that teachers at Prim Middle School will be fair to students of all races/ethnicities?" In our sample, we found high internal consistency for these three items, with Cronbach's alpha = 0.89 and inter-item correlations of 0.8 or higher.

Expectations of Belonging. Our belonging items were adapted from Hailey (2022a, 2022b). These adapted items were: "In general, how welcome do you think your student would feel in Prim Middle School?", "How likely is it that your student would feel socially and

emotionally supported in Prim Middle School?", "How likely is it that your student would academically succeed in Prim Middle School?", and "How likely is it that your student would make friends at Prim Middle School?" In our sample, the items showed moderately high internal consistency with Cronbach's alpha = 0.82 and inter-item correlations ranging from 0.55 to 0.78.

Computing Factor Scores for Racial Climate and Belonging. We conducted exploratory factor analyses using the Weighted Least Squares estimator and oblimin rotation across all the items from the equal status subscale for racial climate and for expectations of student belonging. The rule of thumb to retain factors left of the "elbow" in a scree plot of eigenvalues suggested a single factor, but disagreement about this subjective criteria warrants additional evidence for choosing how many factors to retain (UCLA: Statistical Consulting Group, n.d.). We thus compared the factor loadings for the one-factor and the two-factor solutions. In the one factor solution, all the equal status/racial climate items had larger factor loadings (0.86 to 0.9) than any of the belonging items. We interpret this as evidence that the equal status subscale of racial climate would dominate the one-factor solution. In the two-factor solution, the equal status items loaded strongly and almost exclusively onto the second factor with loadings ranging from 0.82 to 0.95. The items about the student making friends and succeeding loaded strongly and almost exclusively onto another factor with loadings of 0.86 and 0.93. The items about the school being welcoming and social-emotionally supportive clustered more strongly but not exclusively with the items about friendship and success.

We retained two factors—equal status and belonging. Both the one-factor and two-factor solution were justifiable based on the factor analyses. However, the two-factor solution cleanly mapped onto our constructs of interests whereas the one-factor solution would represent the concept of equal status somewhat muddled by items about belonging.

Estimation

Recall our first research question: how do test score gaps and suspension gaps between Black students and their peers affect Black parents' desire to enroll their children in an otherwise higher-achievement middle school? To answer this question, we estimate Equation 1 below using OLS regression, in each wave of data collection and for the pooled sample across waves:

$$(1)\ Y_i = \beta_0 + \beta_1 Test\ Score\ Gap_i + \beta_2 Suspension\ Gap_i + \beta_3 (Test\ Score\ Gap\ x\ Suspension\ Gap)_i + X_i + \alpha_i + \varepsilon_i$$

where, Y_i is a factor score for desire to enroll in (or alternatively the likelihood of choosing) Prim Middle School, *Suspension Gap* is a binary indicator for randomly assigned information about a suspension gap, *Test Score Gap* is a binary indicator for randomly assigned information about a suspension gap, and $Suspension\ Gap\ x\ Test\ Score\ Gap$ is a binary interaction term indicating that both gaps were presented. X represents a vector of covariates and α_i represents a fixed effect for the data collection wave. Our covariates are parent age, gender, education, and marital/cohabitation status; an indicator for K12 experience teaching; region; and locale (rural, town, suburban, city) based on zip code designations by the Census Bureau (Geverdt, 2019). For wave one, estimating Equation 1 yields intent-to-treat (ITT) effects, where most parents (~75%) clicked to view suspension rates. For wave two, there is no non-compliance since all parents were shown suspension rates.

Our second research question is whether test score gaps and suspension gaps affect Black parents' expectations of school racial climate and student belonging. To answer this question, we estimate Equation 1 again, except that Y_i becomes a factor score for the equal status subscale. Then we estimate Equation 1 once more where Y_i is a factor score for school belonging.

We estimate Equation 1 using effect coding, which provides equal statistical power for estimating main effects and interaction effects in a factorial experiment (Kugler et al., 2018). Effect coding represents the suspension gap only treatment as [1, -1], the test score gap only treatment as [-1, 1], both gaps treatment as [1,1], and the no gap condition as [-1,-1]. In a factorial experiment with a roughly equal number of observations in each condition, effect coding yields equal standard errors across factors (Duke Global Health Institute, 2020; Kugler et al., 2018). Effect coding also provides equal standard errors for interaction effects in observational studies to the extent that factor levels have approximately the same number of observations, such as 1:1 ratio of male to female (te Grotenhuis et al., 2017). In expectation, effect coding thus provides equal precision to estimate interactions effects for fathers vs. mothers and by child gender. Although the number of observations in each cell need not be exactly the same to use effect coding (Duke Global Health Institute, 2020; Kugler et al., 2018), we report effects using weighted effect coding (te Grotenhuis et al., 2017) in **Appendix G** to account for slight imbalance in the number of observations. Since our experiment is strongly balanced by condition and parent gender, our main results are almost identical.

Effect coding dictates the interpretation of the estimates from Equation 1. The intercept β_0 represents the grand mean, or the unweighted average of the outcome across conditions in the sample. Since the treatments in this study have two-levels, -1 and 1, the effect-coded regression coefficients represent half the change in the dependent variable when switching between the levels or, equivalently, the deviation from the grand mean (Brehm & Alday, 2022). The coefficient β_1 represents half the effect of a tests score gap, averaged across parents in our sample with and without suspension gaps. Likewise, β_2 represents half the effect of a suspension gap, regardless of whether there was a test score gap. In order to interpret both the main effects

as treatment effects instead of deviations from the grand mean, we multiply the coefficients by a scaling constant of 2; this does not affect hypothesis testing because it applies to the standard errors as well (Kugler et al., 2018). Finally, β_3 represents one fourth of the interaction effect, so we multiply by four.

Table 5 below illustrates how effect coding computes all main and interaction effects using the full sample by pooling together different experimental conditions (Kugler et al., 2018). The shaded rows represent the experimental conditions that the effect coding pools together to estimate each effect. For example, Panel A shows that the effect of the suspension gap on Black parents' school choice is estimated by taking the mean for the both-gaps and the suspension-gap-only conditions pooled together (the shaded rows) and then subtracting the mean for the test score gap only and no gap conditions pooled together (the unshaded rows). Note that this estimate does not compare the suspension gap treatment to the no gaps condition; effect coding does not compare individual treatment arms to a fixed control (Kugler et al., 2018). Effect coding is more efficient, splitting the entire sample into different halves to estimate each coefficient. Panel A also shows that the effect of the suspension gap is averaged across the two levels of the test score gap [-1, 1]. Therefore, the main effect is the effect of a suspension gap regardless of whether there is a test score gap (not the effect of a suspension gap vs. a no gap control). Analogously, effect coding gives the main effect of a test score gap regardless of whether there is a suspension gap.

Table 5: Computation of Main and Interaction Effects with Effect Coding

Panel A	Condition	**Suspension gap**	Test score gap	Suspension gap x test score gap	N
2	(Suspension gap only)	**1**	-1	-1	839
4	(Both gaps)	**1**	1	1	
1	(No gaps)	**-1**	-1	1	836

3	(Test score gap only)	**-1**	1	-1	

Main effect of **suspension gap** = (mean in conditions 2 and 4) - (mean in conditions 1 and 3)

Panel B	Condition	Suspension gap	**Test score gap**	Suspension gap x test score gap	N
3	(Test score gap only)	-1	**1**	-1	837
4	(Both gaps)	1	**1**	1	
1	(No gaps)	-1	**-1**	1	838
2	(Suspension gap only)	1	**-1**	-1	

Main effect of **test score gap** = (mean in conditions 3 and 4) - (mean in conditions 1 and 2)

Panel C	Condition	Suspension gap	Test score gap	**Suspension gap x test score gap**	N
1	(No gaps)	-1	-1	**1**	845
4	(Both gaps)	1	1	**1**	
2	(Suspension gap only)	1	-1	**-1**	830
3	(Test score gap only)	-1	1	**-1**	

Interaction effect of **suspension gap x test score gap** =
(mean in conditions 1 and 4) - (mean in conditions 2 and 3)

Note from Table 5 Panel C that the interaction of suspension gap x test score gap captures the difference between pooling the "both gaps" and "no gaps" conditions and pooling the test score only and suspension gap only conditions. Since both pooled comparisons average over the main effect of the suspension gap (+1 and -1) and the main effect of the test score gap (+1 and -1), the difference between the two pooled comparisons represents the interaction effect. Effect coding allows us to understand whether the effect of either suspension or test score gaps was different depending on the level of the other.

Following our preregistration plan, we conducted exploratory subgroup analyses by parent gender. That is, do the effects of suspension and test score gaps on school choice and expectations of racial climate and belonging differ for Black fathers vs. mothers? In addition, we

explored whether school choices and expectations differed by male vs. female children regardless of parent gender. To answer these questions, we estimate Equation 2 below. Note that we exclude the very small number of non-binary respondents and respondents with non-binary children from the subgroup analysis so that the variable $Male$ has only two levels.

$$(2)\ Y_i = \beta_0 + \beta_1 Test\ Score\ Gap_i + \beta_2 Suspension\ Gap_i + \beta_3(Test\ Score\ Gap\ x\ Suspension\ Gap)_i + \beta_4(Suspension\ Gap\ x\ Male)_i + \beta_5(Test\ Score\ Gap\ x\ Male)_i + \beta_6(Test\ Score\ Gap\ x\ Suspension\ Gap\ x\ Male)_i + X_i + \alpha_i + \varepsilon_i$$

First, we interpret Equation 2 to analyze heterogenous effects by parent gender. As before, Y_i represents the outcome, starting with the desire to enroll factor score. $Suspension\ Gap$ and $Test\ Score\ Gap$ also retain the same interpretation. $Male$ corresponds to whether the parent is a father [1] or a mother [-1]. Note that the main effect of the parent being $Male$ is captured by the vector of covariates X_i. The coefficient β_1 now estimates half the effect of a test score gap, regardless of whether there is a suspension gap and regardless of whether the parent is male or female. The coefficients β_2 estimates half the effect of a suspension gap, regardless of whether there is a test score gap and regardless of whether the parent is male or female. The interaction effect β_3 captures whether either of these effects differed based on the level of the other, regardless of parent gender. The interaction effects β_4, β_5, and β_6 estimate whether each of the previous main effects or the $Test\ Score\ Gap\ x\ Suspension\ Gap$ interaction effect was different for fathers vs. mothers.

We also explore whether the main and interaction effects differed for parents envisioning school choice for their boys vs. girls by changing one variable. We recode $Male$ to correspond to

whether the respondent's oldest child is male [1] or a female [-1]. Then we estimate Equation 2 with all the coefficients interpreted analogously.

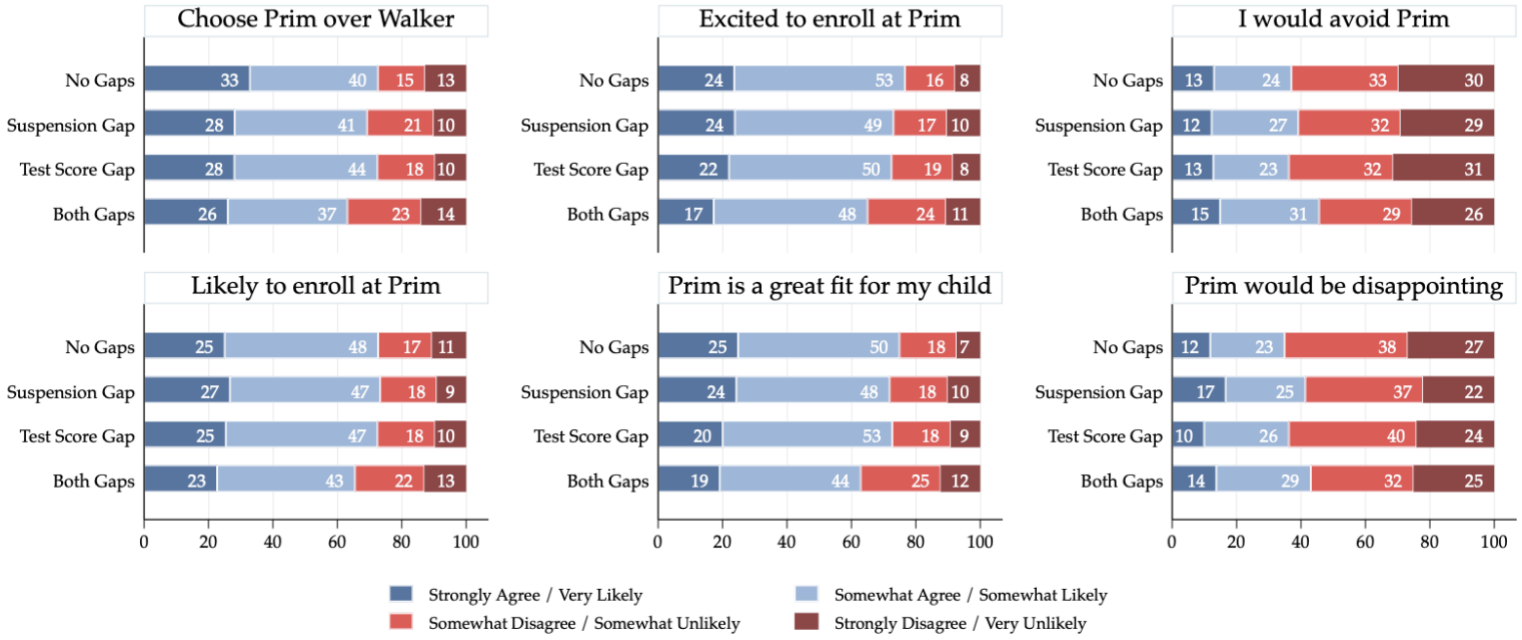
Results

Descriptive Findings

To build intuition for our main results, we plot the raw item distributions by experimental condition before testing our hypotheses in a regression framework. Figure 2 shows the distribution of desire to enroll items. Appendix B provides similar figures for the racial climate and belonging items. Recall that Prim Middle School has better overall student outcomes, so that our primary research question is whether racial disparities discourage Black parents from an otherwise higher-achievement school. As expected, most Black parents preferred Prim Middle School across all conditions. About 73% of the parents in the "No Gaps" condition reported that they would be "Somewhat likely" or "Very likely" to enroll their child at Prim and to choose Prim over the alternative. Even in the "Both Gaps" condition, 63% of parents said they would likely choose Prim over the alternative and 66% said they would likely enroll their child at Prim. That said, parents responded more positively overall to Prim in the "No Gaps" condition and more negatively in the "Both Gaps" condition. Respondents in the suspension gap only and test score gap only conditions fell in between the two other conditions, such that differences were less pronounced for each gap compared to the "No Gaps" condition. Effect coding to maximize power helps analyze these small differences. Factor scoring for our confirmatory outcome captures movement between "Strongly Agree" and "Somewhat Agree" and between "Strongly Disagree" and "Somewhat Disagree" and avoids losing precision by dichotomizing these outcomes.

Figure 2: Desire to Enroll Items by Experimental Condition

N = 1677 Black Parents



Effects of Test Score Gaps and Suspension Gaps

From left to right, Table 8 below reports the effect of suspension gaps and test score gaps on the desire to enroll factor score, the likelihood of enrolling, and the likelihood of choosing Prim over the alternative anchor school. The two columns on the far right remind readers that the effect coding computes each coefficient by comparing the mean of the two pooled experimental conditions that are shaded to the mean of the two pooled experimental conditions that are unshaded. Each effect-coded coefficient comes from approximately splitting the sample in half, so the standard errors remain almost the same for the coefficients within each column. Not shown in Table 8, we also examined whether test score gaps made Black parents more likely to view the suspension rates, which parents in wave one had to click to see.

Table 8 Column 1 shows that suspension gaps and test score gaps each significantly reduced Black parents' desire to enroll by a similar magnitude of about 0.11. Since we used effect coding, we multiply the coefficient of -0.0548 on the test score gap by two to represent the difference between the conditions with a test score gap vs. without a test score gap, regardless of the suspension gap. The standard deviation for the desire to enroll factor score is approximately 0.88 in wave one, so the effect of the test score gap translates to about 0.125 standard deviations. The effect of the suspension gap translates to about 0.129 standard deviations. The interaction effect equal four times the coefficient in Column 1 Row 3, or -0.08. This might suggest that the effect of both gaps on desire to enroll is slightly more negative than expected based on adding the two main effects of the test score and suspension gaps. However, the interaction coefficient was not statistically significant perhaps due to the standard error being roughly as large as the interaction term.

Table 8 Columns 2 and 3 show regression results for two binary outcomes: the likelihood of enrolling and the likelihood of choosing Prim over the alternative anchor school. The test score gap had a negative effect of about 4.4 percentage points (the coefficient of 0.0219 multiplied by two) on the likelihood of enrolling and a negative and a marginally significant negative effect on the likelihood of choosing Prim. The suspension gap reduced Black parents' likelihood of choosing Prim by a significant 6.5 percentage points (the coefficient of 0.0325 multiplied by two) relative to not having a suspension gap but had no effect on the likelihood of enrolling. Though not always significant, the coefficients for the likelihood of enrolling at Prim and choosing Prim over the alternative are consistent with the results for the preregistered desire to enroll factor score.

Table 8: Effects of Suspension and Test Score Gaps on Desire to Enroll and Choice

	Effect-coded Regression Results				
	(1)	(2)	(3)		
VARIABLES	Desire to enroll	P(Enroll)	P(Choose)	N compared	Conditions Compared
Test Score Gap	-0.0548**	-0.0219**	-0.0198*	837	(Test score gap only) (Both gaps)
	(0.0215)	(0.0111)	(0.0113)	838	(No gaps) (Suspension gap only)
Suspension Gap	-0.0569***	-0.0169	-0.0325***	839	(Suspension gap only) (Both gaps)
	(0.0215)	(0.0112)	(0.0113)	836	(No gaps) (Test score gap only)
Test Score Gap x Suspension Gap	-0.0204	-0.0172	-0.0128	845	(No gaps) (Both gaps)
	(0.0215)	(0.0111)	(0.0112)	830	(Suspension gap only) (Test score gap only)
Constant	-0.0766	0.768***	0.696***		
	(0.142)	(0.0712)	(0.0729)		
Observations	1,675	1,675	1,674		
R-squared	0.033	0.023	0.028		

Robust standard errors in parentheses

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

Notes: Results from OLS regressions estimating Equation 1 using effect coding in both waves of data collection with a fixed effect for wave. Column 1 shows effects on our factor score for desire to enroll, which is approximately normal. Columns 2 and 3 show effects from linear probability models on the likelihood of enrolling at Prim or choosing Prim over the alternative anchor school. The covariates were indicators for whether the parent was male, identified as multi-racial in addition to Black, had a BA degree or higher, was married, was cohabiting with a partner, and was currently or previously a K-12 teacher. We also included indicators for region, locale (suburban, town, rural, urban), and age group.

Next, we examine whether test score and suspension disparities affected expectations of belonging and racial climate. Recall that these items were only asked in wave two, which did not ask parents if they wanted to view suspensions and instead showed all respondents both test scores and suspension rates. All the coefficients are negative, but only the decline in expectations of belonging when exposed to a test score gap was marginally significant. Overall, we find null results for both the expectations of belonging factor and the equal status factor, though this could be driven by imprecision. We are likely underpowered for these exploratory analyses. Our pre-registration plan projected that we would need about 1200 respondents to detect an effect on our desire to enroll factor of 0.16 standard deviations prior to any gains in precision from covariates. Since we only asked about belonging and racial climate in wave two, we only have 901 respondents for these outcomes.

Table 9: Effects of Suspension and Test Score Gaps on Expectations of Belonging and Racial Climate

Effect-coded Regression Results

	(1)	(2)		
VARIABLES	Expectations of belonging	Racial climate: equal status subscale	N compared	Conditions Compared
Test Score Gap	-0.0610*	-0.0325	448	(Test score gap only)
	(0.0312)	(0.0310)		(Both gaps)
			453	(No gaps)
				(Suspension gap only)
Suspension Gap	-0.0414	-0.0378	444	(Suspension gap only)
	(0.0310)	(0.0308)		(Both gaps)
			457	(No gaps)
				(Test score gap only)
Test Score Gap x Suspension Gap	-0.0265	-0.00556	445	(No gaps)

	(0.0308)	(0.0307)		(Both gaps)
			456	(Suspension gap only)
				(Test score gap only)
Constant	-0.0878	-0.0572		
	(0.176)	(0.177)		
Observations	901	901		
R-squared	0.036	0.029		

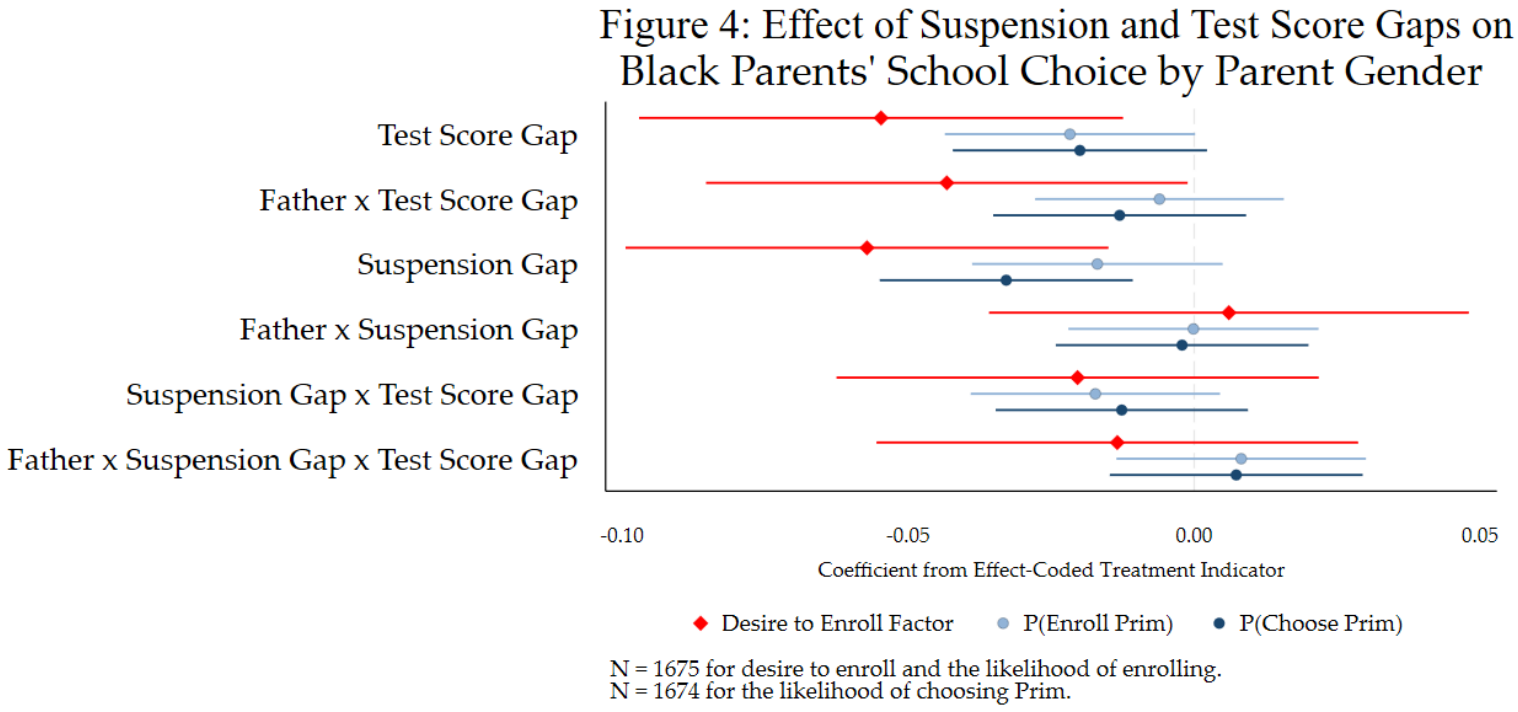
Robust standard errors in parentheses

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

Notes: Results from OLS regressions estimating Equation 1 using effect coding in wave two. Column 1 shows effects on our factor score for desire to enroll, which is approximately normal. Columns 2 and 3 show effects from linear probability models on the likelihood of enrolling at Prim or choosing Prim over the alternative anchor school. The covariates were indicators for whether the parent was male, identified as multi-racial in addition to Black, had a BA degree or higher, was married, was cohabiting with a partner, and was currently or previously a K-12 teacher. We also included indicators for region, locale (suburban, town, rural, urban), and age group.

Subgroup Analyses by Parent and Child Gender

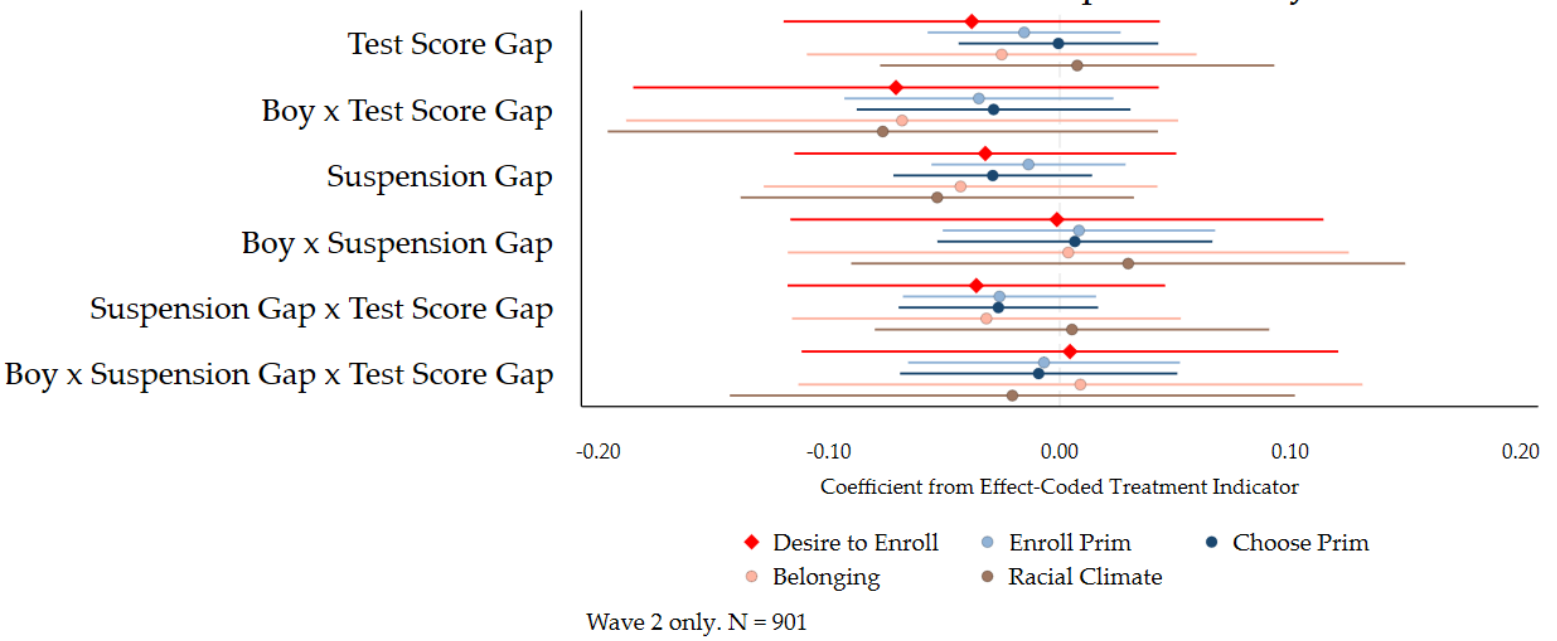
Figure 4 below shows the results of estimating Equation 2 to examine interaction effects by parent gender on school choice. The main effects for suspension gap and test score gap and



the two-way interaction for suspension gap x test score gap essentially replicate the coefficients shown previously in Table 8 above. We find mostly null interactions between racial gaps and being a father versus being a mother. The effects of racial disparities did not vary by parent gender, except that the test score gap reduced desire to enroll by about 0.19 standard deviations more for Black fathers than for Black mothers (the interaction effect was -0.043 x 4 = 0.1732 and the standard deviation for desire to enroll was approximately 0.91).

Next, we turn to subgroup analyses by child gender. We find null results. The effects of test score gaps on desire to enroll, belonging, and racial climate appear somewhat larger in magnitude for boys compared to girls, but none of the coefficients are statistically significant. Statistical power may be especially limited for this subgroup analysis given that only wave two collected data on child gender. With desire to enroll, for example, the standard errors on the coefficients for test score gap, suspension gap, and their two-way interaction are each approximately 0.042. These standard errors are nearly twice as large as the corresponding

Figure 5: Effect of Suspension and Test Score Gaps on Black Parents' School Choice and Expectations by Child Gender



Wave 2 only. N = 901

standard errors in Table 8. We encourage future research that examines how concerns about school climate and expected belonging may vary by child gender and inform school choice, perhaps with survey data from both parents and their children.

Discussion, Limitations, and Implications

We conducted the first survey experiment that randomized racial disparities in student outcomes to assess how they affect school perceptions among a large, national sample of Black parents. Our vignette experiment isolates the causal effect of racial disparities in a way not possible with observational data. We found that test score and suspension gaps each produced a roughly 0.12 standard deviation reduction in a factor score for desire to enroll. Despite these effects, Black parents still preferred the higher-achievement school over the anchor school when both gaps were present. Although parents weigh multiple factors when considering schools in a complex process that this study is not designed to replicate, racial disparities in test scores and suspension rates between Black and non-Black students can meaningfully influence Black parents' desire for otherwise higher-achievement schools, independent of other considerations.

Although racial disparities had a significant deterrent effect, more than two-thirds of Black parents in our sample still chose the higher-achievement school despite both a test score gap and a suspension gap. This implies that, all else equal, Black parents would prefer schools without racial gaps, but many would accept some racial disparities for higher overall achievement and lower overall suspension rates. Policies, including some test-based accountability provisions (Hemelt & Jacob, 2020), that shrink disparities by reducing the test scores of high-achieving students rather than improving outcomes on the lower end of the distribution are unlikely to appeal to Black families.

We acknowledge a key limitation that comparing school profiles may only represent a small, optional part of a longer, more complex and realistic choice process, which might involve feedback from social networks and school visits (Bell, 2009; Lareau et al., 2021). Concerns about racial disparities and school racial climate could be triggered through school visits or word of mouth, rather than through viewing school profiles. Moreover, such concerns likely interact with both parents' knowledge of their child and perceptions of the school's broader reputation. This study does not directly predict which schools Black parents will choose. However, it does suggest that disparities, on their own, can shape parent preferences. Schools might meaningfully shift Black parents' perceptions by reducing either test score or suspension gaps. Future research should examine how disparities operate in real-world school choice contexts.

An alternative interpretation of our results is that Black parents were focused on whether Prim had better outcomes for Black children than Walker, not on racial gaps. To manipulate gaps while holding the overall test scores and suspension rates at Prim constant across experimental conditions, we had to vary Black students' outcomes across conditions. This yielded different contrasts between Prim and Walker across conditions. For instance, in the test score gap condition, Black children at Prim and at Walker had the same proficiency rates. In the no gap control, Black kids had higher test scores at Prim than at Walker.

Despite these different contrasts, we prefer our interpretation that Black parents were responding at least partly to racial disparities rather than only the levels for Black students for two reasons. First, even in the both gaps condition, where Black students had the same test score proficiency rates at Prim and at Walker and Black students had slightly higher suspension rates at Prim as at Walker, Black parents still preferred Prim nearly two-to-one (63%). This implies that Black parents were not only focused on the best levels for Black students, but also on the better

overall student outcomes and perhaps the greater share of experienced teachers at Prim. If instead Black parents only cared about the levels for Black students, respondents in the both-gaps condition would be indifferent between Walker and Prim or even slightly prefer Walker. Second, Black parents were still deterred by the test score gap in the test-score-gap-only condition, when Black students at Prim and at Walker had the same proficiency rates, suggesting that disparities mattered. If instead respondents did not care about Black students' relative achievement, then the test score-gap-only condition should have no effect. These findings suggest that Black parents responded to both higher overall achievement and lower relative achievement for Black students. Third, we asked Black parents in wave two to mark all that apply from a list of reasons for their responses about the school profiles. About 22 percent reported that "gaps in student outcomes" played a role in their responses. This was smaller than the shares that said math test scores (49%), reading test scores (48%), and suspension rates (40%) played a role in their responses. However, the greater emphasis on *overall* levels with a real but lesser emphasis on gaps aligns with our finding that gaps were a significant deterrent but most Black parents preferred Prim even in the both gaps condition.

Our results yield similar practical implications even if some Black parents preferred whichever school had better outcomes for Black students, regardless of whether it had racial gaps. Suppose that schools need not close racial gaps to attract Black families; they only need better outcomes for Black students than the alternatives. However, closing racial gaps between Black and non-Black students would still represent the largest improvement in levels for Black students short of Black students doing better than the school average. Thus, from the perspective of how schools might attract Black families, we can interpret the effect of closing gaps as a constrained upper bound on how improving outcome levels for Black students might shift Black

parents' perceptions. If we replaced the "no gap" condition at Prim with a "smaller gap" condition, Black students would do better at Prim than at Walker by a smaller margin and we would expect effects on desire to enroll that are no larger than 0.125 standard deviations.

Data limitations make it hard to assess how well the choice between Walker and Prim mirrors the real options that Black families face. Our school profiles for Prim manipulate within-school gaps between Black students and overall achievement. Yet much of the existing research describes gaps at the district level (e.g., analyses based on district-level subgroup data from the Stanford Education Data Archive) and focuses on Black-White comparisons (Soland, 2021). With these caveats, we consider how our findings might generalize to choices of district and how future research might more systematically address generalizability.

If Walker Middle School in our experiment were a district, it would likely have higher test scores than real districts that have no Black-overall test score gap. The few real districts with near-zero Black-White test score gaps fall into two categories: large districts with high poverty rates like Detroit where both Black and White students both have very low performance and small districts with few Black students resulting in imprecise estimates of racial achievement gaps (Reardon et al., 2019). Reardon et al. (2019) found not one U.S. school district with moderate Black enrollment, a near-zero Black-White achievement gap, and moderately high overall academic achievement. If our Walker profile instead showed all students doing much more poorly, Black parents in our experiment would likely have preferred Prim by an even wider margin. All else constant, the effects of the test score and suspension gaps may have been smaller because inequality at Prim would have been preferable to very low achievement for all students. If real racial gaps are larger than in our treatment conditions, then the deterrent effects for Black parents could be larger. Estimating an exchange rate between disparities and overall achievement

is beyond the scope of this paper. Future research with a nationally-representative sample could explore this question, perhaps using respondents' local data to calibrate the school profiles.

Our exploratory analyses of racial climate, belonging, and child gender relied on our second wave of data collection, a smaller sample with more imbalance across conditions. These limitations make our exploration of mechanisms and gender more susceptible to low power and potential bias. We found little evidence that either racial disparities in suspensions or test scores reduced Black parents' expectations of racial climate and belonging. Our exploratory subgroup analyses by parent and child gender also yielded mostly null results.

Nevertheless, our exploratory analyses raise two potential areas for future research. First, despite high expectations of fairness, most parents in our sample believed that their child would have felt at least somewhat unwelcome at our focal school. Future research might examine the conditions that make Black families and their children feel welcome. Second, we found some evidence that racial test score gaps especially deterred Black fathers from the higher-achievement school. We refrain from interpreting this exploratory analysis with nine coefficients as clear evidence that Black fathers worried more about test score disparities than Black mothers. However, it does raise a potential counternarrative to the stereotype of uninvolved Black fathers that future research could probe, perhaps in nationally representative data.

While addressing structural barriers such as transportation and segregation, policymakers seeking to advance equity through school choice should also focus on making schools more attractive to Black families by targeting the disparities within them. We encourage further inquiries into how reducing disparities in test scores, suspensions, or both may attract or retain Black families.

Appendix A: Walker Middle School Profile Held Constant for all Respondents

Walker Middle School

General Information

Category: Middle School
Principal: John Thomas
Accreditation: Accredited
Distance to School:
15-minute walk from your address
Student-Teacher Ratio: 12 to 1
Teachers with Three or More Years of Experience: 68%

Academic Achievement

English Pass Rate: 65%
Math Pass Rate: 66%

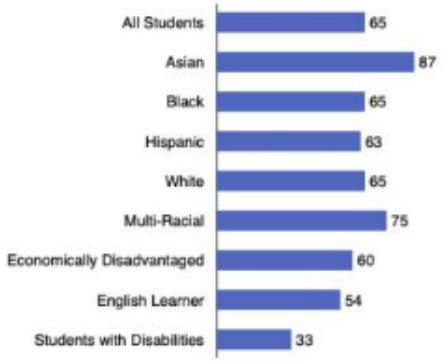
Enrollment

Asian: 5%
Black: 42%
Hispanic: 20%
White: 29%
Multi-Racial: 4%
Economically Disadvantaged: 32%
Students with Disabilities: 13%
English Learners: 16%

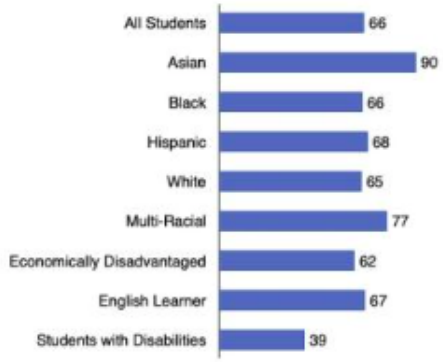
Student Engagement:

Chronic Absenteeism: 8%

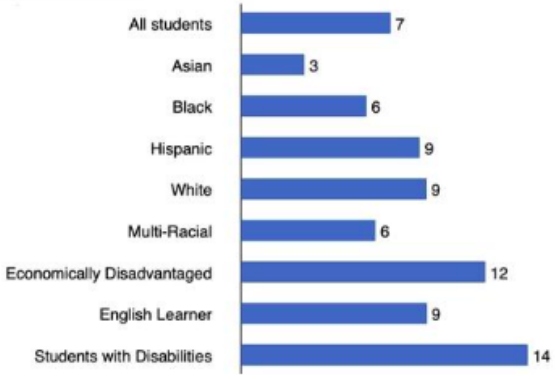
English: Percent of Students at Grade Level



Math: Percent of Students at Grade Level



Suspension Rates



Appendix B: Items

Desire to enroll scale

Variable Name	Question
Excited	I would be excited to enroll my child at Prim Middle School.
Fit	Prim Middle School would be a great fit for my child.
Choose	I would choose Prim Middle School over Walker Middle School.
Enroll	How likely would you be to enroll your child in Prim Middle School?
Disappoint	Enrolling my child in Prim Middle School would be disappointing.
Avoid	I would do everything realistically in my power to avoid Prim Middle School.

Belonging items:

Variable Name	Survey Question
Supported	How likely is it that your student would feel socially and emotionally supported in Prim Middle School?
Friends	How likely is it that your student would make friends in Prim Middle School?
Welcome	In general, how welcome do you think your student would feel in Prim Middle School?
Succeed	How likely is it that your student would academically succeed in Prim Middle School?

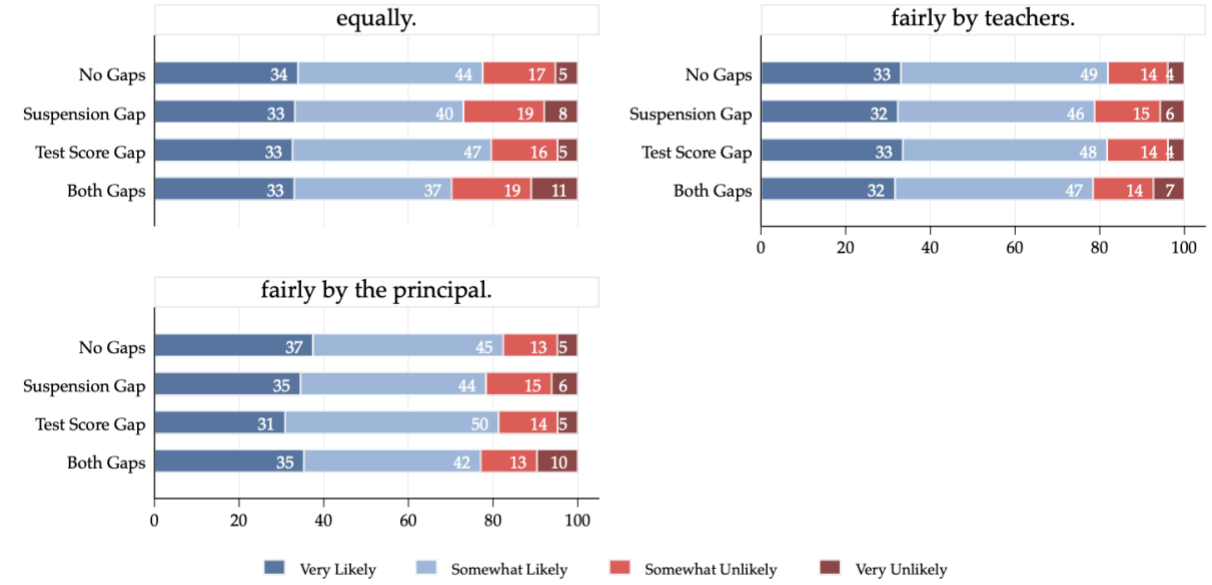
Source: Chantal Hailey, School Choice and COVID Survey Experiment, 2021

Variations of Byrd (2017) Validated Racial Climate Subscales

1. How likely is it that students of all races/ethnicities will be treated equally in Prim Middle School?
2. How likely is it that the principal of Prim Middle School will treat students of all races/ethnicities fairly?
3. How likely is it that teachers at Prim Middle School will be fair to students of all races/ethnicities?

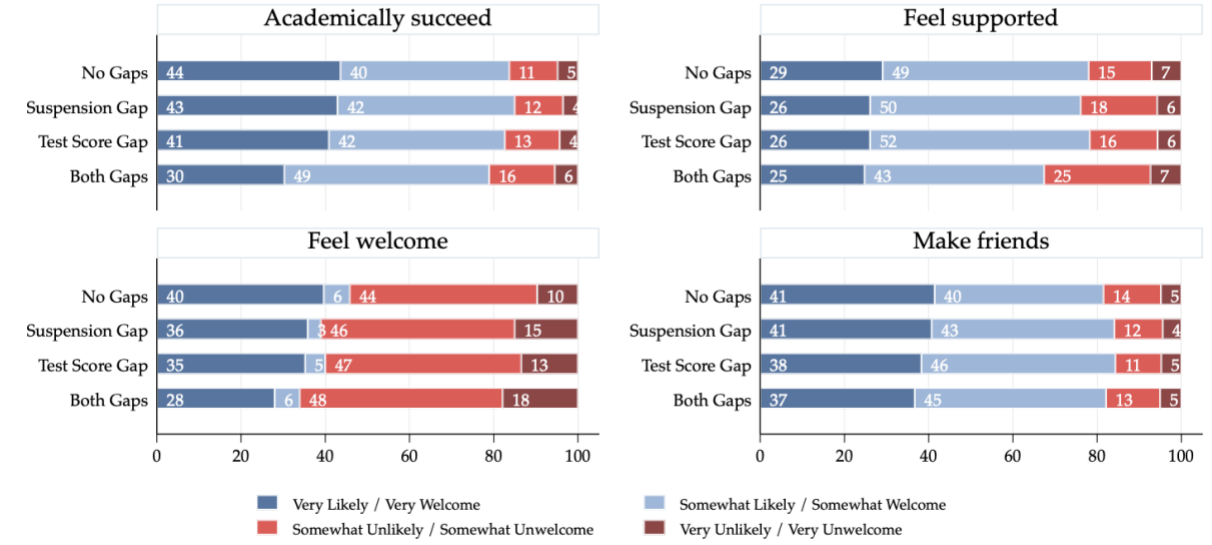
Figures 2 and 3 below show the response distribution for the racial climate items and belonging items by experimental condition. Recall that these measures were only introduced in wave two.

Figure 2: How likely is it that students of all races/ethnicities will be treated...



Note: Wave 2 data only. N = 901 Black Parents

Figure 3: How likely is it that your student would... at Prim?
(and how welcome would they feel at Prim?)



Note: Wave 2 data only. N = 901 Black Parents

Appendix C: Typo in Wave One that was Corrected in Wave Two

Prim Middle School

General Information

Category: Middle School
Principal: Carrie Walsh
Accreditation: Accredited
Distance to School:
15-minute walk from your address
Student-Teacher Ratio: 12 to 1
Teachers with Three or More Years of Experience: 72%

Academic Achievement

English Pass Rate: 75%
Math Pass Rate: 75%

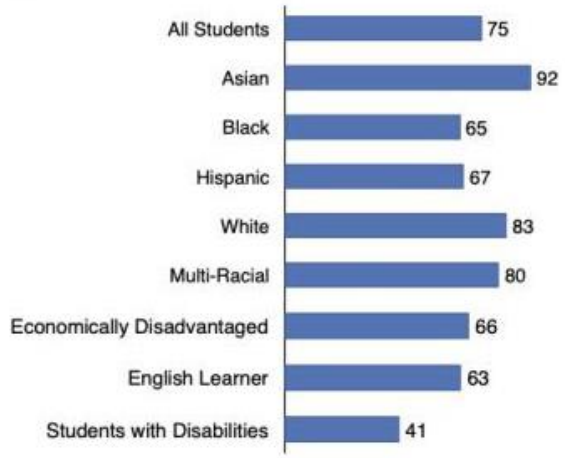
Enrollment

Asian: 9%
Black: 35%
Hispanic: 17%
White: 34%
Multi-Racial: 5%
Economically Disadvantaged: 26%
Students with Disabilities: 12%
English Learners: 13%

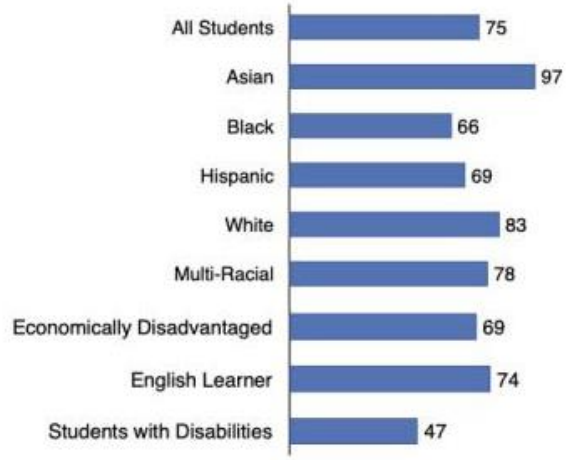
Student Engagement:

Chronic Absenteeism: 6%

English: Percent of Students at Grade Le 15-



Math: Percent of Students at Grade Level



Appendix D

Table 1. Standardized Differences in Baseline Characteristics Between Treatment Conditions and Control

	Standardized Mean Difference by Treatment			
	Test Score Gap Only	Suspension Gap Only	Both Gaps	Obs
Male	0.017	0.006	0.021	1675
Black Multiracial/ethnic	0.026	-0.008	-0.037	1677
Married	0.001	0.023	0.032	1677
Cohabiting	0.009	-0.013	-0.003	1677
Less than High School	-0.047	-0.002	-0.031	1677
High School	0.007	-0.026	-0.002	1677
Associate's Degree	0.010	0.019	0.010	1677
Some College	0.008	0.010	-0.009	1677
BA degree or higher	-0.004	0.004	0.030	1677
Previously taught K-12	-0.010	-0.048	0.017	1677
Teaches K-12 now	-0.030	0.002	-0.016	1677
Age Under 24	-0.031	-0.057	-0.077	1677
Age 25-34	0.021	0.048	0.030	1677
Age 35-44	-0.041	0.010	-0.004	1677
Age 45-54	0.061	-0.034	0.006	1677
Age 55+	-0.017	0.006	0.032	1677
Suburban	-0.049	0.050	0.054	1677
Rural	0.055	0.016	-0.004	1677
Town	0.038	-0.000	0.057	1677
Urban	-0.011	-0.058	-0.050	1677
Northeast	-0.028	-0.005	0.024	1677
Midwest	-0.004	0.006	-0.041	1677
South	0.020	-0.022	-0.015	1677
West	0.013	0.016	0.027	1677
Kids under 18 at Home	-0.051	-0.028	-0.034	1677

Appendix E: Regression Results by Wave

In wave one, our school choice vignette showed all parents the test score proficiency rates by subgroup and offered the option to view suspension rates. Overall, about 75% of parents in wave one chose to view the suspension rates, which were also disaggregated by race/ethnicity, economic disadvantaged, and special education status. The test score gap did not significantly affect whether parents chose to view suspensions.

Table A2. Effects of Suspension Gaps and Test Score Gaps in Wave One

	Effect-coded Regression Results				
	(1)	(2)	(3)		
VARIABLES	Desire to enroll	P(Enroll)	P(Choose)		
				N compared	Conditions compared
Test Score Gap	-0.0365	-0.0112	-0.0297*	389	(Test score gap only)
	(0.0318)	(0.0169)	(0.0171)		(Both gaps)
				384	(No gaps)
					(Suspension gap only)
Suspension Gap	-0.0887***	-0.0267	-0.0398**	395	(Suspension gap only)
	(0.0316)	(0.0169)	(0.0171)		(Both gaps)
				378	(No gaps)
					(Test score gap only)
Test Score Gap x Suspension Gap	-0.0130	-0.00503	0.00505	399	No gaps
	(0.0320)	(0.0168)	(0.0170)		Both gaps
				374	Suspension gap only
					Test score gap only
Constant	0.0428	0.684***	0.793***		
	(0.265)	(0.136)	(0.133)		
Observations	774	774	773		
R-squared	0.044	0.025	0.035		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A3. Effect of Test Score Gaps on the Likelihood of Viewing School Suspension Rates in Wave One

Panel A: Effect-coded Regression Results

VARIABLES	P(Viewed Suspensions)	N compared	Conditions Compared
Test Score Gap	0.0110	389	(Test score gap only)
	(0.0155)		(Both gaps)
		384	(No gaps)
			(Suspension gap only)
Constant	0.620***		
	(0.131)		
Observations	774		
R-squared	0.026		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Panel B: Dummy-coded Regression Results

Experimental Condition	P(Viewed Suspensions)
Suspension Gap	-0.0264
	(0.0442)
Test Score Gap	-0.0453
	(0.0460)
Both Gaps	0.0570
	(0.0411)
Constant	0.632***
	(0.134)
Observations	774
R-squared	0.034

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results for Wave Two

Table A4. Effects of Suspension and Test Score Gaps in Wave Two

Effect-coded Regression Results

VARIABLES	(1) Desire to enroll	(2) P(Enroll)	(3) P(Choose)	(4) Belonging	(5) Racial Climate: Equal Status	N compared	Conditions Compared
Suspension Gap	-0.0332	-0.00926	-0.0257*	-0.0414	-0.0378	444	(Suspension gap only)
	(0.0296)	(0.0151)	(0.0152)	(0.0310)	(0.0308)		(Both gaps)
						457	(No gaps)
							(Test score gap only)
Test Score Gap	-0.0754**	-0.0338**	-0.0156	-0.0610*	-0.0325	448	(Test score gap only)
	(0.0297)	(0.0151)	(0.0153)	(0.0312)	(0.0310)		(Both gaps)
						453	(No gaps)
							(Suspension gap only)
Suspension gap x test score gap	-0.0331	-0.0295**	-0.0313**	-0.0265	-0.00556	445	(No gaps)
	(0.0295)	(0.0150)	(0.0151)	(0.0308)	(0.0307)		(Both gaps)
						456	(Suspension gap only)
							(Test score gap only)
Constant	-0.165	0.810***	0.634***	-0.0878	-0.0572		
	(0.165)	(0.0853)	(0.0879)	(0.176)	(0.177)		
Observations	901	901	901	901	901		
R-squared	0.050	0.040	0.046	0.036	0.029		

Robust standard errors in parentheses

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

Appendix F: Representativeness of our Sample Compared to Black Parents in the U.S.

Our respondents resembled the U.S. population of Black parents or adults in marital status, associates degree and "some college" completion, and urbanness but differed in other aspects. Table E1 compares demographics for the sample pooled across waves to population estimates for the most similar categories available from the U.S. Census Bureau's American Community Survey (ACS) microdata for 2022 and the National Center for Education Statistic's (NCES) Public Elementary/Secondary School Universe Survey for 2019-2020 (Moslimani et al., 2024; National Center for Education Statistics, 2021a, 2023). Our sample had a slightly lower share of multiracial respondents than the share for Black adults nationwide in 2022 (8.4% vs. 11.3%). The marriage rate in our sample came quite close to the marriage rate for parents of Black children nationally (33.2% vs. 35.5%). Our sample's education levels resembled ACS estimates but had fewer BA completers (22.7% vs 32.4%) and more who stopped at high school (29.9% vs. 23.6%). The population benchmark might be mechanically higher, however, because the ACS estimate considered the highest degree of *any related adult* in a Black child's household. The share of respondents from urban areas mirrored the share of Black public-school students attending urban schools (47% vs. 45.4%). However, our sample was less suburban (25% vs. 35.8%), more rural (27.7% vs. 11.9%), and much less likely to reside in a town (0.3% vs. 6.9%) than the nation's Black public K-12 enrollment. Finally, 38.2% of our sample resided in the South compared to more than half the adult Black population.

Table A5. Sample Comparison to National Benchmarks

	Full Sample	National Benchmark Estimate	Population for the National Benchmark	Source
Race				
Black only	91.6%	88.7%	Black adults	Pew Research Center tabulations of the ACS Public Use Microdata Sample, 2022. (Moslimani et al. 2021)
Black multiracial	8.4%	11.3%		
Marital status				
Married	33.2%	35.5%	Parents or householders of Black children under 18	NCES tabulations of (ACS), 1-Year Public Use Microdata Sample, 2022. Digest of Education Statistics 2023, table 102.20. (NCES 2023)
Highest level of education				
BA degree or higher	22.7%	32.4%	Any related adult in the household of Black children under 18	NCES tabulations of (ACS), 1-Year Public Use Microdata Sample, 2022. Digest of Education Statistics 2023, table 104.70. (NCES 2023)
Associates degree	11.1%	12.2%		
Some college	25.9%	26.3%		
High school	29.9%	23.6%		
Less than high school	5.5%	5.5%		
Locale				
Rural	27.7%	11.9%	Black public school students	US Department of Education. Common Core of Data. "Public Elementary/Secondary School Universe Survey," 2019-20. (NCES 2021)
Suburban	25.0%	35.8%		
Town	0.3%	6.9%		
Urban	47.0%	45.4%		
Region				
Northeast	20.4%	16.6%	Black adults	Pew Research Center tabulations of the ACS Public Use Microdata Sample, 2022. (Moslimani et al. 2021)
Midwest	25.4%	17.2%		
South	38.2%	55.8%		
West	15.0%	10.3%		
N	1677			

Although our sample only somewhat represents Black parents nationwide, researchers have found comparable treatment effects across paired survey experiments using online samples and nationally representative samples. Researchers have replicated experiments in psychology, economics, and political science using online samples (Berinsky et al., 2012; Thomas & Clifford, 2017). A series of 15 replications of survey experiments based on priming, information, and framing treatments found a strong correlation (0.85) between treatment effects in nationally representative samples and the same survey experiment conducted using Amazon's Mturk platform. The authors believed that these information experiments had fairly homogenous effects because the background characteristics of the samples differed and the evidence of within-experiment treatment effect heterogeneity was limited (Coppock, 2019). Prior research suggests that data quality from online panels is comparable to that of Mturk (Kennedy et al., 2020; Peer et al., 2022; Smith et al., 2016), which has been used for survey experiments in education (Houston & Henig, 2021). We decided on Centiment, however, because having a pre-screened panel helped ensure that respondents were Black.

Appendix G: Results using Weighted Effect Coding

Table A6. Weighted-effect-coded Regressions Results

VARIABLES	(1) Desire to enroll	(2) P(Enroll)	(3) P(Choose)	(4) Belonging	(5) Racial Climate: Equal Status	N compared	Conditions Compared
Test Score Gap	-0.0549**	-0.0219**	-0.0199*	-0.0611*	-0.0325	837	(Test score gap only)
	(0.0216)	(0.0111)	(0.0113)	(0.0312)	(0.0310)		(Both gaps)
						838	(No gaps)
							(Suspension gap only)
Suspension Gap	-0.0568***	-0.0169	-0.0325***	-0.0413	-0.0377	839	(Suspension gap only)
	(0.0215)	(0.0112)	(0.0113)	(0.0310)	(0.0308)		(Both gaps)
						836	(No gaps)
							(Test score gap only)
Test Score Gap x Suspension Gap	-0.0202	-0.0171	-0.0126	-0.0263	-0.00551	845	(No gaps)
	(0.0213)	(0.0110)	(0.0111)	(0.0305)	(0.0304)		(Both gaps)
						830	(Suspension gap only)
							(Test score gap only)
Constant	-0.0768	0.768***	0.696***	-0.0879	-0.0572		
	(0.142)	(0.0712)	(0.0729)	(0.176)	(0.177)		
Observations	1,675	1,675	1,674	901	901		
R-squared	0.033	0.023	0.028	0.036	0.029		

Robust standard errors in parentheses

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

Appendix H: Effects on Binary Outcomes when "Somewhat Likely" is Recoded as not Choosing or Enrolling in Prim

Table A7. Effects of Suspension Gaps and Test Score Gaps when Recoding "Somewhat Likely" as 0 for Choice and Enrollment

Effect-coded Regression Results

VARIABLES	(1) Very likely to Enroll	(2) Very Likely to Choose Prim	N compared	Conditions Compared
Test Score Gap	-0.00958	-0.0176	837	(Test score gap only)
	(0.0106)	(0.0110)		(Both gaps)
			838	(No gaps)
				(Suspension gap only)
Suspension Gap	-0.00182	-0.0184*	839	(Suspension gap only)
	(0.0106)	(0.0111)		(Both gaps)
			836	(No gaps)
				(Test score gap only)
Test Score Gap x Suspension Gap	-0.00945	0.00957	845	(No gaps)
	(0.0106)	(0.0110)		(Both gaps)
			830	(Suspension gap only)
				(Test score gap only)
Constant	0.200***	0.214***		
	(0.0649)	(0.0682)		
Observations	1,675	1,674		
R-squared	0.030	0.040		

Robust standard errors in parentheses

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

References

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., & Walters, C. R. (2020). Do Parents Value School Effectiveness? *The American Economic Review*, *110*(5), 1502–1539.

Adukia, A., Feigenberg, B., & Fatemeh, M. (2024). *From Retributive to Restorative: An Alternative Approach to Justice in Schools*. https://www.dropbox.com/scl/fi/9rk2481uoi3cbfwli08kt/RP_Paper_July2024.pdf?dl=0&e=1&rlkey=c849wv64vs6mkq32gnxzy36yq

Alesina, A., Ferroni, M. F., & Stantcheva, S. (2024). *Perceptions of Racial Gaps, their Causes, and Ways to Reduce Them*. https://scholar.harvard.edu/files/stantcheva/files/alesina_ferroni_stantcheva_perceptions_racial_gaps.pdf

Alesina, A., Ferroni, M., & Stantcheva, S. (n.d.). *Online Appendix for "Perceptions of Racial Gaps, their Causes, and Ways to Reduce Them."*

Allen, Q. (2013). "They Think Minority Means Lesser Than": Black Middle-Class Sons and Fathers Resisting Microaggressions in the School. *Urban Education*, *48*(2), 171–197. https://doi.org/10.1177/0042085912450575

Angrist, J. D., Pathak, P. A., & Walters, C. R. (2013). Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics*, *5*(4), 1–27. https://doi.org/10.1257/app.5.4.1

Augustine, C., Engberg, J., Grimm, G., Lee, E., Wang, E., Christianson, K., & Joseph, A. (2018). *Can Restorative Practices Improve School Climate and Curb Suspensions? An Evaluation of the Impact of Restorative Practices in a Mid-Sized Urban School District*. RAND Corporation. https://doi.org/10.7249/RR2840

Bañales, J., Marchand, A. D., Skinner, O. D., Anyiwo, N., Rowley, S. J., & Kurtz-Costes, B. (2020). Black Adolescents' Critical Reflection Development: Parents' Racial Socialization and Attributions About Race Achievement Gaps. *Journal of Research on Adolescence*, *30*(S2), 403–417. https://doi.org/10.1111/jora.12485

Bell, C. A. (2009). All Choices Created Equal? The Role of Choice Sets in the Selection of Schools. *Peabody Journal of Education*, *84*(2), Article 2. https://doi.org/10.1080/01619560902810146

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Billingham, C. M., & Hunt, M. O. (2016). School Racial Composition and Parental Choice: New Evidence on the Preferences of White Parents in the United States. *Sociology of Education*, *89*(2), 99–117. https://doi.org/10.1177/0038040716635718

Billingham, C. M., Kimelberg, S. M., Faude, S., & Hunt, M. O. (2020). In Search of a Safe School: Racialized Perceptions of Security and the School Choice Process. *The Sociological Quarterly*, *61*(3), 474–499. https://doi.org/10.1080/00380253.2019.1711257

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*, 104334. https://doi.org/10.1016/j.jml.2022.104334

Butler, A., & Quarles, B. (2024). "It's a Chance, Not a Choice": Black Families, School Choice, and Gentrification in Washington, D.C. *Teachers College Record*, 01614681241242878. https://doi.org/10.1177/01614681241242878

Byrd, C. M. (2017). The complexity of school racial climate: Reliability and validity of a new measure for secondary students. *British Journal of Educational Psychology*, *87*(4), 700–721. https://doi.org/10.1111/bjep.12179

Chin, M. J. (2022). The Stability of Families' Revealed Preferences for Schools. *Educational Researcher*, *51*(9), 588–592. https://doi.org/10.3102/0013189X221125778

Chubb, J. E., & Moe, T. M. (1990). America's Public Schools: Choice Is a Panacea. *The Brookings Review*, *8*(3), 4–12. https://doi.org/10.2307/20080159

Cobanoglu, C., Cavusoglu, M., & Turktarhan, G. (2021). A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of Amazon Mechanical Turk (MTurk). *Journal of Global Business Insights*, *6*(1), 92–97. https://doi.org/10.5038/2640-6489.6.1.1177

Cooper, C. W. (2005). School Choice and the Standpoint of African American Mothers: Considering the Power of Positionality. *The Journal of Negro Education*, *74*(2), 174–189.

Coppock, A. (2019). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. *Political Science Research and Methods*, *7*(3), 613–628. https://doi.org/10.1017/psrm.2018.10

Cucchiara, M. (2013). "Are We Doing Damage?" Choosing an Urban Public School in an Era of Parental Anxiety. *Anthropology & Education Quarterly*, *44*(1), 75–93. https://doi.org/10.1111/aeq.12004

Cuddy, M., Krysan, M., & Lewis, A. (2020). Choosing homes without choosing schools? How urban parents navigate decisions about neighborhoods and school choice. *Journal of Urban Affairs*. http://www.tandfonline.com/doi/abs/10.1080/07352166.2020.1739537

Duke Global Health Institute. (2020). *Core Guide Dummy and Effect Coding in the Analysis of Factorial Designs* (Research Design & Analysis Core). https://sites.globalhealth.duke.edu/rdac/wp-content/uploads/sites/27/2020/08/Core-Guide_Dummy-and-Effect-Coding_16-03-20.pdf

Eisenlohr, A., Kennedy, K., Bulkley, K. E., & Marsh, J. A. (2023). A Tale of Two Systems: Choice and Equity in the District of Columbia's Charter Schools. *Educational Policy*, 08959048231198816. https://doi.org/10.1177/08959048231198816

Fong, K., & Faude, S. (2018). Timing Is Everything: Late Registration and Stratified Access to School Choice. *Sociology of Education*, *91*(3), 242–262. https://doi.org/10.1177/0038040718785201

Ford, J. B. (2017). Amazon's Mechanical Turk: A Comment. *Journal of Advertising*, *46*(1), 156–158. https://doi.org/10.1080/00913367.2016.1277380

Friedman, M., & Friedman, R. D. (1980). *Free to Choose: A Personal Statement* (Darden items stored in Ivy, By Request). Harcourt Brace Jovanovich.

Fryer, R. G., & Howard-Noveck, M. (2020). High-Dosage Tutoring and Reading Achievement: Evidence from New York City. *Journal of Labor Economics*, *38*(2), 421–452. https://doi.org/10.1086/705882

Geverdt, D. (2019). *Locale Boundaries File Documentation* (115; p. 26). National Center for Education Statistics. https://nces.ed.gov/programs/edge/docs/EDGE_NCES_LOCALE.pdf

Gilliam, W. S., & Reyes, C. R. (2018). Teacher Decision Factors That Lead to Preschool Expulsion: Scale Development and Preliminary Validation of the Preschool Expulsion Risk Measure. *Infants & Young Children*, *31*(2), 93–108. https://doi.org/10.1097/IYC.0000000000000113

Glazerman, S., & Dotter, D. (2017). Market Signals: Evidence on the Determinants and Consequences of School Choice From a Citywide Lottery. *Educational Evaluation and Policy Analysis*, *39*(4), 593–619. https://doi.org/10.3102/0162373717702964

Golann, J. W., Debs, M., & Weiss, A. L. (2019). "To Be Strict on Your Own": Black and Latinx Parents Evaluate Discipline in Urban Choice Schools. *American Educational Research Journal*, *56*(5), 1896–1929. https://doi.org/10.3102/0002831219831972

Goldring, E. B., & Hausman, C. S. (1999). Reasons for parental choice of urban schools. *Journal of Education Policy*, *14*(5), 469–490. https://doi.org/10.1080/026809399286161

Haderlein, S. A. K. (2022). How Do Parents Evaluate and Select Schools? Evidence From a Survey Experiment. *American Educational Research Journal*, *59*(2), 381–414. https://doi.org/10.3102/00028312211046360

Hailey, C. A. (2022a). Racial Preferences for Schools: Evidence from an Experiment with White, Black, Latinx, and Asian Parents and Students. *Sociology of Education*, *95*(2), 110–132. https://doi.org/10.1177/00380407211065179

Hailey, C. A. (2022b). Racialized Perceptions of Anticipated School Belonging. *Educational Policy*, *36*(4), 879–910. https://doi.org/10.1177/08959048221087211

Hanson, R., Pugliese, C., & Grady, S. (2020). *Parent and Family Involvement in Education: 2019* (NCES 2020-076; p. 42). U.S. Department of Education. https://nces.ed.gov/pubs2020/2020076full.pdf

Hastings, J., Kane, T., & Staiger, D. (2009). *Heterogeneous Preferences and the Efficacy of Public School Choice*. NBER Working Paper. https://justinehastings.com/wp-content/uploads/2020/01/2010-06_NBER-Working-Paper-No12145-and-11805_Heterogeneous-Preferences-and-Efficiency-of-Public-School-Choice.pdf

Hemelt, S. W., & Jacob, B. A. (2020). How Does an Accountability Program that Targets Achievement Gaps Affect Student Performance? *Education Finance and Policy*, *15*(1), 45–74. https://doi.org/10.1162/edfp_a_00276

Herelle, T. I. (2022). To protect and to prepare: Black mothers' school-selection decision-making. *Race Ethnicity and Education*, *0*(0), 1–19. https://doi.org/10.1080/13613324.2022.2106474

Holmes Erickson, H. (2017). How do parents choose schools, and what schools do they choose? A literature review of private school choice programs in the United States. *Journal of School Choice*, *11*(4), 491–506. https://doi.org/10.1080/15582159.2017.1395618

Houston, D. M., & Henig, J. R. (2021). The Effects of Student Growth Data on School District Choice: Evidence from a Survey Experiment. *American Journal of Education*, *127*, 33.

Houston, D. M., & Henig, J. R. (2022). *The 'Good' Schools: Academic Performance Data, School Choice, and Segregation*. https://doi.org/10.26300/7N5S-KM15

Houston, D. M., & Henig, J. R. (2023). The "Good" Schools: Academic Performance Data, School Choice, and Segregation. *AERA Open*, *9*, 23328584231177666. https://doi.org/10.1177/23328584231177666

Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J., & Asare-Marfo, D. (2020, February 18). Assessing the Risks to Online Polls From Bogus Respondents. *Pew Research Center Methods*. https://www.pewresearch.org/methods/2020/02/18/assessing-the-risks-to-online-polls-from-bogus-respondents/

Kerwin, J., Rostom, N., & Sterck, O. (2024). Striking the Right Balance: Why Standard Balance Tests Over-Reject the Null, and How to Fix it. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4926535

Kimelberg, S. M. (2014). Chapter 7. Middle-Class Parents, Risk, and Urban Public Schools. In A. Lareau & K. Goyette (Eds.), *Choosing Homes, Choosing Schools: Residential Segregation and the Search for a Good School.* Russell Sage Foundation. https://muse-jhu-edu.proxy1.library.virginia.edu/book/30872

Kraft, M. A. (2015). How to Make Additional Time Matter: Integrating Individualized Tutorials into an Extended Day. *Education Finance and Policy*, *10*(1), 81–116. https://doi.org/10.1162/EDFP_a_00152

Kugler, K. C., Dziak, J. J., & Trail, J. (2018). Coding and Interpretation of Effects in Analysis of Data from a Factorial Experiment. In L. M. Collins & K. C. Kugler (Eds.), *Optimization of Behavioral,*

Biobehavioral, and Biomedical Interventions: Advanced Topics (pp. 175–205). Springer International Publishing. https://doi.org/10.1007/978-3-319-91776-4_6

Lareau, A., Weininger, E. B., & Warner-Griffin, C. (2021). Structural constraints and the school choice strategies of black American middle-class parents. *British Journal of Sociology of Education*, *42*(4), 490–505. https://doi.org/10.1080/01425692.2021.1872367

Lewis-McCoy, R. L. (2016). Boyz in the 'Burbs: Parental Negotiation of Race and Class in Raising Black Males in Suburbia. *Peabody Journal of Education*, *91*(3), 309–325. https://doi.org/10.1080/0161956X.2016.1182839

Liu, J., Hayes, M. S., & Gershenson, S. (2022). JUE insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline. *Journal of Urban Economics*, 103453. https://doi.org/10.1016/j.jue.2022.103453

Mellon, G., & Siegler, B. (2023). New Experimental Evidence on Anti-Asian Bias in White Parents' School Preferences. *Sociology of Education*, *96*(4), 255–274. https://doi.org/10.1177/00380407231173933

Morris, M. (2016). *Pushout: The Criminalization of Black Girls in Schools*. New Press, The.

Moslimani, M., Tamir, C., Budiman, A., Noe-Bustamante, L., & Mora, L. (2024, January 18). Facts About the U.S. Black Population. *Pew Research Center's Social & Demographic Trends Project*. https://www.pewresearch.org/social-trends/fact-sheet/facts-about-the-us-black-population/

National Center for Education Statistics. (2021a). *Digest of Education Statistics, 2021* (214.40). National Center for Education Statistics. https://nces.ed.gov/programs/digest/d21/tables/dt21_214.40.asp

National Center for Education Statistics. (2021b). *Parent and Family Involvement in Education Survey of the National Household Education Surveys Program.* (PFI-NHES:2019; Digest of Education

Statistics 2020, Table 206.40.). National Center for Education Statistics, U.S. Department of Education. https://nces.ed.gov/programs/digest/d20/tables/dt20_206.40.asp

National Center for Education Statistics. (2021c). *Public school choice programs* (Fast Facts). Institute of Education Sciences. https://nces.ed.gov/fastfacts/display.asp?id=6

National Center for Education Statistics. (2023). *Digest of Education Statistics, 2023.* (Table 104.70). National Center for Education Statistics. https://nces.ed.gov/programs/digest/d21/tables/dt21_214.40.asp

Okonofua, J. A., & Eberhardt, J. L. (2015). Two Strikes: Race and the Disciplining of Young Students. *Psychological Science*, *26*(5), 617–624. https://doi.org/10.1177/0956797615570365

Pattillo, M. (2015). EVERYDAY POLITICS OF SCHOOL CHOICE IN THE BLACK COMMUNITY. *Du Bois Review: Social Science Research on Race*, *12*(1), 41–71. https://doi.org/10.1017/S1742058X15000016

Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are Achievement Gaps Related to Discipline Gaps? Evidence From National Data. *AERA Open*, *5*(4), 2332858419875440. https://doi.org/10.1177/2332858419875440

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Phi Delta Kappan International. (2022). 2022 Poll Results | PDK Poll. *PDK Poll of the Public's Attitudes Toward the Public Schools The 54th Annual PDK Poll*. https://pdkpoll.org/2022-pdk-poll-results/

Posey-Maddox, L. (2017). Schooling in suburbia: The intersections of race, class, gender, and place in black fathers' engagement and family–school relationships. *Gender and Education*, *29*(5), 577–593. https://doi.org/10.1080/09540253.2016.1274389

Posey-Maddox, L., de Royston, M. M., Holman, A. R., Rall, R. M., & Johnson, R. A. (2021). No Choice Is the "Right" Choice: Black Parents' Educational Decision-Making in Their Search for a "Good" School. *Harvard Educational Review*, *91*(1), 38–61. https://doi.org/10.17763/1943-5045-91.1.38

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The Geography of Racial/Ethnic Test Score Gaps. *American Journal of Sociology*, *124*(4), 1164–1221. https://doi.org/10.1086/700678

Reuther, K. (n.d.). *Respondent Sourcing and Security in 3 Simple Steps—Centiment*. Centiment. Retrieved April 28, 2024, from https://www.centiment.co/blog/www.centiment.co/blog/respondent-sourcing-security-in-3-simple-steps

Reynolds, R. E., Howard, T. C., & Jones, T. K. (2015). Is this what educators really want? Transforming the discourse on Black fathers and their participation in schools. *Race Ethnicity and Education*, *18*(1), 89–107. https://doi.org/10.1080/13613324.2012.759931

Rowley, K. J., & McNeill, S. M. (2021). Choosing Predominantly Black Magnet Schools: The Importance of School Racial Composition in Parental Decision-Making. *Journal of School Choice*. https://www.tandfonline.com/doi/abs/10.1080/15582159.2020.1863136

Saporito, S., & Lareau, A. (1999). School Selection as a Process: The Multiple Dimensions of Race in Framing Educational Choice*. *Social Problems*, *46*(3), 418–439. https://doi.org/10.2307/3097108

Schneider, M., & Buckley, J. (2002). What Do Parents Want From Schools? Evidence From the Internet. *Educational Evaluation and Policy Analysis*, *24*(2), Article 2. https://doi.org/10.3102/01623737024002133

Schneider, M., Marschall, M., Teske, P., & Roch, C. (1998). School Choice and Culture Wars in the Classroom: What Different Parents Seek from Education. *Social Science Quarterly*, *79*(3), 489–501.

Sempelles, E., Cui, J., & McNamara, M. (2024). *First Look: Parent and Family Involvement in Education: 2023* (NCES 2024-113). National Center for Education Statistics. US Department of Education. http://nces.gov/pubsearch/pubsinfo.asp?pubid=2024113.

Simms, A., & Talbert, E. (2019). Racial Residential Segregation and School Choice: How a Market-Based Policy for K-12 School Access Creates a "Parenting Tax" for Black Parents. *Phylon (1960-)*, *56*(1), 33–57.

Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research*, *69*(8), 3139–3148. https://doi.org/10.1016/j.jbusres.2015.12.002

Soland, J. (2021). Are Schools Deemed Effective Based on Overall Student Growth Also Closing Achievement Gaps? Examining the Black–White Gap by School. *Teachers College Record*, *123*(12), 211–236. https://doi.org/10.1177/01614681211070876

te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2017). When size matters: Advantages of weighted effect coding in observational studies. *International Journal of Public Health*, *62*(1), 163–167. https://doi.org/10.1007/s00038-016-0901-1

Teske, P., Fitzpatrick, J., & Kaplan, G. (2006). "The Information Gap?" *Review of Policy Research*, *23*(5), 969–981. https://doi.org/10.1111/j.1541-1338.2006.00245.x

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197. https://doi.org/10.1016/j.chb.2017.08.038

UCLA: Statistical Consulting Group. (n.d.). *A Practical Introduction to Factor Analysis: Exploratory Factor Analysis*. Retrieved January 4, 2025, from

https://stats.oarc.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/

Ukanwa, K., Jones, A. C., & Turner, B. L. (2022). School choice increases racial segregation even when parents do not care about race. *Proceedings of the National Academy of Sciences*, *119*(35), e2117979119. https://doi.org/10.1073/pnas.2117979119

Virginia Department of Education. (2022, March 2). *Viriginia School Quality Profiles: Richmond City Public Schools*. Virginia School Quality Profiles. https://schoolquality.virginia.gov/divisions/richmond-city-public-schools#desktopTabs-3

Waitoller, F. R., Nguyen, N., & Super, G. (2019). The irony of rigor: 'No-excuses' charter schools at the intersections of race and disability. *International Journal of Qualitative Studies in Education*, *32*(3), 282–298. https://doi.org/10.1080/09518398.2019.1576939

Waitoller, F. R., & Super, G. (2017). School Choice or the Politics of Desperation? Black and Latinx Parents of Students with Dis/Abilities Selecting Charter Schools in Chicago. *Education Policy Analysis Archives*, *25*(55). https://eric.ed.gov/?id=EJ1144449

Wallace, D. (2017). Distinctiveness, deference and dominance in Black Caribbean fathers' engagement with public schools in London and New York City. *Gender and Education*, *29*(5), 594–613. https://doi.org/10.1080/09540253.2017.1296118

Welsh, R. O. (2023). Up the down escalator? Examining a decade of school discipline reforms. *Children and Youth Services Review*, *150*, 106962. https://doi.org/10.1016/j.childyouth.2023.106962

Welsh, R. O., & Little, S. (2018). The School Discipline Dilemma: A Comprehensive Review of Disparities and Alternative Approaches. *Review of Educational Research*, *88*(5), 752–794. https://doi.org/10.3102/0034654318791582

Whitehurst, G. J. (2017). *Education Choice and Competition Index 2016: Summary and Commentary*. Center on Children and Families at Brookings. https://www.brookings.edu/articles/the-2016-education-choice-and-competition-index/