education policy analysis archives

A peer-reviewed, independent,
open access, multilingual journal

Arizona State University

Volume 31 Number 43 April 25, 2023 ISSN 1068-2341

A Seat at the Table: Lessons from Tennessee's Rapid Achievement and Equity Gains[1]

Kaylee T. Matheny
Stanford University
United States

Citation: Matheny, K. T. (2023). A seat at the table: Lessons from Tennessee's rapid achievement and equity gains. *Education Policy Analysis Archives, 31*(43). https://doi.org/10.14507/epaa.31.7481

Abstract: This study complements literature on state policymaking by exploring one state's achievement trends and policymaking during a period of remarkable achievement gains. I use mixed methods, drawing on hierarchical linear modeling and semi-structured interviews with policy actors, to explore the case of Tennessee in the post-recession period. Results indicate that Tennessee's school districts improved faster than districts in other states during this period without doing so at the expense of historically disadvantaged student groups. I argue that Tennessee's trends are a result of strong policies and a robust approach to the policymaking process.

Keywords: state policy; achievement, achievement gaps; hierarchical linear modeling; mixed methods

[1] The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B140009 to the Board of Trustees of the Leland Stanford Junior University. The opinions expressed are those of the author and do not represent views of the Institute, the U.S. Department of Education, or the Board of Trustees of the Leland Stanford Junior University.

Journal website: http://epaa.asu.edu/ojs/
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 2/22/2022
Revisions received: 7/18/2022
Accepted: 7/18/2022

Un asiento en la mesa: Lecciones de los rápidos logros y ganancias de equidad de Tennessee
Resumen: Este estudio complementa la literatura sobre la formulación de políticas estatales mediante la exploración de las tendencias de logro y la formulación de políticas de un estado durante un período de logros notables. Utilizo métodos mixtos, basados en modelos lineales jerárquicos y entrevistas semiestructuradas con actores políticos, para explorar el caso de Tennessee en el período posterior a la recesión. Los resultados indican que los distritos escolares de Tennessee mejoraron más rápido que los distritos de otros estados durante este período sin hacerlo a expensas de los grupos de estudiantes históricamente desfavorecidos. Argumento que las tendencias de Tennessee son el resultado de políticas sólidas y un enfoque sólido para el proceso de formulación de políticas.
Palabras clave: política estatal; logros; brechas de logros; modelado lineal jerárquico; métodos mixtos

Um lugar à mesa: Lições das rápidas conquistas e ganhos de equidade do Tennessee
Resumo: Este estudo complementa a literatura sobre a formulação de políticas estaduais, explorando as tendências de realização de um estado e a formulação de políticas durante um período de notáveis ganhos de realização. Utilizo métodos mistos, com base em modelagem linear hierárquica e entrevistas semiestruturadas com atores políticos, para explorar o caso do Tennessee no período pós-recessão. Os resultados indicam que os distritos escolares do Tennessee melhoraram mais rapidamente do que os distritos de outros estados durante esse período, sem prejudicar grupos de alunos historicamente desfavorecidos. Eu argumento que as tendências do Tennessee são resultado de políticas fortes e uma abordagem robusta do processo de formulação de políticas.
Palavras-chave: política de estado; desempenho; lacunas de desempenho; modelagem linear hierárquica; métodos mistos

A Seat at the Table: Lessons from Tennessee's Rapid Achievement and Equity Gains

In 2013, the National Assessment for Educational Progress (NAEP) results indicated Tennessee had seen the greatest score increase since 2011 among all 50 states (Zubrzycki, 2013). A few years later, in 2017, a national study on district-level educational opportunity in early and middle childhood made headlines when it showed Tennessee as a beacon of growth in an ocean of declining test scores as children progress through school (Aldrich, 2017; Camera, 2017; Drum, 2017; Reardon, 2017/2019). These reports beg the question, what happened in Tennessee?

In this study, I use hierarchical linear modeling to identify whether Tennessee's achievement gains in the post-recession period hold after accounting for demographic changes and coarse policy indicators and whether these achievement gains occurred through privileging the achievement of historically advantaged or disadvantaged students. I contextualize these findings using semi-structured interviews with state policymakers about what occurred in Tennessee over the study period. Finally, I probe the narrative policymakers construct by disaggregating overall achievement trends and trends in disparities by year and compare results to the timeline in participants' narratives.

This study offers three key contributions to education policy literature. First, this study offers a key case of applying the conceptual policymaking process in practice. This is particularly important to understand at the state level, where high-impact policies are legislated (e.g., Jackson et al., 2016; Spencer & Maldonado, 2021; Wong et al., 2008). Second, this study complements existing

literature on state-level education policy by using a mixed methods approach to understand one state in-depth. Mixed methods are growing more popular in education policy research (see Chesnut et al., 2018), and this study illustrates how quantitative and qualitative methods can inform one another through an iterative analytic approach. Finally, by explicating the case of Tennessee, this study identifies features of state policy contexts that may facilitate achievement growth while narrowing achievement disparities. Identifying places with these achievement patterns is an important area for future work (see Atteberry et al., 2021; Matheny et al., 2023). Although this analysis may offer lessons for education policy implementation broadly, it may be particularly meaningful in the present moment; over the last few years, decades of national achievement gains have come to a grinding halt (U.S. Department of Education, 2019b). Further, early reports suggest the COVID-19 pandemic has had disproportionate impacts on low-income communities and communities of color, potentially widening achievement disparities between these historically disadvantaged communities and their more advantaged counterparts (Dorn et al., 2020).

Literature Review

State-Level K-12 Education Policy Change, 2009 to 2015

To provide important context for the education policy landscape, I begin by describing national policy reforms in the years following the recession. The first came as part of the Obama administration's American Recovery and Reinvestment Act (ARRA) (2009), through which $4.35 billion were allocated to fund Race to the Top. The U.S. Department of Education asked Race to the Top applicants to propose reforms in four areas: developing competitive standards and assessments, building data systems to measure student success and inform education leaders, recruiting and retaining effective teachers and leaders, and turning around low-achieving schools (ARRA, 2009). All told, 18 states and Washington, DC won Race to the Top grants of between $17 and $700 million. Even many states that did not win Race to the Top still implemented versions of these policies (Howell, 2015). However, there is some descriptive evidence that suggests winning Race to the Top did impact states' policy landscapes. For example, Howell (2015) finds that states that won Race to the Top implemented more policies overall and developed more rigorous standards than both their non-winning applicant and non-applicant counterparts. Tennessee was one of the first winners of Race to the Top.

Despite President Obama's Race to the Top initiative, which privileged states' agency in developing reforms that worked for their contexts, there were still provisions from 2001's No Child Left Behind (NCLB) that many states found constraining (Ayers & Owen, 2012). Beginning in 2012, 43 states were granted NCLB waivers (U.S. Department of Education, 2016), requiring states to set new goals for student achievement, adopt college- and career-ready standards, hold schools accountable for increasing student achievement, and improve teacher and principal effectiveness (U.S. Department of Education, 2016). As part of these requirements, many states adopted a framework naming the lowest-performing 5% of schools as "Priority Schools" and the 10% of schools with the largest achievement disparities between historically advantaged and disadvantaged student groups (considering racial and ethnic background, economically disadvantaged status, and English learner status) as "Focus Schools" (Tennessee Department of Education, 2017). This was a key reform associated with NCLB waivers, as some evidence suggests accountability systems targeting specific student groups can encourage schools to direct efforts toward those students (Figlio & Loeb, 2011). Many states instituted this reform, with mixed results. For example, Dee and Dizon-Ross (2019) argue that in Louisiana, assigning schools to "focus" status did not have an identifiable impact on school performance. However, Bonilla and Dee (2020) find that in Kentucky, ostensibly the same reform significantly raised proficiency rates for disadvantaged students. For both

studies, authors argue that results may at least partially be attributable to fidelity of policy implementation. This underscores the importance of analyzing policy implementation processes to understand heterogeneous impacts of education policies. Tennessee, too, adopted this reform.

One notable feature of the "focus school" framework is that states were given significant flexibility in implementing this reform (Tennessee Department of Education, 2017). For example, Bonilla and Dee (2020) argue that a potential explanation for the policy's effectiveness in Kentucky is that Kentucky used a "super subgroup" measure to aggregate disadvantaged students so schools with smaller student populations would still be held accountable for structurally disadvantaged students' performance. Tennessee's accountability framework had a similar concept. In addition to designating focus schools, the state also adopted a district accountability framework (with "reward," "priority," and "focus" districts) that identified districts in need of improvement based on student achievement and the achievement of disadvantaged student groups (Jones, 2014). This means that a single school in a district may not have a large enough population of disadvantaged students to warrant focus school status, but the district can still be held accountable for the performance of student subgroups across schools. Tennessee's district accountability system, then, offers a unique adaptation of the NCLB waiver requirements. It is possible this may play a role in Tennessee's district-level achievement.

In addition to new accountability approaches, nearly all states implemented a turnaround strategy that impacted school governance. In Tennessee, approximately half of "priority" schools were subsumed under the Achievement School District, which included both state-run and charter schools (Stockard, 2018). Other "priority" schools were placed under the management of districts ("iZone" schools; Stockard, 2018). While studies have found mostly null impacts of the Achievement School District on achievement, iZone schools saw achievement gains (Pham et al., 2020; Zimmer et al., 2017).

Like many states granted NCLB waivers, Tennessee also adopted a rigorous teacher evaluation framework. Teacher quality is critical for student performance (Opper, 2019). Although Tennessee was the first state to calculate value-added scores, with principals and teachers receiving information about student growth in their schools and classes since 1993 (Tennessee SCORE, 2017), it was not until 2010's First to the Top Act that Tennessee incorporated these data in teacher evaluation (Tennessee SCORE, 2017). In fact, Tennessee's commitment to teacher evaluation reform was a key component of the state's award-winning Race to the Top application; with the help of the award's $500 million, the state's teacher evaluation system underwent a makeover to not only include student growth data but also significantly more teacher observations: Tennessee went from requiring two teacher observations over 10 years to four or six per year, depending on the experience of the teacher (Tennessee Department of Education, 2012).

Evidence suggests that rigorous teacher evaluation improves student achievement (Taylor & Tyler, 2012). However, recent work from Tennessee argues assigning teachers a high number of evaluations—a result of Tennessee's teacher evaluation reforms—results in lower first evaluation scores, particularly for early-career teachers (Hunter, 2020). This has implications for teacher morale and retention and could lead to teacher turnover (Paufler, 2018), which is negatively associated with achievement trends and trends in disparities (Matheny et al., 2023). Rodriguez et al. (2020) find that in Tennessee, teacher evaluation reforms did indeed cause turnover; however, most of this turnover was concentrated among less effective teachers, improving the overall effectiveness of the teacher workforce. This likely has positive implications for Tennessee's achievement over the study period. However, recent reports from Tennessee also indicate that ineffective teachers are still concentrated among disadvantaged students (Podesta, 2019), suggesting additional room for growth in effectively distributing educator talent.

In addition to considering important policies that took place over the study period, it is important to note that policy change does not take place in a vacuum. In their policy brief outlining what has worked in Tennessee's low-performing schools, Henry et al. (2018) assert that the ingredients in Tennessee's reforms "are mutually supportive components of a comprehensive strategy for effective school reform" (p. 1). Essentially, the state's recipe is neither prescriptive nor able to be easily disassembled.

Conceptual Framework

The Public Policy Process

Due to provisions associated with Race to the Top grants and NCLB waiver receipt, other states adopted many of the same reforms as Tennessee. However, Tennessee stands alone in the achievement gains made in the post-recession period (2009-2015; see U.S. Department of Education, 2019a). I argue here the state's approach to policy implementation is at least as important as the policies themselves. For the qualitative portion of this study, to explicate the narrative of what occurred in Tennessee and identify potential indicators of a fruitful policy context, I use Buck's (2006) conceptual framework for the policymaking process. Here, I describe each stage of the policymaking process, the ways in which the organizational structure of the k-12 education system exacerbates the already-challenging nature of policymaking, and potential indicators for a successful policy context.

Agenda-Setting

This framework begins with agenda-setting, or the process of identifying priority areas for institutional policy attention. Buck (2006) distinguishes between two types of agendas: "systemic agendas," which are not constrained by time or resources and include "all the issues that a political community agree need to be resolved and that they also agree are within governmental authority"; and "institutional agendas," which often "work within strict time constraints" and are "specific, concrete, and limited" (p. 41). Typically, items first appear on the systemic agenda; when they become too problematic to be ignored, they advance to the institutional agenda.

Because systemic agendas are not constrained by time or resources, they can theoretically include all items of importance to stakeholders at all levels of the education hierarchy. However, spots on the institutional agenda are much more exclusive, requiring the allocation of limited resources. This necessitates a more unified (or more authoritative) effort at agenda-setting. This is difficult in education, as not all stakeholders agree about the purpose of educating students (Labaree, 1997). Because there are numerous potential goals for education, stakeholders at different levels in the educational hierarchy—from the U.S. Secretary of Education to students themselves—may have different ideas of what they hope education can accomplish. One indicator of successful policymaking, then, is establishing a unified agenda of policy priorities across stakeholders.

Policy Formulation and Legitimation

The second step in the public policy process is policy formulation and legitimation, or the process through which policies materialize from abstract ideas to concrete proposals. This necessitates bringing different stakeholders to the table to agree on how to define the problem, compromise on a solution, and build stakeholders' trust in the policy (Buck, 2006). This suggests building trust and establishing legitimacy are additional indicators of successful policymaking. Further, defining the problem requires stakeholders to develop shared language around what exactly the problem is and how to address it. As Elmore and McLaughlin (1988) note, "[R]eforms cannot be expected to have large-scale or long-term effects unless they involve substantial dialogue among

policy, administration, and practice" (p. vi). Effective dialogue helps minimize ambiguity and brings actors with different roles to the table for problem-defining and policy formulation. Therefore, another indicator of successful policymaking is pursuing intentional dialogue across actors in the education policy space.

Policy Implementation

Policy implementation, too, is a stage of compromise. Policymakers and policy implementers often have different stores of knowledge, relationships with stakeholders, and levels of engagement with the decisions made in the day-to-day implementation of a policy (Buck, 2006; Tseng, 2012). The "loosely-coupled" nature of the education system, whereby actors at each level—from the state, to the district, to the school, to the classroom, to the family—have significant autonomy, exacerbate the challenges of policymaking (Weick, 1976). This structure makes communication difficult as policies trickle down through tiers in the education hierarchy (Weick, 1976). This, researchers argue, is the key challenge for reforming education (Elmore & McLaughlin, 1988; Labaree, 2010). Places that have successfully implemented reforms will likely have scaffolded policy rollouts to overcome communication challenges. In addition, bureaucracies maintain legitimacy through the collection and dissemination of institutional knowledge (Weber, 1922/2019). Collecting knowledge and strategically informing relevant stakeholders, then, may be another feature that scaffolds successful policy implementation.

Even if policies are clearly communicated, there is likely to be some uncertainty in how schools and teachers should implement reforms. This uncertainty, in turn, can stifle reform efforts. Podolny (2001) argues that one way of overcoming an institution's egocentric uncertainty (in this case, actors in a school not knowing how to effectively interpret or enact a policy) is by bridging structural holes (holes in the information flow between groups; see Burt, 2004). In the education system, this might look like having dedicated roles for communication between tiers (e.g., a designated person whose job it is to maintain contact between the school district and the state department). Having bridges between relevant policy actors to mitigate uncertainty is another potential indicator of successful policymaking.

Policy Evaluation

Policy evaluation is about much more than analytical questions of whether X affects Y. While these types of summative evaluations are important, there are formative evaluations that occur along the way, allowing for "mid-course correction if the policy goals are being bypassed or if new and unintended consequences seem imminent" (Buck, 2006, p. 56). Thus, while evaluation is the last stage here, it often leads to circling back to other stages—giving this seemingly linear perspective of the policymaking process a more iterative nature. In practical terms, this could mean that an education context with successful policy reforms also incorporates formative evaluation into its policy process to make improvements over time.

Methods

Research Questions

In this study, I investigate the following questions:

1. Do Tennessee's achievement gains in the post-recession period hold after accounting for demographic changes and national policy reforms? Further, are these gains shared by both historically advantaged and historically disadvantaged student groups?
2. How do Tennessee policymakers account for these achievement gains?

3. To what extent does the narrative Tennessee's policymakers construct align with the achievement gains and disparity narrowing in the study period?

Mixed Methods Approach

I take a mixed methods approach, following the dialectic mixed methods paradigm (Greene, 2007). Here, I use the quantitative and qualitative analyses to provide insight on aspects of Tennessee's achievement patterns from different perspectives—quantitatively, through the broad view of what the achievement patterns were and the results that onlookers see, and qualitatively, through the more detailed, insider view of how critical stakeholders argue these patterns were made manifest by actors in the state's policy space. In this study, a mixed methods inquiry serves two purposes among those Greene et al. (1989) outline. The first is development, whereby (a) the quantitative analysis established the extent to which both gains and disparity-narrowing were shared across districts in Tennessee, supporting a qualitative design focused on state-level factors rather than specific district-level strategies; and (b) the chronology that emerged in the qualitative analysis yielded subsequent disaggregated quantitative analyses by year. The second purpose among those Greene et al. (1989) outline is expansion, whereby I quantitatively establish that Tennessee's achievement patterns are worth investigating and qualitatively explore potentially replicable and scalable strategies for state-level education reforms. These purposes led me to integrate data iteratively. First, I quantitatively analyzed state-level trends in achievement, then qualitatively investigated the policy context in which Tennessee's state education reforms were implemented, and then quantitatively probed the narrative policymakers construct by disaggregating achievement by year.

Quantitative Data and Method

I provide descriptive evidence of trends in Tennessee's achievement and achievement disparities between historically advantaged and historically disadvantaged groups. I also investigate the extent to which these trends are associated with demographic changes and state-adopted reforms. To do so, I use data from the Stanford Education Data Archive (SEDA) (Fahle et al., 2021). SEDA encompasses 430 million standardized test performance observations in reading and math for students in Grades 3-8 in nearly every geographic school district from 2009 to 2019. SEDA uses these observations to generate standardized and grade-level-equivalent estimates of student achievement and achievement disparities for every district for which there are sufficient observations. Note that SEDA estimates are not reported for district-grade-year-subject observations with fewer than 20 students; for disparities, estimates are reported only if both groups' scores are based on at least 20 students. This results in smaller sample sizes for disparity trends than for the overall trend. Additionally, I use district-level trends, rather than school-level trends, because SEDA does not report estimated disparities at the school level. Still, districts are a key unit through which policies are disseminated (Roza, 2010); further, in Tennessee, district superintendents are key stakeholders in the policy process, and the district accountability framework was key to its reforms.

Outcomes

The outcomes in this study are the changes in average district-level achievement and average district-level achievement disparities between poor and non-poor students, Black and White students, and Hispanic and White students. For modeling, I use the standardized estimates from SEDA; however, for interpretation, I contextualize findings with grade-level-equivalent units, also available in SEDA.

Key Predictor

The key predictor is a binary indicator for whether a school district is in Tennessee. The coefficients of interest represent the extent to which Tennessee's achievement patterns differ from those in other U.S. states (and Washington, DC).

Covariates

In this study, I am interested not only in Tennessee's raw achievement differences, but also the extent to which these differences may be explained by increasing or decreasing proportions of historically advantaged students and state context variables, including the adoption of national policy reforms. Therefore, I include measures for these constructs in my models.

I construct district-level demographic covariates and state-level means of demographic covariates from the SEDA covariate data and the Common Core of Data (CCD), including the CCD's Local Education Agency Finance Survey (F-33). I control for student and community demographic characteristics, including the proportion and change in the proportion of Asian students, Black students, Hispanic students, and economically disadvantaged students, as well as the mean and change in the community's socioeconomic context (see Fahle et al., 2021). I use this strategy to isolate the impact of the state's policy landscape as opposed to levels and changes in the enrollment of historically advantaged or disadvantaged students. For example, if students facing structural barriers (such as racism or poverty) to educational opportunity exited Tennessee at high rates over the study period, this could "inflate" Tennessee's estimated achievement without any substantive change in state policy. I also control for the initial achievement (or disparity, in disparity-trend models) in each district and state to control for any impact that starting the period as a low- or high-achieving district or state may have on trends in achievement.

In addition, I use coarse controls for the major national policies described in the literature review. These include the amount of Race to the Top funds won (in per pupil dollars) and binary variables indicating the round in which funds were won (Howell, 2015) as well as a binary variable for whether a state secured an NCLB waiver (U.S. Department of Education, 2016). If improvement gains are primarily among those states receiving NCLB waivers, for example, this may be driving some of the improvement in Tennessee. Further, for additional state policy context, I include the average and the change in state per pupil educational expenditures. For descriptive statistics of the variables included in this analysis, see Appendix Table 1. For specific states' coarse policy indicators, see Appendix Table 2.

Analytic Sample

Due to Tennessee's statewide testing obstacles brought on by using a new vendor in 2016 (Tatter, 2016), I restrict my sample to observations between the spring 2009 and 2015 school years. It is worth noting that SEDA's design links it to NAEP to make comparisons across states, as NAEP is one of the only assessments administered to a representative sample of students in every state (Reardon et al., 2021). However, NAEP gives only a bird's eye view of what is happening at the state level; SEDA offers a more granular district-level look, so SEDA is helpful to understand achievement patterns in the average school district—a critical unit for state policy dissemination (Roza, 2010).

Modeling Strategy

I use precision-weighted hierarchical linear models (with districts nested within states) to estimate the extent to which being in Tennessee predicts district-level achievement gains. I conduct this analysis in two steps, as SEDA estimates are broken down into state-district-grade-year-subject

estimates (and subgroup, for subgroup estimates). In step one, I use district-grade-year-subject observations of student achievement to estimate the annual district-level *trend*, or change, in student achievement. Grades and subjects are pooled together, indicating this is the shared trend across grades (3, 4, 5, 6, 7, and 8) and subjects (math and reading). The parameter of interest is the slope across years, indicating how much students learn, in standard deviations, each year. This modeling strategy follows other approaches using SEDA (see Jang & Reardon, 2019; Matheny et al., 2023; Reardon, 2019). For this model, see Appendix A.

Step one yields each school district's estimated annual change in achievement from 2009 to 2015. In step two, I identify the relationship between Tennessee and district-level trends using the following model:

$$\hat{\alpha}_{ds} = \beta_0 + \beta_n(\mathbf{X} - \overline{\mathbf{X}})'_{ds} + e_{ds} + \varepsilon_{ds}$$

$$\beta_0 = \theta_{00} + \theta_{01} Tennessee_s + \theta_{0k}(\mathbf{Z} - \overline{\mathbf{Z}})'_s + r_0$$
$$\beta_n = \theta_{n0} + r_n$$

$$e_{ds} \sim N(0, \psi^2_{ds}); \varepsilon_{ds} \sim N(0, \hat{\pi}^2_{ds}); [r_0, \dots, r_n] \sim MVN(0, \boldsymbol{\tau})$$

Where $\hat{\alpha}_{ds}$ represents the estimated annual achievement trend for district *d* in state *s*; $(\mathbf{X} - \overline{\mathbf{X}})'_{ds}$ represents a vector of *n* mean-centered district covariates; $(\mathbf{Z} - \overline{\mathbf{Z}})'_s$ represents a vector of *k* mean-centered state-level covariates; and $Tennessee_s$ is an indicator for whether or not the state is Tennessee. The first error term, e_{ds}, represents random error. The second term, ε_{ds}, represents the square root of the error variance. Specifically, in the estimation, $\hat{\pi}^2_{ds}$ is treated as a known parameter: It is the error variance of the estimated district-level trend in achievement, estimated in step one. Importantly, some achievement estimates are more reliable than others (see Fahle et al., 2021), so in this model, I upweight estimates with lower error variance. This provides the precision-weighting for the model, indicated as ε_{ds}. ψ^2_{ds} and $\boldsymbol{\tau}$ are estimated. I fit the model via maximum likelihood using the HLM v7 software.

The parameter of interest is θ_{01}, the difference between the average state-level trend (θ_{00}) and the state-level trend in Tennessee, accounting for mean-centered district-level covariates $(\mathbf{X} - \overline{\mathbf{X}})'_{ds}$ and state-level covariates $(\mathbf{Z} - \overline{\mathbf{Z}})'_s$ in successive models. Covariates are grand-mean-centered such that the constant in Table 1 can be understood as the increase in achievement in the average U.S. district (outside of Tennessee) over the study period.

Qualitative Data and Method

For the qualitative analysis, I conduct a narrative inquiry (Creswell & Poth, 2017), which foregrounds chronology and includes identifying themes across participants' stories of an event (Cortazzi, 1993). Chronology is central to this study's quantitative analysis, as the very concept of an achievement *trend* implies change over time, as well as the qualitative work, through the story stakeholders tell about what happened in Tennessee through the study period. To craft this story, I collect data through interviews with prominent stakeholders in Tennessee's education policy landscape who could describe the policies and processes enacted over the study period. I treat interviews with participants as oral histories, or participants' reflections on the events that occurred in Tennessee over the study period (Plummer, 1983).

The primary data for the qualitative analysis come from 15 interviews with 22 different leaders in Tennessee's education policy sector conducted in spring 2019. While conducting interviews in this period means that participants had time to reflect on the full extent of the policymaking process over the study period (2009-2015)—an asset for this study—it also means

potential concerns for "recall bias," through which participants misremember events due to delayed recollection. To account for this, I validated facts using publicly available documents (e.g., Frist, 2009), online newspaper and magazine articles, and the Department of Education website. I also cross-referenced key narrative details across interviews, and I report findings based on consistency across interviews. These strategies help establish trustworthiness of the qualitative findings.

Participants include ten current and former affiliates with the state's department of education; six administrators in one of Tennessee's large, urban districts; three researchers; two policy advocates; and one legislator (referred to throughout this study as "policymakers"[2]). I recruited participants primarily through directly contacting people with roles in policy implementation and those with a history of leadership in Tennessee's education policy sector, though I also used snowball sampling to collect a smaller number of cases. This recruitment strategy and sample size is appropriate given the limited population from which to draw. Additionally, I follow Small (2009) in applying Yin's (2002) case study logic to this in-depth interview study, for which the objective of sampling is saturation. Based on the consistency of even the earliest interviews, I argue that this sample size is appropriate for the purposes of this study. Notably, there are perspectives missing from this sample, including those of teachers, district leaders in rural and suburban settings, and parents. Because this inquiry is about the education policy process broadly, and because the interviews I describe did reach a level of consistency indicative of saturation, this sample is sufficient for the current study.

In interviews, I asked questions designed to understand participants' roles in agenda-setting, policy development, and implementation; to what participants would attribute the state's achievement gains; and which features of the state might facilitate effective policy implementation. Interviews lasted between 20 minutes and one hour, with an average length of 44 minutes. I transcribed interviews and coded them in Dedoose, a qualitative analysis program. I employed a combination of coding schema. First, I used thematic coding, with both deductive (e.g., "organizational structure") and inductive (e.g., "alignment, consensus, or unification") codes. I developed deductive codes using my interview guide; I developed inductive codes by documenting patterns in language, sentiment, and narrative points that occurred across three or more interviews. In addition to thematic coding, I also employ narrative coding, with codes that indicate chronology (e.g., "2012") and narrative turning points (e.g., "catalyst"; Saldaña, 2016). I use the deductive and inductive codes as they align with narrative codes to construct the narrative in the findings section.

Although participants in this study do not represent a traditionally vulnerable population, a state's policy space lends itself to being a "small world," with a relatively small population from which to draw. Further, I wanted participants to feel they could speak freely regarding both the positive aspects of Tennessee's policy trends and potential areas for improvement. For this reason, when I recruited participants, I informed them that I would not use direct quotes from interviews when reporting results, choosing instead to craft a narrative from overarching themes and summaries of participants' responses. This is consistent with the narrative inquiry approach employed here (Creswell & Poth, 2017) and further substantiates the trustworthiness of participants' accounts.

---

[2] Throughout the study, I refer to state-level policy actors, including the participants in my study, as "policymakers." While the sample in the qualitative analysis includes researchers and district-level practitioners, I use "policymakers" as opposed to "practitioners" due to the narrative's centrality of the policymaking process. Further, I use "policymakers" to differentiate from the more general "stakeholders," which would necessarily include teachers, district leaders in rural areas, students, and parents, none of whom are represented in the interview sample.

My positionality as a graduate student seemed to be an asset when conducting this study. Many participants recognized my university and advisor and expressed enthusiasm about meeting with me. I am not from Tennessee, which is both an asset and a constraint to my inquiry: I had less knowledge of context-specific details (for example, in Tennessee, I found that people use the term "directors of schools" rather than "superintendents"), but I was also not personally invested in Tennessee's story. This allowed useful distance between myself as the researcher and the narrative I constructed with participants' stories. I may have also had greater access to the actors I interviewed given my status as a White woman in professional attire; in the elevator at the Tennessee Department of Education, I was mistaken for someone who was interviewing for a job, suggesting a social fit with the context in which several of my participants' interviews took place.

Results

Quantitative Results

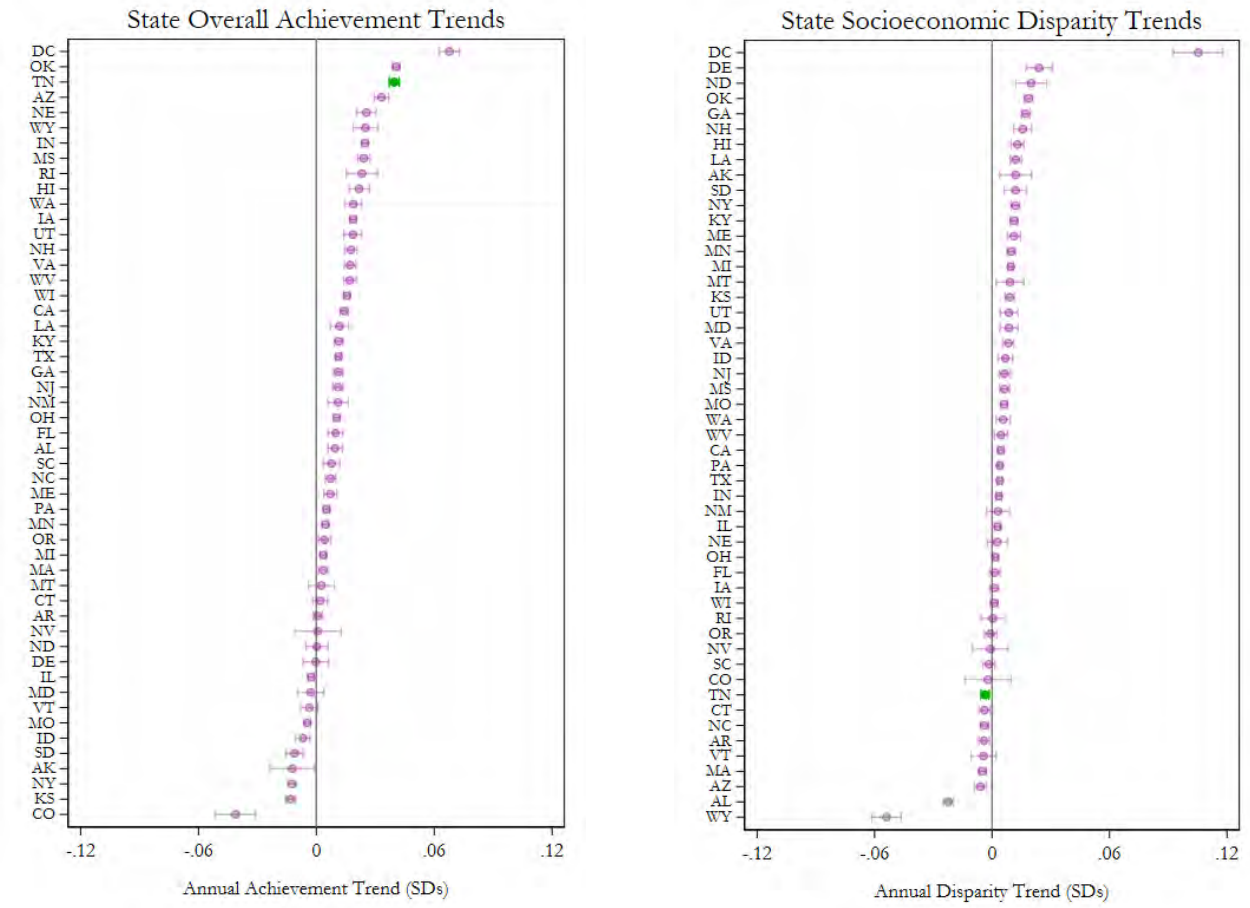
I first illustrate the extent to which Tennessee is an important case by showing its annual district-level achievement and disparity trends over the study period. Figure 1 includes four caterpillar plots that show how Tennessee's achievement trends compare to those of other states. From 2009 to 2015, Tennessee was the fastest-improving state in the nation (at a little more than 0.04 standard deviations per year, or nearly half a standard deviation over a decade); its achievement trends were second only to those of Washington, DC. This alone is notable, but it is only part of the story. Note that in the other plots, most states fall to the right side of the center vertical (x=0) line, indicating a *growing* achievement disparity (with the exception of the White-Hispanic disparity, for which most states see narrowing over the study period). Tennessee, conversely, narrows all three disparities measured here. It is eighth among states in narrowing its socioeconomic achievement disparity, second in narrowing its White-Black achievement disparity, and first in narrowing its White-Hispanic achievement disparity. This optimal combination—increasing overall achievement while narrowing all three disparities—occurs in only seven other states (Alabama, Florida, Mississippi, Oregon, North Carolina, South Carolina, and Rhode Island) over the study period, and none of these surpass the achievement gains nor the racial/ethnic disparity narrowing of Tennessee. The caterpillar plots show what is happening in each state on average, but they do little to illustrate the variation across districts within states. Indeed, it is possible that Tennessee's gains are driven by a small number of optimal districts with particularly significant trends. For example, this is the case in South Carolina, where average state-level achievement gains and narrowing achievement disparities make it appear among the most optimal places. However, very few districts experience this achievement pattern: Only 6.83% of South Carolina's districts are "optimal," compared to 14% of districts across the US.[3] Instead, its average state-level performance is driven by a few high-performing districts.

To illustrate the variation across districts within states with a focus on Tennessee, Figure 2 shows a panel of scatterplots, with Tennessee's districts in green. Figure 2 shows that in most districts in Tennessee, all students are improving: This is the case for all districts in the upper-right quadrants, where the achievement change for both the historically advantaged group and the historically disadvantaged group is greater than zero. Further, Tennessee's districts are primarily above the dashed line, indicating that in most places, the historically disadvantaged group is improving faster than the historically advantaged group, narrowing the achievement disparity. In fact, more than half of Tennessee's school districts both experienced achievement gains and narrowed achievement disparities over the study period—a greater share than in any other U.S. state.

[3] Source: Author's calculations using the Stanford Education Data Archive, v 4.1, years 2009-2015.

Figure 1

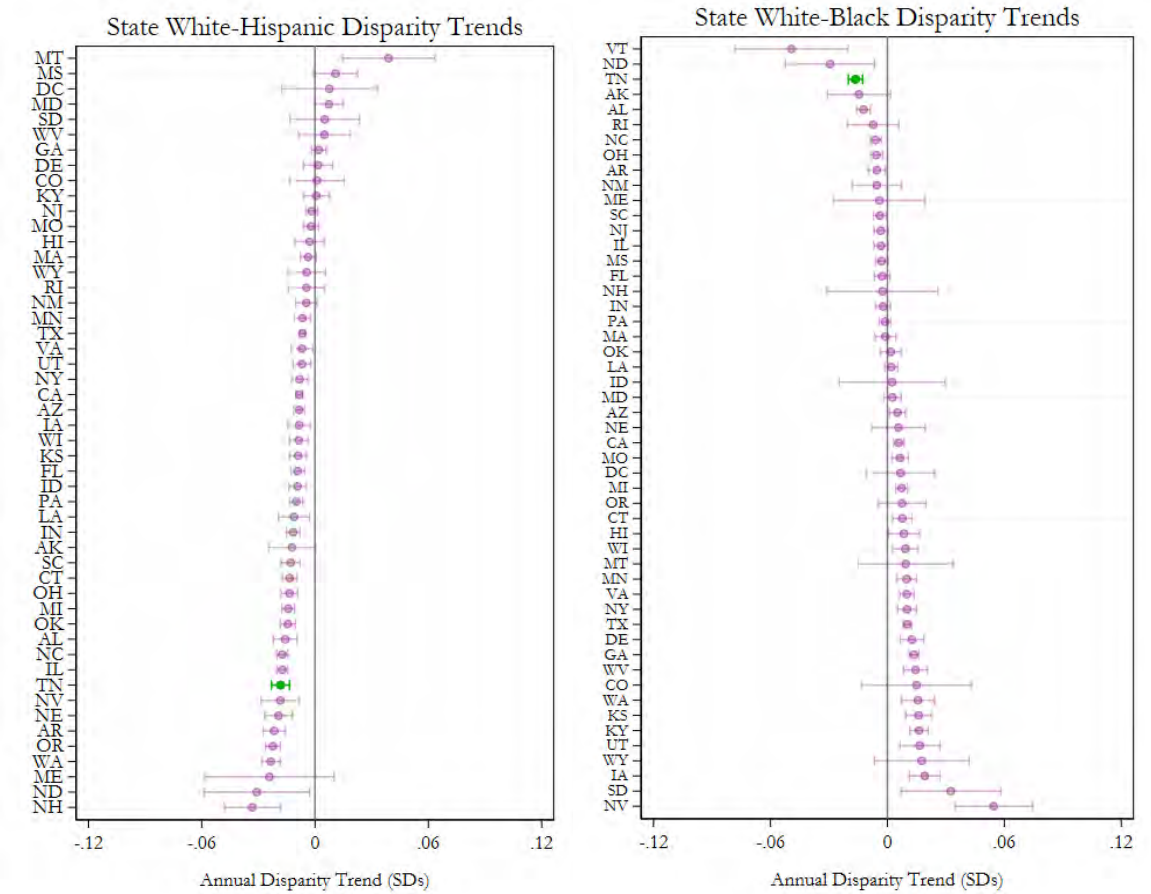
Caterpillar Panel of State Achievement Trends (Tennessee in Green)



Note: In this plot, states are ranked on the y-axis in accordance with their average annual achievement change or their disparity trends (positive to negative), and the x-axis represents the annual trend estimate in standard deviations. Dots indicate the estimated annual change in the district-level achievement or disparity for each state, and bars indicate 95% confidence intervals. For example, student performance in the average district in Tennessee increases by about 0.041 standard deviations each year over the study period. Vermont is omitted from the White-Hispanic disparity trend panel due to lack of available disparity data.

Figure 1 cont.

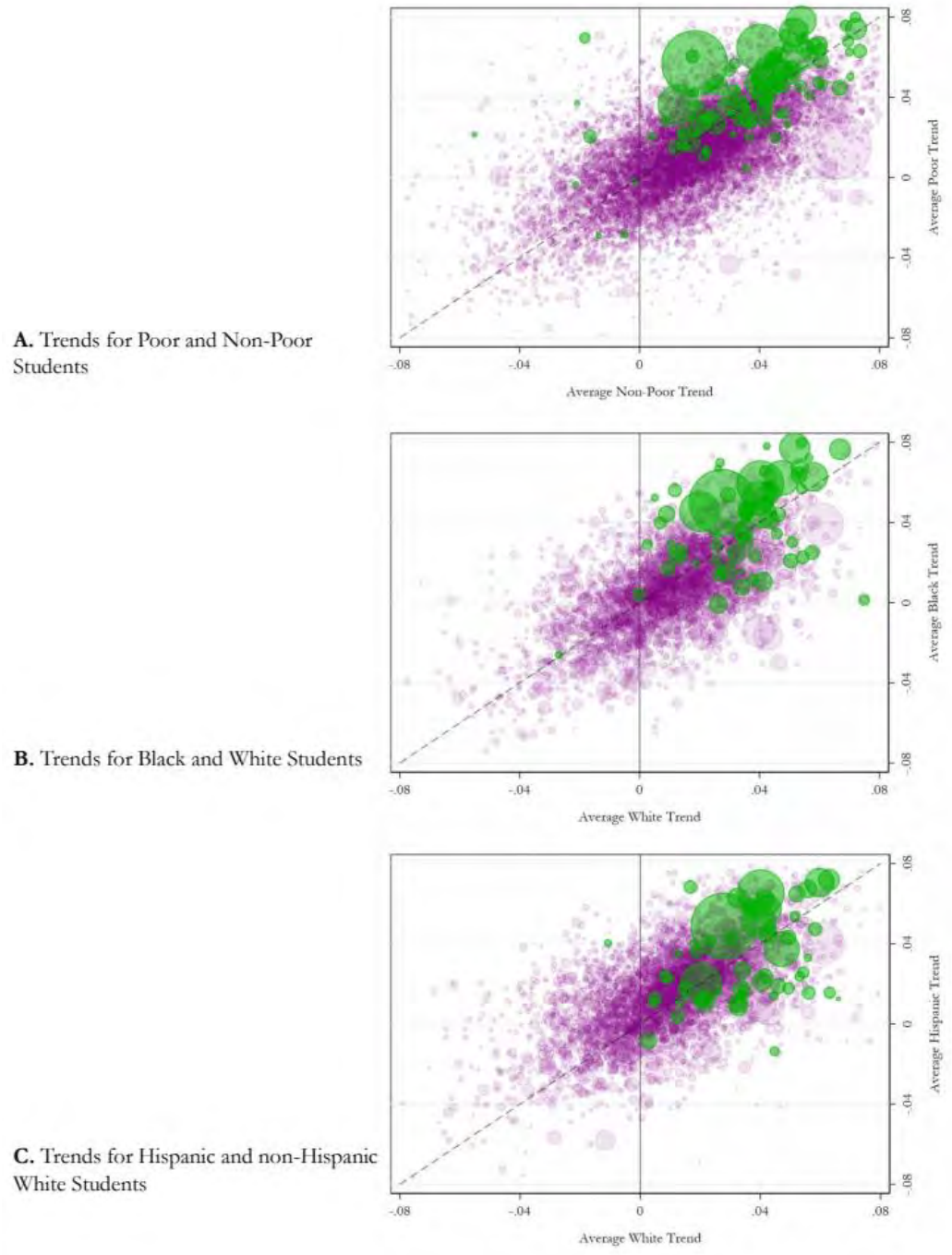
Caterpillar Panel of State Achievement Trends (Tennessee in Green)



Note: In this plot, states are ranked on the y-axis in accordance with their average annual achievement change or their disparity trends (positive to negative), and the x-axis represents the annual trend estimate in standard deviations. Dots indicate the estimated annual change in the district-level achievement or disparity for each state, and bars indicate 95% confidence intervals. For example, student performance in the average district in Tennessee increases by about 0.041 standard deviations each year over the study period. Vermont is omitted from the White-Hispanic disparity trend panel due to lack of available disparity data.

Figure 2

Scatterplot Panel Comparing Student Groups' Achievement Trends (Tennessee in Green)



Note: This figure shows the variation in achievement trends. Each circle is a school district, with larger circles indicating higher enrollment. The x-axis is the historically advantaged group's achievement, and the Y-axis is the historically disadvantaged group's achievement, so when Y > X (when a district is plotted above the Y = X line), the achievement disparity between the advantaged and disadvantaged group is narrowing over the study period. These plots show that for nearly every district in Tennessee, poor students', Black students', and Hispanic students' test scores increased over the study period. In many places, and particularly for Tennessee's largest districts, these gains were made while narrowing disparities.

Next, I describe the results from the hierarchical linear models, displayed in Table 1. In Panel A, for the overall achievement trend, Model 1 indicates that districts in Tennessee achieved at 0.033 standard deviations higher each year than districts in other U.S. states[4] (roughly equivalent to 1/10th of a grade level, or a full grade level's worth of additional learning over 10 years). This estimate does not change much with the inclusion of demographic characteristics or coarse policy indicators, though including the coarse policy indicators does double the standard error.

Table 1

Tennessee's Relationship to District-Level Achievement Trends

Outcome of Interest	Model 1	Model 2	Model 3	Model 4
Panel A. Overall	0.033*	0.032*	0.027	0.027
Achievement Trend	(0.014)	(0.014)	(0.015)	(0.031)
	[*0.008*]	[*0.008*]	[*0.007*]	[*0.008*]
(N)	11,317	11,317	11,317	11,317
Panel B. Socioeconomic	-0.011	-0.004	-0.002	-0.022
Disparity Trend	(0.012)	(0.012)	(0.013)	(0.030)
	[*0.006*]	[*0.041*]	[*0.022*]	[*0.017*]
(N)	8,347	8,347	8,347	8,347
Panel C. White-Black	-0.019*	-0.020*	-0.017	-0.042
Disparity Trend	(0.010)	(0.009)	(0.011)	(0.025)
	[*0.004*]	[*0.023*]	[*0.024*]	[*0.029*]
(N)	2,665	2,665	2,665	2,665
Panel D. White-Hispanic	-0.032**	-0.032**	-0.022**	-0.021
Disparity Trend	(0.010)	(0.010)	(0.012)	(0.024)
	[*-0.006*]	[*0.018*]	[*0.006*]	[*-0.005*]
(N)	3,455	3,455	3,455	3,455
Controls	None	District-Level Demographics	District- and State-Level Demographics	District- and State-Level Demographics, Coarse Policy Indicators

Note: This table displays the relationships between Tennessee and school districts' overall achievement (Panel A), socioeconomic achievement disparities (Panel B), White-Black achievement disparities (Panel C), and White-Hispanic achievement disparities (Panel D). Values represent coefficient (standard error) [*comparison group mean*]. Standard errors are clustered through HLM. Sample size reports the number of districts included in the analysis. M1 has no covariates. M2 controls for district-level student characteristics, including the proportion and change in the proportion of Asian students, Black students, Hispanic students, and poor students as well as district-level achievement at the start of the study period and average socioeconomic context. M3 includes these controls and adds state-level averages of the same variables. M4 includes all

[4] Note that the coefficient for Tennessee in each model indicates the difference for Tennessee relative to other states. For the population average of what is occurring in Tennessee, readers must add the constant (what is happening in the average U.S. state, without Tennessee) to the Tennessee coefficient. Thus, in Table 2 Model 1, readers can infer that in the average US state, students experience 0.008 standard deviations of increased achievement every year. Tennessee improves at 0.033 standard deviations faster than the average US state, or at the rate of approximately 0.041 standard deviations every year.

aforementioned controls and adds coarse policy indicators for how much money states won in the Race to the Top competition (in $100 per pupil dollars), which round states won in the Race to the Top competition, whether the state received an NCLB waiver, and the average and change in state educational expenditures (in $100 per pupil dollars). ǂ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

In addition to estimating the relationship between Tennessee and district improvement, I also estimate the relationship between Tennessee and narrowing socioeconomic achievement disparities, White-Black achievement disparities, and White-Hispanic achievement disparities. Note that negative coefficients indicate a narrowing (or a slower widening) of the achievement disparity relative to what is happening in the rest of the United States.

Table 1, Panel B reports the results for the estimated annual socioeconomic disparity narrowing in Tennessee. The coefficients for all models are insignificant due to the size of their standard errors (a byproduct of the estimation strategy used here and the variation across school districts in Tennessee), but they are all negative, toward disparity narrowing. This is not necessarily the experience for all districts, as the variation suggests, but this is what is occurring in the average school district in Tennessee. It is worth noting that because SEDA draws on population-level data from nearly every school district, coefficients do not represent estimates of a population average from a representative sample in the way that NAEP does; they represent the actual population average. The coefficient thus indicates a comparison of population-level averages—an important feature of the analyses here. This finding is notable because it suggests that Tennessee's gains, larger than those of any almost every other U.S. state over the same period, do not come at the expense of poor students' achievement. For a contrasting case, consider Washington, DC, the only place in the nation where achievement gains over the study period are even higher. Here, the socioeconomic disparity is increasing faster than in any other state, suggesting D.C.'s gains are primarily driven by non-poor students. In Tennessee, increases in standardized test scores among both poor and non-poor students drive achievement gains.

The estimates in Panel B are sensitive to the coarse policy indicators, which make the estimate even more negative and double its standard error. One potential explanation is that states that received NCLB waivers tend to see an increase in their socioeconomic achievement disparities over the study period. Because Tennessee follows a different trend and is, instead, experiencing little change in its socioeconomic achievement disparity, the estimate inflates.

In addition to the socioeconomic achievement disparity, I estimate Tennessee's relationship to the district-level White-Black achievement disparity. I report results in Table 1, Panel C. Tennessee experiences 0.019 standard deviations more closure each year than the average district in other US states. This equates to roughly half a grade level over 10 years. This coefficient is mostly robust to the inclusion of covariates, though it loses significance in Model 3. In Model 4, the coefficient is even more negative, suggesting the coarse policy indicators do little to attenuate the relationship. Across models, all coefficients are negative, indicating that on average, Black students are seeing higher gains than White students over the study period.

Finally, I estimate Tennessee's relationship to the White-Hispanic achievement disparity. I report results in Table 1, Panel D. Tennessee experiences 0.032 standard deviations more closure each year than the average district in other U.S. states. This equates to roughly three-quarters of a grade level over 10 years. First, it is worth noting that this is the first disparity panel in which the comparison group mean (the average trend in the White-Hispanic disparity in districts elsewhere in the US) has been negative across all observations, indicating that in the average district outside of Tennessee, the White-Hispanic disparity is narrowing. Second, the White-Hispanic disparity narrowing coefficient is larger (in Models 1, 2, and 3) than the socioeconomic or White-Black disparity narrowing coefficients. This is true despite the broader U.S. trend toward a narrowing

White-Hispanic disparity. In this case, too, the coarse policy indicators do little to explain Tennessee's trend.

One important consideration is whether districts with increasing achievement are also narrowing disparities. I gain traction on this question by identifying pairwise correlations between districts' improving achievement and achievement disparity narrowing for districts both inside and outside of Tennessee. I report these results in Appendix C. In general, I find no significant correlation between achievement gains and disparity narrowing, suggesting that Tennessee's achievement trends are not being driven by distinctly different "types" of districts (e.g., "high-performing" districts versus "disparity-narrowing" districts).

Qualitative Results

The results that follow are narrated in four parts, each a stage of the public policy process described in this study's conceptual framework. I begin by narrating how Tennessee's policymakers crafted a unified agenda leading up to the study period. I follow by describing how policy priorities were selected and legitimized in the earliest years of the study period, then how policies were implemented throughout the study period. I close by sharing the state's orientation toward ongoing evaluation. In the narrative, I report themes that were prominent across participants' interviews by providing the percentage of interviews in which that theme occurred. This is not meant to be representative of all the actors in Tennessee's policy space; rather, it is an indication of salience. Note, too, that percentages are reported based on the number of interviews and not the number of participants, as in two cases, there were multiple participants in a single interview.

Agenda-Setting

On February 28, 2007, the U.S. Chamber of Commerce released the first installment of its Leaders and Laggards report card series. Tennessee received three Fs: one in academic achievement of low-income and minority students, one in postsecondary and workforce readiness, and one in truth in advertising about student proficiency (U.S. Chamber of Commerce, 2007). This report meant students were considered proficient in Tennessee but would not have been in most other states (for a full list of categories and Tennessee's scores, see Appendix Table 3).

Governor Bredesen (D) was scarcely a month into his second term at the time the report was released, and the results helped set the course of education policy in Tennessee for the rest of his time in office. Tennessee's proficiency standard, in particular, was the lowest among all U.S. states, and among the categories assessed on the U.S. Chamber of Commerce report card, it was the "Truth in Advertising" rating that seemed to hit the hardest; in 67% (n=10) of interviews, participants mentioned this score on the report card as key to setting the course of education policy in Tennessee for years to come. Tennessee's leaders took that score at face value: They had not been honest with educators, parents, or students; Tennessee's students were being labeled "proficient" when they would not have been in most other states. This report card put academic standards at the top of the education policy agenda for nearly all policymakers.

Less than a year after Tennessee received an F for Truth in Advertising, in January 2008, the Tennessee Department of Education capitalized on the unified momentum around academic standards and established new graduation requirements. At the same time, former U.S. Senate Majority Leader Bill Frist (TN-R) concluded his term in the Senate and directed his efforts to the state's education policy space. In early 2009, Frist led a team, the State Collaborative on Reforming Education (SCORE), in interviewing researchers and policy experts and conducting more than 60 town hall meetings with policymakers, business leaders, educators, and parents. He hoped to write a "road map" for Tennessee's next steps in education. Through this work, Frist and SCORE crafted an education policy agenda that featured the voices of stakeholders from all levels of the education

system. Although Frist was only mentioned directly in 20% (*n*=3) of interviews, SCORE was mentioned in nearly all of them—and in 67% (*n*=10), participants described the non-profit, non-partisan advocacy work SCORE began in 2009 (and has continued since) as critical to building relationships across stakeholders and getting policymakers to share the same priorities.

Policy Formulation and Legitimation

In July 2009, as Frist's team put together its report, President Barack Obama announced the Race to the Top competition. Race to the Top incentivized states to design strategies for education reform in four areas: adopting standards and assessments for postsecondary success, recruiting and retaining effective teachers and principals, turning around low-achieving schools, and building data systems that facilitate students' academic success (ARRA, 2009). Three months later, Frist released his team's road map. Nine months of investigating Tennessee's education landscape resulted in the following four recommendations: embrace high standards, cultivate strong leaders, ensure excellent teachers, and utilize data to enhance student learning (Frist, 2009). There was significant overlap in Race to the Top's priorities and the needs Frist's team identified. This cemented the plans that Tennessee proposed in its Race to the Top application.

Through this process, Tennessee established the four policy priorities that carried through the next decade: standards, assessment, accountability, and teacher evaluation. Although there were other priorities participants discussed—response to intervention was mentioned in 13% (n=2) of interviews, for example—80% (n=12) of interviews included the mention of all four of these policy priorities, and 93% (n=14) included at least one. Across all themes that arose in interviews, the policy priorities that participants identified were the most consistent. This suggests both a unified agenda and shared language around policy formulation.

The federal government's explicit support of these policy priorities substantiated the policies' legitimacy. In January 2010, Tennessee submitted its Race to the Top application, capitalizing on the well-timed findings of Frist and his team's road map and the fortuitous interest convergence between the state and federal government's priorities for education. However, Tennessee's leadership did not wait to hear whether they won the grant before taking action. Governor Bredesen called a special session of the legislature that same month to pass the bipartisan First to the Top bill, which gave provisions for restructuring low-achieving schools, created a 15-person teacher evaluation advisory committee (which had to have at least five k-12 teachers), and established a teacher professional development fund, among other initiatives (Tennessee First to the Top Act, 2010). Two months later, in March 2010, Delaware and Tennessee became the first Race to the Top winners, receiving $100 million and $500 million, respectively (Tennessee State Government, 2010).

The $500 million Tennessee won through Race to the Top is a nontrivial amount by any standard and especially critical during the Great Recession. However, it was not central to participants' narratives. When participants discussed Race to the Top in interviews, they mentioned that the money was helpful to steadfastly pursue the state's policy priorities, but they emphasized that the state was on that trajectory anyway. It is important to note that the legislature passed First to the Top before legislators knew they would receive federal funds to support the bill's initiatives. In addition, after the funding ran out, Tennessee continued to "stay the course," with no rollbacks in policy or student achievement. This is not to make light of the considerable investment that the federal government made in Tennessee, only to say that funding alone was not enough, and it was not central to participants' narratives.

By passing First to the Top, Tennessee set the stage for change. Then, that fall, Tennessee elected Governor Bill Haslam (R). Participants described some initial uncertainty about whether Haslam would build on the foundation that Bredesen's administration had overseen, especially since the two governors were from different political parties. However, Bredesen and Haslam met

frequently as Haslam transitioned into office, and it soon became apparent that political differences did not take precedent over shared values of student achievement. In 53% (*n*=8) of interviews, participants mentioned the alignment of and consensus between state leaders as being particularly remarkable or important to the state's achievement gains and consistency in direction over the study period. Although this feature would likely be helpful in many states' policy landscapes, Tennessee's policy context makes it particularly important: Participants most frequently named the governor as the most prominent stakeholder voice. This occurred in 53% (*n*=8) of interviews (tied only with directors of schools, Tennessee's term for superintendents). Thus, the continuity across governors was particularly important given the governor's position as the most prominent stakeholder voice.

This consensus across leaders, even those from different political parties, allowed Tennessee to "stay the course" on policy priorities. This phrase was repeated in 73% (*n*=11) of interviews, demonstrating the consistency in language not only across policy priorities but in orientation to those priorities, as well. Key policy actors from across the state, from legislators to the SCORE advocacy organization to leaders in the Department of Education, may have disagreed on the way particular policies were implemented, but they all remained aligned on the policy priorities advanced in the First to the Top bill.

The policy priorities did not inspire as much tension as some of the policies formed to address them. After winning Race to the Top, in spring 2010, Tennessee pursued strong standards and assessments by joining other states in adopting the newly released Common Core standards and utilizing Common Core-aligned assessments designed by the Partnership for Assessment of Readiness for College and Careers (PARCC). By taking part in initiatives other states were pursuing, Tennessee was able to scaffold efforts at building a common language around standards and assessment. However, using standards and assessments that did not originate in Tennessee fostered tension in the state policy space. Common Core and PARCC were vestiges of the Bredesen administration, and Haslam withdrew from PARCC in 2014 only to return to TCAP, the Tennessee Comprehensive Assessment Program, for the state's assessments in the relevant grades and subjects. Around the same time, state leaders began a comprehensive review of the Common Core state standards; two years later, in 2016, the state approved the Tennessee Academic Standards, for implementation in the 2017-18 school year. I use these examples here, detailed to me in interviews, only to illustrate that consensus among state leaders did not necessarily mean consensus on what policy priorities looked like in practice. Rather, Tennessee's leadership "stayed the course" in terms of elevating standards, assessment, accountability, and teacher evaluation, and this philosophy guided the trial and error that accompanies policymaking.

From the First to the Top bill and the Race to the Top award, Tennessee had the consensus across leaders, the statewide momentum, and the resources to pursue strong standards, assessments, accountability, and teacher evaluation. In doing so, Tennessee adopted Common Core, re-evaluated its state assessments, and reimagined its teacher evaluation process. Additionally, as part of the No Child Left Behind waiver Tennessee received in 2012, the state adopted both a school-level and district-level accountability framework. In many cases, these reforms were not overly unique: Many states adopted Common Core, updated their teacher evaluation processes, and implemented new accountability frameworks over the study period, since these helped states meet fundamental provisions in federal initiatives. However, I argue here that the consistent focus on key policy priorities and the way education policies were implemented were just as critical for the state's remarkable achievement gains.

Policy Implementation

Theory suggests successful policy implementation includes scaffolding policy rollouts, allowing some level of flexibility and autonomy in implementing policies, effectively collecting and

disseminating institutional data, and fostering successful communication across different parts of an organization (Buck, 2006). Despite the obstacles that accompany shifting administrations, the state's prominent stakeholders, including Governor Haslam and newly appointed Commissioner of Education Kevin Huffman, rose to the challenge of turning policy into practice. Participants mentioned each of these strategies in some way.

In 60% (*n*=9) of interviews, participants mentioned that professional development to help teachers deliver content using the Common Core standards was essential. It was important that the state did not just bring in new standards, but also provided support for the transition so teachers knew what changes they needed to make in the classroom. Professional development was also built into the new teacher evaluation system, cementing its importance.

Professional development needs were in part identified using Tennessee's regional support model ("CORE"), a model participants cited in 73% (*n*=11) of interviews as critical to policy implementation and to the state department's relationship with school districts. Although Tennessee had regional field offices before 2012, the Tennessee Department of Education reimagined the CORE offices to better serve the new policy rollout. In doing so, the Department of Education and the CORE offices transitioned from being compliance-based to support-based, a change that participants mentioned in 60% (*n*=9) of interviews.

While the department shifted its orientation and practices from compliance-based to support-based, it also experienced a fundamental reorganization that brought together offices focused on similar workflows. Instead of being siloed, offices in which people were working on complementary priorities were consolidated. In 53% (*n*=8) of interviews, participants mentioned that reorganizing the department informed how initiatives were re-conceptualized and pursued. In discussing the department's change, participants were clear that it was not just a physical or logistical reorganization, but a philosophical one. Before, conversations had often started by discussing financial matters; now, conversations started with students.

In changing its orientation to be more support-focused, the department also created opportunities for districts to implement policies in ways that best met their own needs and contexts. This "structured flexibility," mentioned in 40% (*n*=6) of interviews, allowed the state to hold districts accountable while also giving districts oversight over their own practices. For example, districts did not have to adopt the statewide teacher evaluation model. Instead, they were given the option to adopt it or to develop their own using guidelines put together by the teacher evaluation advisory committee. In this way, districts were given some level of autonomy over the way policies made it to their classrooms, increasing district buy-in.

There were other philosophical changes that the Department of Education experienced over the study period, as well. After receiving an F for Truth in Advertising, Tennessee's education leaders developed an orientation around transparency, a value participants expressed was important to the state's education landscape in 40% (*n*=6) of interviews. In practice, this meant being transparent about data. State leaders, including the Department of Education, wanted stakeholders to know if students were proficient; they wanted to know if schools and districts had achievement disparities; they wanted teachers to have access to their students' achievement data so that teachers could pursue evidence-based practices in the classroom; and they wanted school leaders and district offices to use data to inform decisions about funding and programming (in 60%, *n*=9, of interviews, participants mentioned the importance of data-informed decision-making for teachers and school leaders). Participants articulated using data to shed light in corners that had previously been ignored and making sure that relevant stakeholders had access to data insights—that to fix a problem, they had to know it was there in the first place.

In addition to ensuring teachers and school district leaders had access to student performance data, the state Department of Education also regularly communicates with these

groups. In 67% (*n*=10) of interviews, participants mentioned at least one method of regularly communicating research, effective practices, or upcoming policies to educators and directors of schools (superintendents). For example, some divisions have newsletters they send to teachers or regular meetings with superintendents that both help solidify the connection between the state department and teachers in classrooms and help keep key stakeholders informed about important issues. This is critical for overcoming one of the greatest challenges of a loosely coupled system: the elaborate game of "telephone" that comes from poorly communicating policy priorities as they trickle through different levels of education's institutional structure.

Each of these changes was important to Tennessee's policy implementation, but they might all be for naught if not for the people. When I asked participants what made Tennessee unique or what Tennessee's story could not be told without, in 53% (*n*=8) of interviews, participants cited the people. A few mentioned how fortuitous the state has been to have had two consecutive governors from different political parties who placed such a high premium on educational inputs and outcomes; other participants mentioned the exceptionally driven education commissioners; and most mentioned the talent coming to the Department of Education, especially when it was dramatically expanded with Race to the Top funds.

Ongoing Policy Evaluation

A critical component of effective policy evaluation is evaluating the policy throughout the implementation process. One way Tennessee's leaders engaged in effective policy evaluation is through communicating with educators and directors of schools. Importantly, this communication was not a one-way street. Beginning in 2012, the Department of Education began administering a district and teacher survey to elicit feedback on policies and practices. One such policy was the new teacher evaluation model. In its first year, 2012, only 28% of teachers agreed that the teacher evaluation process led to improvements in student learning, and only 38% agreed that it led to improvements in their teaching; by 2016, those numbers had reached 66 and 71%, respectively (Bailey & Booker, 2018). Despite poor initial reception, the state department "stayed the course" in teacher evaluation and set out to improve the system rather than abandon it. Listening to teachers led to using teachers' growth scores in lieu of overall achievement scores if it would benefit the teacher, an impactful change for teachers making strides in low-performing schools. This is just one example. In 73% (*n*=11) of interviews, participants mentioned at least one (and often multiple) instance in which the Department of Education sought feedback from stakeholders in the policy process.

Finally, it is possible that Tennessee's story would have already come to a close if leaders did not maintain their interest in continued improvement. In 60% (*n*=9) of interviews, participants emphasized that while Tennessee has made gains, its student achievement and equity still fall short of the state's goals. Some districts may have narrowed achievement disparities, but disparities remain wide. Strides in data systems and teacher evaluation appear to have enhanced teacher and student performance, but these strides have also made it apparent that low-income students and students of color tend to have less-effective teachers. There is more work to do, then, to advance equity. In general, participants expressed an optimism that the state's leaders would continue to build on the momentum begun over a decade ago.

Probing the Narrative

Thus far, I have described Tennessee's achievement trends over the study period and narrated the process through which Tennessee set a unified agenda, formulated and legitimized policies, implemented policies, and engaged in evaluation. Importantly, I am unable to conclude the extent to which these features played a role in Tennessee's achievement patterns over the 2009 to

2015 period. Such an endeavor is challenging for a few reasons. First, many states implemented similar reforms to Tennessee over the study period, though in different ways and at different degrees of implementation fidelity; while the reforms themselves are quantifiable, the nuances of implementation are less so. Additionally, as detailed in the qualitative results section, many reforms in Tennessee occurred simultaneously. The Department of Education underwent organizational change, schools brought in new data systems, teachers endured new evaluation models, and the department began to evaluate districts with the new district accountability system all around the same period.

Despite these complications, it is possible to look at Tennessee's achievement patterns and consider how closely related they are to the timeline in policymakers' narratives. The state's reforms largely began with the First to the Top bill in January 2010, continued through NCLB waiver adoption and the implementation of the teacher evaluation system in 2012, and are driven by the same priorities today. Notably, most of Tennessee's improving achievement manifests in 2013 (see Table 2; Figure 3) and increases more slowly through the rest of the study period. This suggests that the reform story may explain the state's overall improving achievement.

In addition to considering whether the narrative here explains Tennessee's overall achievement trends, it is also worth considering whether it explains Tennessee's narrowing of disparities. Recall that disparity narrowing is negative, indicating that in 2009, Tennessee's socioeconomic disparity is wider than in the rest of the United States and its racial/ethnic disparities are statistically indistinguishable from those in the rest of the United States. The greatest change in the socioeconomic and White-Black disparities manifests in spring 2011—before most policies I describe in this study would have taken effect. These disparities narrow a little more in 2012, such that the socioeconomic disparity is no longer statistically different from the disparity in the rest of the United States and the White-Black disparity is distinctly narrower. These disparities then stagnate (for the White-Black disparity) or begin to widen (for the socioeconomic disparity). The White-Hispanic disparity has a slightly different trend, narrowing faster in Tennessee than in the rest of the United States over time. In 2009, it is no different from the disparity in the rest of the United States, and by 2015, it is distinctly narrower than in the rest of the United States. It is worth noting again that Tennessee's White-Hispanic disparity patterns were largely explained by the coarse policy indicators, shared by many other states; this, perhaps, explains the broader national trend toward White-Hispanic disparity narrowing.

This pattern casts some doubt on the idea that Tennessee's reforms drove narrowing socioeconomic and White-Black disparities. However, there are a few notable events that may help explain these achievement patterns. First, Tennessee raised the proficiency standard on its statewide assessment for the spring 2010 test administration, which may have spurred district-, school-, and classroom-level mobilization around Black students' and low-income students' achievement in particular. Another possibility is that the state's preparation for its reforms may have disproportionately impacted these students. For example, in the 2010-11 school year, Tennessee's department of education "provided introductory training to more than 4,000 educators statewide" on the Common Core standards and "trained more than 6,000 educators on the new teacher evaluation instrument" (U.S. Department of Education, 2012, p. 4). Although these reforms were not fully implemented until the following year, the preparation that took place may have had its strongest impact in classrooms serving historically disadvantaged students. It is also worth mentioning that Tennessee's achievement gains are driven by advantaged and disadvantaged students alike—so it may be the case that implementing the statewide reforms in the 2011-12 school year simply eclipsed the impacts of pre-existing efforts to support the achievement of historically disadvantaged students.

Table 2

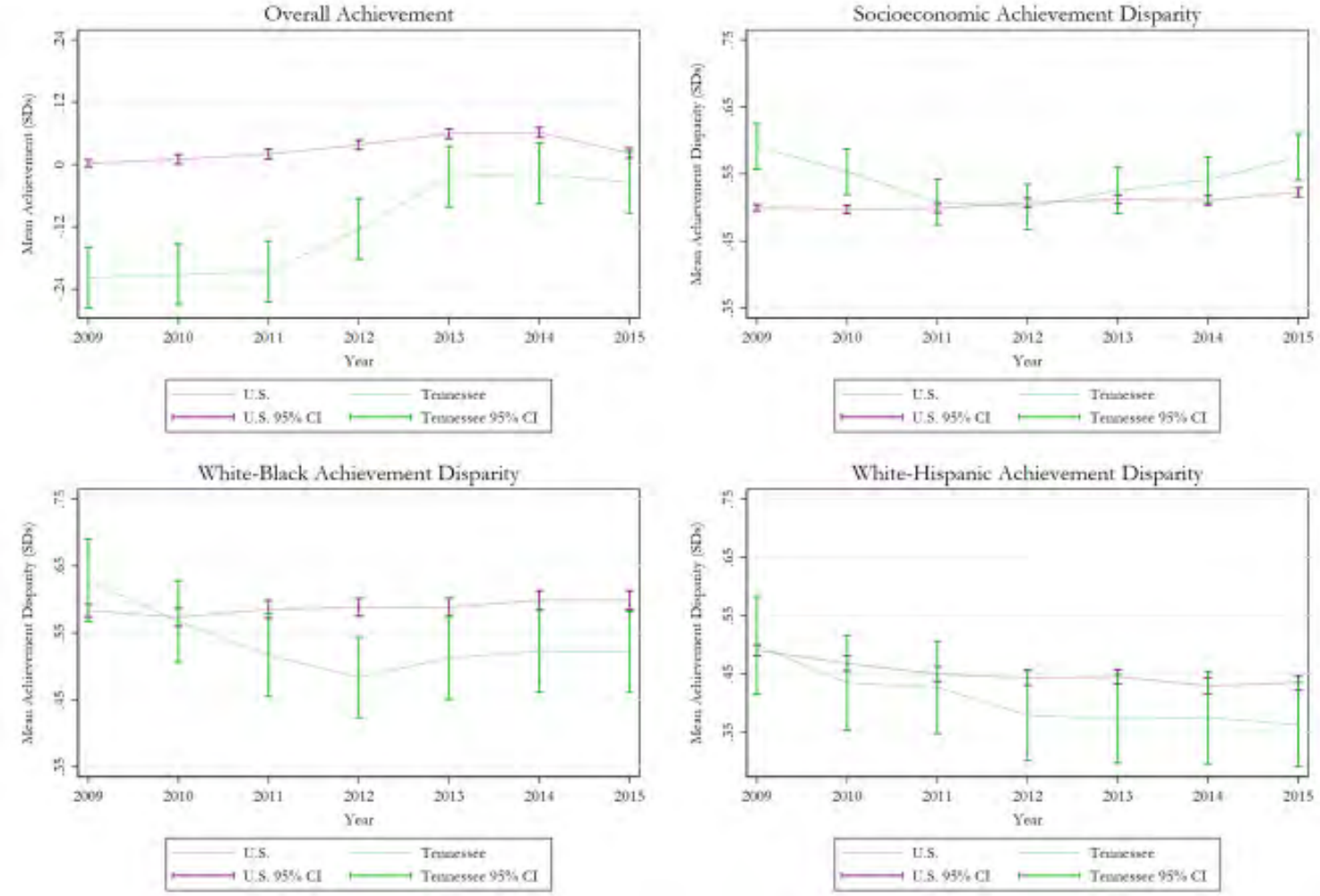
District-Level Achievement Trends Disaggregated by Year

	All	Socioeconomic	White-Black	White-Hispanic
TN x 2015	-0.054	0.054**	-0.079*	-0.076
	(0.030)	(0.018)	(0.031)	(0.037)
TN x 2014	-0.078**	0.031	-0.075*	-0.047
	(0.030)	(0.018)	(0.031)	(0.040)
TN x 2013	-0.080**	0.014	-0.078*	-0.072
	(0.030)	(0.018)	(0.031)	(0.039)
TN x 2012	-0.159***	-0.004	-0.107***	-0.064
	(0.030)	(0.017)	(0.030)	(0.040)
TN x 2011	-0.223***	0.010	-0.068*	-0.022
	(0.030)	(0.017)	(0.031)	(0.040)
TN x 2010	-0.220***	0.057**	-0.006	-0.034
	(0.030)	(0.017)	(0.031)	(0.042)
TN x 2009	-0.218***	0.091***	0.046	0.010
	(0.030)	(0.017)	(0.031)	(0.043)
2015	0.019***	0.022***	0.016*	-0.054***
	(0.005)	(0.003)	(0.007)	(0.007)
2014	0.061***	0.010**	0.014	-0.061***
	(0.005)	(0.003)	(0.007)	(0.007)
2013	0.057***	0.011***	0.005	-0.045***
	(0.005)	(0.003)	(0.007)	(0.006)
2012	0.035***	0.006	0.003	-0.048***
	(0.005)	(0.003)	(0.007)	(0.006)
2011	0.018***	-0.002	-0.000	-0.040***
	(0.005)	(0.003)	(0.007)	(0.006)
2010	0.008	-0.004	-0.011	-0.022***
	(0.005)	(0.003)	(0.007)	(0.007)
2009 (US Mean)	-0.054	0.054**	-0.079*	-0.076
	(0.030)	(0.018)	(0.031)	(0.037)
Adj. R-squared	0.67%	0.25%	0.25%	0.64%
Covariates	None	None	None	None
Observations	72,013	49,278	15,754	19,209

Note: This table displays the relationships over time between Tennessee and overall achievement as well as socioeconomic, White-Black, and White-Hispanic disparities. All Tennessee estimates are relative to the U.S. estimate in the respective year. For example, the first row indicates the difference in achievement in Tennessee in 2015 and average achievement elsewhere in the U.S. in 2015. This means that in Tennessee, in 2015, students were performing at -0.054 standard deviations lower than students elsewhere in the United States in 2015 (not significant). Standard errors are in parentheses. ǂ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Figure 3

Line Graphs of Achievement Over Time (Tennessee in Green)



Note: This figure exhibits the cross-year variation in Tennessee's trends in achievement and U.S. trends more broadly. The estimates and confidence intervals are from Table 2, the year disaggregation. Tennessee's confidence intervals are much larger than those for the rest of the United States due to the significant difference in sample size (the number of districts in Tennessee versus outside of Tennessee).

Discussion

As with most things, there is no magic formula to galvanizing achievement gains and narrowing achieving disparities. There is, however, a particular strategy to the policymaking process such that rapid statewide improvement is possible. The policy priorities that have taken center stage in Tennessee are many of the same priorities articulated in federal policy. This includes policies that constituted criteria for Race to the Top awards, which 18 states won, and those mandated by NCLB waivers, which 43 states received. However, no state has seen parallel gains, and those that have seen gains seem to have done so, at least in part, at the expense of structurally disadvantaged students. Tennessee offers an important contrasting case.

Because many of Tennessee's reforms were adopted in tandem, it is difficult to determine the extent to which the state's gains are attributable to specific features of its policy landscape. These features may even have interactive effects that are unobservable without cross-state variation. For example, it is possible that Tennessee's district-level student subgroup reforms were only effective when implemented along with the state's shift to support-based accountability. This study cannot adjudicate the effectiveness of specific implementation strategies.

What this study can do, however, is suggest mechanisms underlying effective policies and offer hypotheses for understanding heterogeneity in policy effectiveness. This is a crucial next step in education policy research. As Coburn (2016) argues, education policy research "suffers from the propensity to learn the same lessons over and over again. Leadership matters. Teachers' learning communities matter. Teacher quality matters. If we are to move the field forward, it is important to dig beneath what are now truisms to uncover the underlying mechanisms by which these factors matter" (p. 473). This study is an effort to address this call. In doing so, I set a foundation for causal work that might isolate the relationship between strategies described here and intended policy outcomes. Policymakers may also learn from this study, drawing on Tennessee's implementation strategies for their own work. This, too, can help us learn more about how these strategies interact with local context.

There are many features to Tennessee that were unique to the time and place of the study period and are not easily replicable. For example, the U.S. Chamber of Commerce report card catalyzed a statewide orientation toward improving student standards; this constitutes an "origin story" that led to setting ambitious goals. Further, in Tennessee, the commissioner of education and the state board of education are appointed by the governor, facilitating consensus across prominent actors. The combination of governors, commissioners, legislators, Department of Education employees, researchers, and others in the policymaking process cannot be replicated.

However, there are many other critical features that are transferrable to other contexts. For example, there is a specific organization, Tennessee SCORE, whose very job it is to convene stakeholders who might otherwise have little chance to interact or mutually participate in agenda-setting. Additionally, maintaining similar messaging and staying the course on policy priorities allowed leaders to make productive changes to policies without abandoning them and invited other stakeholders to engage throughout the policymaking process. When pursuing policy change, many participants mentioned there was some form of eliciting public feedback, often through town halls held throughout the state. The Department of Education experienced important reorganizations, too, in which roles became centered around responsibilities, the conversations became centered around students (rather than concerns like finances and classroom conditions), and the philosophy transitioned from compliance-based to support-based. The department scaffolded policy transitions for school districts and began to use its regional model differently, with regions focused on identifying and meeting district needs rather than punishing schools and districts for not meeting expectations. These are important pieces of Tennessee's story.

This study has a few limitations. Quantitatively, I only coarsely control for relevant policies over the study period. This offers limited variation since many states implemented similar reforms to Tennessee and many reforms were implemented over a similar timeframe. I attempt to overcome this limitation by describing the process through which Tennessee implemented policy reforms in the qualitative results. On the other hand, there were some policies in which Tennessee was the only or one of the only places implementing that reform (e.g., the district-level accountability framework, only implemented in three states). In these cases, it is challenging to assess the impact of a particular policy due to the timing (how it occurred not in isolation but with a group of other policies at the same time) and the "selection effects." That is, nothing here was randomly assigned. Tennessee very much "selected into" or chose its interventions. This means that at every analytical turn, it is difficult to differentiate between the effects of specific state policies and the impacts of other features, like Tennessee's ripeness for reform, its hunger for improvement, and its newfound commitment to "truth in advertising."

There are some limitations qualitatively, as well. Although the story my interviewees narrated was particularly consistent (often drawing on identical language, like "compliance-based to support-based" and "staying the course"), I draw on a small sample size (22 participants over 15 interviews). I argue that due to linguistic consistency and the small population (state-level education policy actors in Tennessee between 2009 and 2015), this sample size is sufficient for an initial investigation of Tennessee's unique trends. Further, this sample shows the perspective of policymakers and policy implementers at the state level, since this is the story of statewide change, but the narrative may be different from the perspectives of teachers, parents, and students. These actors are absent from this sample. Nonetheless, my participants are core to the Tennessee reform story. Finally, I do not focus extensively on comparison cases, such as Washington DC, South Carolina, or Delaware. Future work might leverage both quantitative and qualitative methods to determine the degree of similarity (or difference) between policy implementations in these places, as well as how these differences relate to achievement trends.

In this study, I explicate the conceptual policymaking process using a particular state's approach to policymaking in a critical period of achievement gains. Because state policy has such a profound impact on students' experiences, it is essential to understand how states pursue policy; Tennessee offers an important case. Further, in this study, my iterative analytic approach—using the quantitative findings to inform my qualitative interview protocol and using the narrative that emerged to return to the quantitative results with a critical eye toward the timeline—offers an example of how mixed methods can inform education policy. Finally, by explicating the case of Tennessee, this study identifies features of state policy contexts that may facilitate achievement growth while narrowing achievement disparities. Future work might isolate the impact of these features using quasi-experimental methods. In addition, this study only considers single subgroups of students, such as Hispanic students or economically disadvantaged students. Future studies might consider intersectional analyses to understand if socioeconomic disparities are primarily widening among White students, for example, or if White-Black disparities are primarily widening among non-poor students.

Acknowledgements

In bringing this work to publication, I appreciate the thoughtful comments from Stanford University's Advanced Topics in Quantitative Policy Analysis workshop, as well as from Ilana M. Horwitz, sean f. reardon, David Labaree, Tom Dee, Carrie Townley-Flores, Marissa E. Thompson, Sonia Giebel, and Crystal Moore. I also appreciate feedback on earlier versions of this work received

through presenting at the annual meetings for the Association for Education Finance and Policy (AEFP) and the Association for Public Policy Analysis and Management (APPAM).

References

Aldrich, M. W. (2017, December). *Here's why people are talking about Tennessee, a 'bright green rectangle' on a new U.S. map of student growth.* Chalkbeat Tennessee. https://tn.chalkbeat.org/2017/12/20/21104039/here-s-why-people-are-talking-about-tennessee-a-bright-green-rectangle-on-a-new-u-s-map-of-student-g

American Recovery and Reinvestment Act of 2009 [ARRA], Pub. L. No. 16 U.S.C. § 111-5 (2009).

Atteberry, A., Bischoff, K., & Owens, A. (2021). Identifying progress toward ethnoracial achievement equity across U.S. school districts: A new approach. *Journal of Research on Educational Effectiveness, 14*(2), 410–441. https://doi.org/10.1080/19345747.2020.1868032

Ayers, J., & Owen, I. (2012*). No Child Left Behind waivers: Promising ideas from second round applications* (Issue July). https://cdn.americanprogress.org/wp-content/uploads/issues/2012/07/pdf/nochildwaivers.pdf?_ga=2.182613542.1415543435.1645578849-648075419.1642899047

Bailey, I., & Booker, L. (2018). *Reflections over time: Tennessee education survey 2018 results in context.* https://www.tn.gov/content/dam/tn/education/data/data_survey_report_2018.pdf

Bonilla, S., & Dee, T. S. (2020). The effects of school reform under NCLB waivers: Evidence from focus schools in Kentucky. *Education Finance and Policy, 15*(1), 75–103. https://doi.org/10.1162/edfp_a_00275

Buck, S. J. (2006). *Understanding environmental administration and law* (3rd ed.). Island Press.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*(2), 349–399. https://doi.org/10.1086/421787

Camera, L. (2017, December). Tennessee nails student growth. *U.S. News & World Report*. https://www.usnews.com/news/best-states/articles/2017-12-22/tennessee-nails-student-growth

Chesnut, C. E., Hitchcock, J. H., & Onwuegbuzie, A. J. (2018). Using mixed methods to inform education policy research. In C. R. Lochmiller (Ed.), *Complementary research methods for educational leadership and policy studies* (pp. 307–324). Palgrave MacMillan. https://doi.org/10.1007/978-3-319-93539-3_15

Coburn, C. E. (2016). What's policy got to do with it? How the structure-agency debate can illuminate policy implementation. *American Journal of Education, 122*(3), 465–475. https://doi.org/10.1086/685847

Cortazzi, M. (1993). *Narrative analysis* (1st ed.). Routledge.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications, Inc.

Dee, T. S., & Dizon-Ross, E. (2019). School performance, accountability, and waiver reforms: Evidence from Louisiana. *Educational Evaluation and Policy Analysis, 41*(3), 316–349. https://doi.org/10.3102/0162373719849944

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020). *COVID-19 and learning loss—disparities grow and students need help*. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19-and-learning-loss-disparities-grow-and-students-need-help

Drum, K. (2017, December). *Lots of poor school districts do pretty well, but we don't know why*. Mother Jones. https://www.motherjones.com/kevin-drum/2017/12/lots-of-poor-school-districts-do-pretty-well-but-we-dont-know-why/

Elmore, R. F., & McLaughlin, M. W. (1988). *Steady work: Policy, practice, and the reform of American education.* The RAND Corporation. https://www.rand.org/pubs/reports/R3574.html

Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2021). *Stanford Education Data Archive technical documentation* (version 4.1). http://purl.stanford.edu/db586ns4974

Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook in economics* (Vol. 3, pp. 343–421). North-Holland. https://doi.org/10.1016/B978-0-444-53429-3.00008-9

Frist, B. (2009). *A roadmap to success: A plan to make Tennessee schools #1 in the Southeast within five years* (Issue October). https://tnscore.org/wp-content/uploads/2018/09/Roadmap-to-Success_VisionReport2009.pdf

Greene, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Henry, G. T., Pham, L., Guthrie, J. E., & Harbatkin, E. (2018). *Guiding principles for improving the lowest-performing schools in Tennessee*. https://peabody.vanderbilt.edu/TERA/Guiding_Principles_Low_Performing_Schools_FINAL.pdf

Howell, W. G. (2015). Results of President Obama's Race to the Top. *Education Next, 15*(4). https://www.educationnext.org/results-president-obama-race-to-the-top-reform/

Hunter, S. B. (2020). The unintended effects of policy-assigned teacher observations: Examining the validity of observation scores. *AERA Open, 6*(2). https://doi.org/10.1177/2332858420929276

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics, 131*(1), 157–218. https://doi.org/10.1093/qje/qjv036

Jang, H., & Reardon, S. F. (2019). States as sites of educational (in)equality: State contexts and the socioeconomic achievement gradient. *AERA Open, 5*(3). https://doi.org/10.1177/2332858419872459

Jones, E. (2014). *Tennessee's No Child Left Behind waiver*. https://comptroller.tn.gov/office-functions/research-and-education-accountability/publications/prek-12/tennessee-s-no-child-left-behind-waiver.html

Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal, 34*(1), 39–81. https://doi.org/10.3102/00028312034001039

Labaree, D. F. (2010). *Someone has to fail: The zero-sum game of public schooling.* Harvard University Press.

Matheny, K. T., Thompson, M. E., Townley-Flores, C., & Reardon, S. F. (2023). Uneven progress: Recent trends in academic performance among U.S. school districts. *American Educational Research Journa*l. https://doi.org/10.3102/00028312221134769

Opper, I. (2019). *Teachers matter: Understanding teachers' impact on student achievement.* RAND. https://doi.org/10.7249/rr4312

Paufler, N. A. (2018). Declining morale, diminishing autonomy, and decreasing value: Principal reflections on a high-stakes teacher evaluation system. *International Journal of Education Policy and Leadership, 13*(8). https://doi.org/10.22230/ijepl.2018v13n8a813

Pham, L. D., Henry, G. T., Kho, A., & Zimmer, R. (2020). Sustainability and maturation of school turnaround: A multiyear evaluation of Tennessee's Achievement School District and Llcal Innovation Zones. *AERA Open, 6*(2). https://doi.org/10.1177/2332858420922841

Plummer, K. (1983). *Documents of life: An introduction to the problems and literature of the humanistic method.* G. Allen & Unwin.

Podesta, K. (2019). *Students in Tennessee instructed by consecutive ineffective teachers.* https://comptroller.tn.gov/office-functions/research-and-education-accountability/publications/prek-12/student-placement.html#:~:text=After being taught by ineffective,and inclusion in identified subgroups.

Podolny, J. M. (2001). Networks as the pipes and prisms of the market. *American Journal of Sociology, 107*(1), 33–60. https://doi.org/10.1086/323038

Reardon, S. F. (2019). *Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age.* The Russell Sage Foundation Journal of the Social Sciences, 5(2), 40–68. https://doi.org/10.7758/RSF.2019.5.2.03

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics, 46*(2), 138–167. https://doi.org/10.3102/1076998619874089

Rodriguez, L. A., Swain, W. A., & Springer, M. G. (2020). Sorting through performance evaluations: The influence of performance evaluation reform on teacher attrition and mobility. *American Educational Research Journal, 57*(6), 2339–2377. https://doi.org/10.3102/0002831220910989

Roza, M. (2010). *Educational economics.* The Urban Institute Press.

Saldaña, J. (2016). *The coding manual for qualitative researchers.* SAGE.

Small, M. L. (2009). "How many cases do I need?": On science and the logic of case selection in field-based research. *Ethnography, 10*(1), 5–38. https://doi.org/10.1177/1466138108099586

Spencer, G., & Maldonado, M. (2021). Determinants of dual enrollment access: A national examination of institutional context and state policies. *AERA Open, 7*(1), 1–18. https://doi.org/10.1177/23328584211041628

Stockard, S. (2018, March 23). Parkinson to introduce bill phasing out state's Achievement School District. *The Daily News, 133*(60). https://www.memphisdailynews.com/news/2018/mar/23/parkinson-to-introduce-bill-phasing-out-states-achievement-school-district/

Tatter, G. (2016, April). *Tennessee fires TNReady testmaker, suspends tests for grades 3-8*. Chalkbeat Tennessee. https://tn.chalkbeat.org/2016/4/27/21098160/tennessee-fires-tnready-testmaker-suspends-tests-for-grades-3-8

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628–3651. https://doi.org/10.1257/aer.102.7.3628

Tennessee Department of Education. (2012). *Teacher evaluation in Tennessee: A report on year 1 implementation* (Issue July). https://files.eric.ed.gov/fulltext/ED533726.pdf

Tennessee Department of Education. (2017). *2017 school accountability protocol.* https://www.tn.gov/content/dam/tn/education/data/data_school_accountability_protocol_2017.pdf

Tennessee First to the Top Act of 2010 (2010). https://wapp.capitol.tn.gov/apps/BillInfo/default.aspx?BillNumber=HB7010&GA=106

Tennessee SCORE. (2017). *Understanding TVAAS: Student growth data and Tennessee's teacher evaluation system* (Issue March). https://tnscore.org/wp-content/uploads/2018/10/TVAAS_PolicymakerBrief2017.pdf

Tennessee State Government. (2010). *Tennessee wins Race to the Top grant.* [Press Release]. https://www.tn.gov/news/2010/3/29/tennessee-wins-race-to-the-top-grant.html#:~:text=According to the U.S. Department,Race to the Top reforms

Tseng, V. (2012). The uses of research in policy and practice. *Social Policy Report, 26*(2), 1–24. https://doi.org/10.1002/j.2379-3988.2012.tb00071.x

U.S. Chamber of Commerce. (2007). *Leaders and laggards: A state-by-state report card on educational effectiveness*. https://www.uschamberfoundation.org/leaders-and-laggards/#/2007/tn/state-profile

U.S. Department of Education. (2012). *Race to the Top Tennessee report, year 1: School year 2010-2011*. https://files.eric.ed.gov/fulltext/ED529330.pdf

U.S. Department of Education. (2016). *ESEA Flexibility*. https://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education. (2019a). *State performance compared to the nation: Data table*. https://www.nationsreportcard.gov/profiles/stateprofile?chort=1&sub=MAT&sj=&sfj=NP&st=MN&year=2017R3

U.S. Department of Education. (2019b). *The nation's report card: State profiles*. https://www.nationsreportcard.gov/profiles/stateprofile/overview/TN?cti=PgTab_GapComparisons&chort=2&sub=RED&sj=TN&fs=Grade&st=MN&year=2003R3&sg=Race%2FEthnicity%3A+White+vs.+Black&sgv=Difference&ts=Cross-Year&tss=2019R3-2003R3&sfj=NP

Weber, M. (2019). *Economy and society* (K. Tribe, Ed.). Harvard University Press. https://doi.org/10.4159/9780674240827

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*(1), 1. https://doi.org/10.2307/2391875

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*(1), 122–154. https://doi.org/10.1002/pam.20310

Yin, R. K. (2002). *Case study research: Design and methods*. SAGE Publications.

Zimmer, R., Henry, G. T., & Kho, A. (2017). The effects of school turnaround in Tennessee's Achievement School District and Innovation Zones. *Educational Evaluation and Policy Analysis, 39*(4), 670–696. https://doi.org/10.3102/0162373717705729

Zubrzycki, J. (2013, November). *Tennessee students lead the nation in growth on NAEP*. Chalkbeat Tennessee. https://tn.chalkbeat.org/2013/11/7/21091461/tennessee-students-lead-the-nation-in-growth-on-naep

About the Author

Kaylee T. Matheny
Stanford University
kmatheny@stanford.edu
https://orcid.org/0000-0002-9939-4155
Kaylee T. Matheny is a PhD candidate at the Stanford Graduate School of Education, an incoming postdoctoral fellow at Johns Hopkins University, and an incoming assistant professor of public policy at Georgetown University. Her research uses mixed methods to understand socioeconomic inequality in educational opportunities.

education policy analysis archives

Volume 31 Number 43 April 25, 2023 ISSN 1068-2341

Appendix

Appendix A. Estimating District-Level Achievement Trends

Step 1. Estimating the Annual District-Level Trend

$$A_{dgyb} = \beta_0 + \beta_1(grade_{dgyb} - 5.5) + \beta_2(year_{dgyb} - 2012) + \beta_3(math_{dgyb} - 0.5) + u_{dgyb} + \varepsilon_{dgyb}$$

$$\beta_0 = \gamma_{00} + v_{0d}$$
$$\beta_1 = \gamma_{00} + v_{1d}$$
$$\beta_2 = \gamma_{20} + v_{2d}$$
$$\beta_3 = \gamma_{30} + v_{3d}$$

$$u_{dygb} \sim N(0, \omega^2_{dygb})$$
$$\varepsilon_{dygb} \sim N(0, \sigma^2_{dygb})$$
$$[v_{0d}, v_{1d}, v_{2d}, v_{3d}] \sim MVN(0, \boldsymbol{\tau})$$

The parameter of interest in step one is β_2, the slope of the relationship between years $(year_{dgyb} - 2012)$ and student achievement (A_{dgyb}) (e.g., how the third graders in a particular district in 2015 perform on the state's standardized math test compared to the third graders in 2014; note that the estimate here pools observations across grades, years, and subjects, so it is the *average* change in the *average* grade in the *average* subject). If successive cohorts of students in a particular district earn higher scores on average over time, this number will be positive; if they get lower scores on average over time, this number will be negative. A_{dgyb} represents SEDA's estimate of the average student achievement for a given district-grade-year-subject, in standard deviations; $grade_{dgyb} - 5.5$ is the centered grade (between 3rd and 8th grades); $year_{dgyb} - 2012$ is the centered year (between 2009 and 2015); and $math_{dgyb} - 0.5$ is the centered subject. ω^2_{dygb} is a known parameter; σ^2_{dygb} and $\boldsymbol{\tau}$ are estimated.

For step two, I use the estimate of β_2 (the estimated annual district-level trend) from step one as the outcome. For simplicity's sake, due to the number of Greek letters and subscripts, the nestedness of the models, and in the interest of general readability, I refer to this as $\hat{\alpha}_{ds}$ in Step 2 (the estimate of a district *d*'s annual achievement trend in state *s*). I use the variance of the district-level residuals (ψ^2_{ds}), produced in the $\boldsymbol{\tau}$ matrix estimated in step one, for the precision-weighting in step two.

Trends in achievement gaps are estimated similarly, except I use three steps to estimate them instead of two. In these cases, the first model estimates the average achievement gap, nesting state-district-grade-year-subject-subgroup observations within state-district-grade-year-subject observations; subsequent models are identical to those above.

One potential source of concern for the quantitative analysis is that Tennessee changed its state proficiency cut-off after 2009. However, this does not impact analyses in this study due to how SEDA estimates are first standardized within each state-subject-grade-year (Fahle et al., 2021), putting all estimates on a common scale within each year (and state, subject, and grade). State-subject-grade-year distributions were then placed on the common NAEP scale in the years for which there is NAEP data (2009, 2011, 2013, 2015) for interpretation. Additionally, because Tennessee raised the proficiency level, we might expect the number of students estimated to be proficient to decrease (if all else was equal). Even if this change in proficiency did impact estimates here, then, this impact would bias downward Tennessee's achievement gains.

Appendix B. Testing the Ecological Fallacy

In addition to identifying broad trends in achievement and disparity narrowing, it is also important to assess whether districts in which students experience increasing achievement also those in which students experience achievement disparity narrowing. I gain traction on this question by identifying pairwise correlations between districts' improving achievement and achievement disparity closures for districts both inside and outside of Tennessee. If the districts with increasing achievement also experience increasing achievement disparities, this would yield a positive correlation; if districts with increasing achievement are experiencing declining achievement disparities (the more optimal outcome), this would yield a negative correlation. Appendix Table 4 reports these results.

In Tennessee, there is no significant correlation between achievement and disparity narrowing. This suggests that Tennessee's achievement trends are not being driven by distinctly different "types" of districts (e.g., "high-performing" districts versus "disparity-narrowing" districts). Put another way, whether a district is high performing appears not to have a relationship with whether that district is closing disparities. This indicates that districts can work toward narrowing achievement disparities without doing so at the expense of increasing average achievement. In the rest of the United States, correlations between racial/ethnic disparity trends and overall achievement trends suggest there is a small but significant relationship such that districts that tend to increase achievement also tend to decrease disparities. This is even more promising.

Appendix Table 1

Descriptive Statistics of Covariates and Outcomes

	Mean	SD	Min	Max	Count
Outcomes					
Overall Trend	0.009	0.048	-0.942	2.007	11,317
Socioeconomic Disparity Trend	0.004	0.055	-1.022	1.144	8,297
White-Black Disparity Trend	0.003	0.046	-0.383	0.629	2,665
White-Hispanic Disparity Trend	-0.007	0.050	-0.611	0.630	3,455
Predictor of Interest					
Tennessee	0.012	0.108	0	1	11,317
District-Level Demographic Variables					
Mean Achievement	-0.000	0.329	-2.013	1.215	11,317
Mean Socioeconomic Achievement Disparity	0.485	0.174	-0.782	1.596	8,787
Mean White-Black Disparity	0.570	0.209	0.027	1.670	2,872
Mean White-Hispanic Disparity	0.460	0.216	-0.598	1.580	3,794
Neighborhood SES	0.100	0.906	-4.398	2.936	11,317
Change in Neighborhood SES	-0.098	0.338	-2.062	1.291	11,317
% Poor Enrollment	0.477	0.219	0.000	1.000	11,317
Change in % Poor Enrollment	0.011	0.016	-0.144	0.160	11,317
% Asian Enrollment	0.022	0.049	0.000	0.730	11,317
Change in % Asian Enrollment	0.000	0.003	-0.042	0.036	11,317
% Black Enrollment	0.079	0.163	0.000	0.998	11,317
Change in % Black Enrollment	-0.001	0.004	-0.075	0.051	11,317
% Hispanic Enrollment	0.134	0.203	0.000	0.999	11,317
Change in % Hispanic Enrollment	0.005	0.007	-0.090	0.100	11,317
State-Level Demographic Variables					
State Mean Achievement	-0.025	0.206	-0.415	0.470	48
State SES Disparity	0.498	0.065	0.342	0.655	48
State White-Black Disparity	0.577	0.067	0.415	0.776	48
State White-Hispanic Disparity	0.469	0.067	0.299	0.623	48
State SES	0.017	0.526	-1.170	0.964	48
State SES Change	-0.112	0.104	-0.325	0.189	48
State % Poor Enrollment	0.497	0.132	0.201	0.762	48
State % Poor Enrollment Change	0.012	0.010	-0.007	0.043	48
State % Asian Enrollment	0.017	0.014	0.003	0.067	48
State % Asian Enrollment Change	-0.000	0.001	-0.001	0.002	48
State % Black Enrollment	0.106	0.140	0.004	0.565	48
State % Black Enrollment Change	-0.001	0.001	-0.007	0.001	48
State % Hispanic Enrollment	0.112	0.115	0.009	0.549	48
State % Hispanic Enrollment Change	0.005	0.003	-0.001	0.011	48
State-Level Policy Variables					
State Mean Per Pupil Expenditures (in $100s)	89.473	22.868	58.109	178.722	48
State Per Pupil Expenditure Change (in $100s)	-0.062	1.246	-1.819	3.241	48
State Won Race to the Top Round 1	0.042	0.202	0	1	48
State Won Race to the Top Round 2	0.167	0.377	0	1	48
State Won Race to the Top Round 3	0.146	0.357	0	1	48
State Mean Per Pupil Race to the Top Award (in $100s)	1.754	3.714	0	17.144	48
State Received NCLB Waiver	0.875	0.334	0	1	48

Note: Trend estimates reflect the annual estimated change in achievement between spring 2009 and spring 2015. These estimates were assessed in step one of the modeling strategy. See Appendix A.

Appendix Table 2

Policy Indicators by State, 2009-2015

	Annual PP Expenditures (in $100s)	Change in PP Expenditures (in $100s)	Won RTT Round 1	Won RTT Round 2	Won RTT in Round 3	RTT PP Award (in $100s)	Received NCLB Waiver
Alabama	69.56	-0.83	0	0	0	0	1
Alaska	178.72	2.95	0	0	0	0	1
Arizona	78.13	-0.42	0	0	1	0.50	1
Arkansas	76.42	-0.15	0	0	0	0	1
California	73.03	-0.62	0	0	0	0	0
Colorado	85.25	-0.20	0	0	1	0.47	1
Connecticut	109.65	2.90	0	0	0	0	1
Delaware	87.85	-0.25	1	0	0	17.14	1
Florida	71.83	-1.12	0	1	0	5.79	1
Georgia	71.52	-1.24	0	1	0	5.17	1
Idaho	74.13	-1.44	0	0	0	0	1
Illinois	75.99	1.46	0	0	1	0.47	1
Indiana	70.16	-0.65	0	0	0	0	1
Iowa	81.30	0.51	0	0	0	0	0
Kansas	94.70	-0.63	0	0	0	0	1
Kentucky	74.59	-0.76	0	0	1	0.56	1
Louisiana	84.89	-1.08	0	0	1	0.54	1
Maine	121.85	-0.62	0	0	0	0	1
Maryland	88.60	-0.85	0	1	0	6.72	1
Massachusetts	92.23	0.47	0	1	0	5.84	1
Michigan	74.39	-0.49	0	0	0	0	1
Minnesota	84.72	-0.20	0	0	0	0	1
Mississippi	67.52	-0.64	0	0	0	0	1
Missouri	79.08	-0.42	0	0	0	0	1
Montana	115.84	-0.53	0	0	0	0	0
Nebraska	111.24	2.74	0	0	0	0	0
Nevada	94.68	-1.82	0	0	0	0	1
New Hampshire	117.84	2.64	0	0	0	0	1
New Jersey	98.67	0.39	0	0	1	0.64	1
New Mexico	109.46	0.01	0	0	0	0	1
New York	134.79	1.97	0	1	0	5.86	1
North Carolina	69.29	-1.46	0	1	0	5.74	1
North Dakota	112.89	3.24	0	0	0	0	0
Ohio	72.69	-0.25	0	1	0	5.10	1
Oklahoma	73.14	-0.76	0	0	0	0	1
Oregon	95.43	-0.86	0	0	0	0	1
Pennsylvania	87.75	1.08	0	0	1	0.52	1
Rhode Island	101.14	0.11	0	1	0	11.85	1
South Carolina	71.57	-0.52	0	0	0	0	1
South Dakota	85.42	0.35	0	0	0	0	1
Tennessee	63.63	-0.06	1	0	0	11.25	1
Texas	78.83	-1.45	0	0	0	0	1
Utah	58.11	-0.53	0	0	0	0	1
Virginia	71.89	-0.77	0	0	0	0	1
Washington	88.36	-0.25	0	0	0	0	1

	Annual PP Expenditures (in $100s)	Change in PP Expenditures (in $100s)	Won RTT Round 1	Won RTT Round 2	Won RTT in Round 3	RTT PP Award (in $100s)	Received NCLB Waiver
West Virginia	87.63	-1.16	0	0	0	0	1
Wisconsin	84.87	-0.82	0	0	0	0	1
Wyoming	143.41	0.05	0	0	0	0	0

Note: This table includes each state's observations of the coarse policy indicators included in this study. For instance, column 4 includes an indicator for whether a state won the first round of Race to the Top. The annual per pupil expenditures (in $100s) are the mean per pupil state expenditures over the study period, and the change in per pupil expenditures is the annual change in per pupil expenditures (in $100s). The column "RTT PP Award" indicates the Race to the Top award amount per pupil, using the number of students enrolled in the year following the award. For example, Race to the Top awards were announced in March 2010, so I use the student enrollment for the 2010-11 school year to estimate the per pupil allocation.

Appendix Table 3

U.S. Chamber of Commerce Report Card, Tennessee, 2007 & 2014

	2007	**2014**
Data Quality	B	A
Truth in Advertising About Student Proficiency	F	A
21st Century Teaching Force	B	B
Return on Investment	C	C
Postsecondary and Workforce Readiness	F	C
Academic Achievement	D	D
Academic Achievement of Low-Income and Minority Students	F	D
Rigor of Standards	C	N/A
Flexibility in Management and Policy	C	N/A

Note: Rigor of standards and flexibility in management and policy were not rated in 2014 nor in any other year that the report has run. The most recent year for which grades were released for the corresponding areas is 2014.

Appendix Table 4

Pairwise Correlations of Improving Achievement and Disparity Narrowing

Disparity	Tennessee	US
Socioeconomic	-0.026	-0.008
(N)	130	8,258
White-Black	0.108	-0.043*
(N)	64	2,625
White-Hispanic	0.184	-0.045**
(N)	47	3,442

Note: Results in this table show the extent to which each disparity (socioeconomic, White-Black, and White-Hispanic) is correlated with overall achievement improving over time in Tennessee and the US broadly (without Tennessee). For example, the first column of correlations shows that in Tennessee, there is a non-significant correlation of -0.026 between districts' socioeconomic disparity trends and overall achievement trends. ǂ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$