EdWorkingPaper No. 24-989

Democratic Policymaking in Schools: The Influence of Teacher Empowerment on Student Achievement

Sara R. Sands
University of Houston

Despite the popularity of teacher leadership since the 1980s, little research examines its effects on student achievement. In this paper, I assess the influence of the New York City Department of Education's Teacher Career Pathways program, a teacher leadership initiative, on student achievement in grades three through eight. Using difference-in-difference approaches, including new event study estimators, I find that where school leaders staffed teacher leaders into formal roles with defined responsibilities, positional authority, and commensurate salary increases, student achievement in ELA and math improves. Moreover, the improvement in scores compounds over time, with schools exhibiting increasing gains in each year following the initial introduction of teacher leaders. Schools that do not staff teacher leaders do not observe similar outcomes. I consider these results in the context of democratic policymaking and teacher empowerment, suggesting that teachers must be formally empowered in schools to lead meaningful changes that ultimately improve student achievement.

VERSION: July 2024

Democratic Policymaking in Schools: The Influence of Teacher Empowerment on Student Achievement

Sara R. Sands

University of Houston

Abstract

Despite the popularity of teacher leadership since the 1980s, little research examines its effects on student achievement. In this paper, I assess the influence of the New York City Department of Education's Teacher Career Pathways program, a teacher leadership initiative, on student achievement in grades three through eight. Using difference-in-difference approaches, including new event study estimators, I find that where school leaders staffed teacher leaders into formal roles with defined responsibilities, positional authority, and commensurate salary increases, student achievement in ELA and math improves. Moreover, the improvement in scores compounds over time, with schools exhibiting increasing gains in each year following the initial introduction of teacher leaders. Schools that do not staff teacher leaders do not observe similar outcomes. I consider these results in the context of democratic policymaking and teacher empowerment, suggesting that teachers must be formally empowered in schools to lead meaningful changes that ultimately improve student achievement.

Introduction

As the gatekeepers of their classrooms, teachers are active makers of policy in practice. This includes interpreting the meaning of policy, carrying out its demands, sharing their thoughts and practices on policies with peers and school leaders, and bridging their professional work to policies where they find it is copacetic or beneficial to their practice and buffering against policies when that connection is not seen (Gallagher, 2008; Honig & Hatch, 2004). Given the vital role teachers play in dictating policy at the "street-level" (Lipsky, 1971), democratic policymaking in the context of education posits that for education policy to improve, teachers should be actively involved as authentic partners in its design and implementation from the outset (Good et al., 2017). It suggests that decisions made in schools are strengthened and more likely to be implemented when teachers are engaged, as these decisions benefit from the ideas, abilities, and experiences of the whole staff (Woods & Gronn, 2009). In short, by applying democratic practices and distributing power to teachers, schools might enhance their organizational capacity, including "increased responsiveness and sustained improvement" and greater organizational capability to deal with challenges of complexity in their work (Woods & Gronn, 2009, p. 438).

This paper is the second of three in a broader study exploring how the New York City Department of Education (NYCDOE) attempted to institute and scale distributed leadership practices and improve student achievement through Teacher Career Pathways (TCP), a teacher leadership program. In this quantitative study employing traditional difference-in-difference and new event study estimators, I answer the question, "Do schools participating in TCP observe improvement of student achievement scores over time that differ from schools that do not participate in TCP?" As part of my analysis, I apply theories of democratic participation in policymaking and teacher empowerment to better conceptualize how teacher leadership may foster more democratization in schools, which can in turn influence student achievement. The findings in this paper support claims that the formalization of teacher leadership roles in schools can be linked to improvement in student achievement.

Teacher Leadership

Since the 1980s when teacher leadership first gained traction amongst policymakers, educators, and funders as an option to improve teaching and learning and, ultimately, student outcomes, an array of definitions, paradigms, and programs have proliferated. Teacher leadership can be defined as "the process by which teachers, individually or collectively, influence their colleagues, principals, and other members of school communities to improve teaching and learning practices with the aim of increased student learning and achievement" (York-Barr & Duke, 2004, p. 287-288). In its current form, teacher leadership is guided by three principles. The first is the idea that districts can use teacher leadership to increase instructional capacity and spread effective instructional practice within schools and across districts, simultaneously "moving the needle" on student performance and advancing larger school improvement efforts (Wenner & Campbell, 2017). This emphasis on instructional leadership seeks to leverage teachers' connections to classrooms and their ability to directly influence instruction, combined with their ability to interact with other teachers and influence their instruction, to lead instructional improvement efforts and improve student achievement (Mangin & Stoelinga, 2008).

The second guiding concept is the idea that those who are most impacted by policy should be involved in the decision-making process to improve decisions and increase the likelihood of implementation (Chrispeels, 2004). As "street-level bureaucrats," teachers typically have the discretion to make decisions about student learning in their classrooms, bridging themselves to external demands when doing so enhances implementation of desired goals and buffering when external demands might derail their own decision-making (Honig & Hatch, 2004; Lipsky, 1971). Positioned to influence policy within and beyond classrooms, teacher leadership invites a sort of hybridization, explicitly distributing and decentralizing decision-making to teachers so they might leverage their knowledge of students and instructional practices in their classrooms and across the school community (Chrispeels, 2004; Sands et al., 2022). This notion of hybridity has led to the emergence of new definitions of teacher leadership, including the idea of "hybrid teacher leaders," which is the focus of this study: "a teacher whose official schedule includes both teaching K-12 students and leading teachers in some capacity" (Margolis, 2012, p. 292).

A third motivating principle is the professionalism of teaching, including incentives to enter teaching and stay in the classroom. To attract and retain teachers, programs and initiatives turned to recognizing teachers' leadership abilities, going so far as to designate some teachers as models who might guide and influence peers (Berg et al., 2019). This public recognition is often extended in at least one of three ways: "bestowing an award, conferring a credential, and/or offering an opportunity for influence" (Berg et al., 2019, p. 18). For some teacher leaders, this designation as a teacher leader allowed them to access roles that shared their expertise and, in return, receive "structural supports that included protected time, space, materials, training, and/or compensation" (Berg et al., 2019, p. 19), creating real career advancement opportunities.

The propagation of teacher leadership programs over the past few decades has introduced several paradigms of teacher leadership. These have ranged from organic, or informal, teacher leadership, wherein teachers step up day to day to fulfill a range of responsibilities in their schools without any clear directive, differentiation, or mandate to lead, to fully formal teacher leadership, where positions are "titled, ranked, differently structured, differently authorized, and differently paid" (Smith, 2019, p. 1; Supovitz & Comstock, 2023; Supovitz, 2017). Both researchers and practitioners have been divided over whether informal teacher leadership or formal teacher leadership should be the preferred model of teacher leadership. Those on the informal side of the continuum assert the teacher leader roles and positions should be less structured, empowering more teachers to participate in leadership activities across several functional areas of school life and, thus, improve the overall school environment. On the formal side, it is argued that role definition and a clear position within the hierarchical structures of schools is necessary to give teachers real power to effectuate change and scale the roles across multiple school sites (Margolis, 2020). As I'll return to momentarily, while myriad studies have considered the impact of informal teacher leaders, few have examined the potential of formal teacher leadership.

Regardless of paradigm, teacher leadership inhabits a sort of duality as both an outcome of district policies intended to elevate the voice of teachers in school-based decision-making and influence over instructional practice, while also serving to achieve other valued outcomes, like improved teacher retention, peer collaboration, and, potentially, student achievement (Margolis & Deuel, 2009; Katzenmeyer & Moller, 2011). Studies examining the impact of teacher leadership have focused on elucidating a multiplicity of areas critical to shaping teacher leadership programs, including teacher leaders' roles, responsibilities, and activities; the processes that govern how leadership is distributed; and the context, including networks and relationships, through which leadership practices are forged and executed (for example, Eckert & Daughtrey, 2019; Muijs & Harris, 2003; Chrispeels, 2004; Hulpia et al, 2009). These studies have in turn found teacher leadership to be positively associated with the achievement of a variety of desirable outcomes, including gains in overall school improvement (e.g., Taylor et al., 2011; Chrispeels, 2004; Heck & Hallinger, 2009; Sands, 2023), improvement of school culture (e.g., Taylor et al., 2011), and increased job satisfaction (Hulpia et al., 2009).

With respect to student achievement, it has proven difficult to explicitly link teacher leadership programs to the improvement of student and, subsequently, the field has lacked studies on the effectiveness of teacher leadership vis-à-vis student achievement (Mangin & Stoelinga, 2008). Several studies have focused on informal teacher leadership, finding that schools with higher levels of teacher leadership had greater student achievement (Ingersoll et al., 2017; Heck & Hallinger, 2009). Relatively few studies assess the effects of formal teacher leadership programs on student achievement outcomes and establish a clear link (Harris, 2005). Where formal teacher leadership is examined, the evaluations tend to be of models where teacher leaders are no longer classroom instructors. For example, Supovitz and Comstock (2023) analyzed the effects on student achievement of a formal teacher leadership model leveraging instructional and technology coaches and subject matter specialists. Using an interrupted time series design, they found mixed results but with notable statistically significant improvement in math for students in teacher leader supported schools for the full five-year period of the study. Similarly, Yost et al. (2010) examined the effects of a teacher leader program designed to train coaches to work with teachers at a middle school in Pennsylvania. They found that after two years, the school where teacher leaders worked with teachers returned better scores in reading and math in contrast to the comparison school. However, in both studies, the focus was on teacher leaders no longer embedded in classrooms.

Only two evaluations appear to examine teacher leadership initiatives where teachers endowed with leadership roles also remained in the classroom. The American Institute of Research evaluation of the Iowa Teacher Leadership & Compensation Program examined the effect of rewarding effective teachers with leadership opportunities and higher pay (Citkowicz et al., 2017). Looking across cohorts, districts, and years of implementation, results in math and reading remained largely unchanged. Finally, Leading Educators, a non-profit focused on developing teacher leaders in school communities, conducted an evaluation of their work in Louisiana and Michigan establishing teacher-led, school-based professional learning structures. Using propensity score matching and difference-in-difference approaches, it found positive, statistically significant effects on math scores in both states and on ELA scores in Michigan (Leading Educators, 2019).

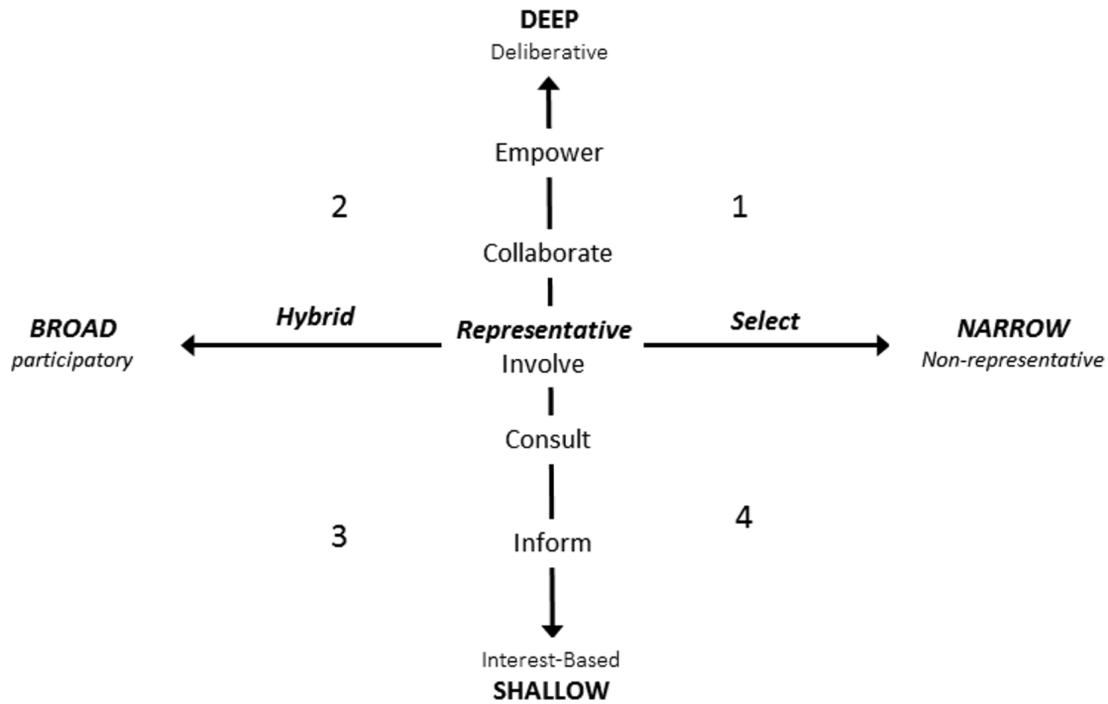
Considering these results, this study makes an important contribution to the literature on teacher leadership by explicitly examining whether and to what degree formal teacher leaders who are in hybrid roles—serving in leadership positions and in the classroom–influence student achievement.

Theorizing Democratic Policymaking and Teacher Empowerment

In the context of education, theories of democratic policymaking draw on two broad bodies of literature: democratic theory and teacher empowerment (Marsh & Hall, 2018; Marks & Louis, 1997). On the democratic theory side, attention has been drawn recently to democratization in organizations, bestowing it greater meaning through newfound emphasis on conversation, reflection, and leadership rooted in connectedness (Woods, 2011). This reflects a changing of tides, stemming from a desire to usher in educational policy regimes that are "professionally engaging, democratically empowering, and organizationally sustainable" (Hargreaves & Shirley, 2009, p. 23). Several models have been developed to help illustrate democratic participation in policymaking in public administration, including in education. These models typically portray democratic decision-making as a process that falls along two continua. Marsh and Hall (2018) developd one such framework of engagement (Figure 1) that elucidates "*who* is involved, *what* is the purpose and scope of engagement, and *how* engagement operates" (p. 245).

Figure 1

Models of Stakeholder Engagement.



Source. Marsh & Hall, 2018, p. 246.

In Marsh and Hall's (2018) model, the horizontal spectrum shows who is involved. On the one end, there is participatory democracy, where all individuals impacted by a decision share equal power to determine decisions. Amongst the many claimed benefits of participatory democracy is the increased likelihood that individuals will support the collectively determined decisions, a prime motivator in the case of education. On the other end, there is non-representative democracy, which pushes for limiting engagement in the policymaking process to only well-informed decision makers. The vertical spectrum maps what the purpose of decision-making is and how it should proceed. At the bottom pole is interest-based democracy, which characterizes decision-making as based on competition for the advancement of private interests. At the top pole is deliberative democracy, which promotes decision-making based on public discourse in which members of the collective consider each other's claims and work to identify the common good. Applying this framework to teacher leadership, it is possible to consider who participates as falling closer to "select" representation, where engagement is expanded beyond district and school administrators but limited to only teachers in the teacher leader role. However, with respect to how decision-making should proceed and its purpose, teacher leadership might fall closer to "empower" as, in its ideal form, teacher leaders have direct authority over some decisions.

The idea of empowerment has long shaped political debates on education (Elmore, 1990). In line with the ethos of democratic policymaking, I interpret empowerment to mean the decentralization of decision-making authority to teachers to promote meaningful involvement in decision-making processes that determine the design and implementation of school wide changes and resource allocation (Robertson et al., 1995). Empowerment encompasses three qualities: real changes in the treatment and exercise of teachers' professional authority, increases in autonomy and involvement in decision-making processes that have a material consequence on teachers' working conditions, and increased authority and input over broader organizational issues beyond teachers' typical daily tasks and routines (Balyer et al., 2017; Bogler & Nir, 2012). Such issues might include budgetary allocations, teacher hiring, selection and implementation of curriculum and instructional practices, organizational goal setting, determination of program participation, and measurement and evaluation of student success (Balyer et al., 2017).

An important component of teacher leadership is the element of teacher empowerment. Empowerment is, in many respects, intended to serve as an antidote to the movement of power away from teachers to higher levels of the educational bureaucratic hierarchy (e.g., school leaders, district leaders, and state leaders). The argument is asserted that the expansion of school bureaucracy at the federal, state, and local levels has limited and undermined the authority of school personnel, especially teachers, in their mission to focus on teaching and learning, and to serve students and their caregivers (Elmore, 1990). Darling-Hammond (1997) classifies this diminution of power as “the symptoms of excessive bureaucratization,” exemplified in the “lack of school-level flexibility for allocating resources—dollars, people, and time—to meet students' needs;” diminished flexibility at the classroom level for determining content, methods, and materials; and increased paperwork aligned to reporting systems designed to monitor school activities in accordance with external directives (p. 64). As teachers are called on to implement and be accountable for policies and procedures that are highly standardized, they experience disempowerment alongside decreasing leeway to exercise expertise (Darling-Hammond, 1997). This, in turn, puts teachers in a bind: They must be responsive to the individual needs of “clients” (students and their caregivers) while conforming to highly prescriptive policies and practices established beyond their realm of influence (Darling-Hammond, 1997; Lipsky, 1971).

In this paper, I examine the influence of teacher leadership on student outcomes through the lens of democratic empowerment. For student achievement to be impacted by democratic policymaking and teacher empowerment, teacher leaders must have the ability to influence instructional vision and professional collaboration (Marks & Louis, 1997). In the case of TCP, teacher leaders are first and foremost supposed to serve as instructional leaders in their school community, advising on curriculum and instructional decisions, aligning schoolwide professional development to those decisions, and developing a culture of collaboration and professional that supports attainment of instructional goals (TCP, 2024). Subsequently, considering empowerment in the context of democratic policymaking, I would expect to see greater gains in student achievement in teacher leadership schools.

Data & Methods

In this section, I detail the data and methods applied to study the impact of TCP on student achievement. As part of this, I include a discussion of the new research on the limitations of traditional difference-in-difference models and the capabilities of new event study approaches to address them.

Data

TCP Program Data

In this study, I rely on the program data from TCP and New York state student achievement scores from grades 3 through 8 on math and English Language Arts (ELA) tests. The primary data source used to establish the sample for the analyses was TCP school program participation data, which were approved for release by the NYCDOE. This data was maintained by the NYCDOE's Office of Teacher Leadership, which managed TCP.[1]

This data included the years a school had qualified teacher leaders and the number of qualified teacher leaders at a school site, and the years a school staffed teacher leaders and the number of staffed teacher leaders. A school with qualified teacher leaders was a school where at least one teacher passed the district's application process to qualify as a teacher leader but were not placed (staffed) into a formal position. A school with staffed, or appointed, teacher leaders was a school where the principal chose to appoint at least one qualified teacher leader into a formal teacher leadership role. Principals appointing a teacher leader were responsible for paying the salary addition to compensate them and agreed to cover the release time for teacher leaders to perform their duties. During the year, teacher leaders spent part of the school day in the classroom with students and the remainder supporting their colleagues through various mechanisms targeting instructional practice and professional growth, such as "1:1 coaching, classroom inter-visitations, facilitating teacher teams, content- or strategy-based workshops, or[and] representative advocacy" between teachers and school leaders (TCP, 2020). The district outlined a broad scope of

[1] Sands (2023a, 2023b) detail the TCP teacher leadership selection process and program in depth.

responsibilities and activities for each kind of role, commensurate with its level and pay. How the role was implemented and what specific activities teacher leaders engaged in, either individually or in teams, was decided by the school.

Based on the above distinctions, the program data was used to establish the three main test groups for my analysis: schools that had staffed teacher leaders; schools that had qualified teacher leaders who were not staffed; and schools that did not have any teacher leaders.

Student Achievement Data

Every year in the spring, students in grades 3 through 8 participate in ELA and math state assessments (NYCDOE, 2022). Unless otherwise stated, tests are administered in all schools, except for specialized school districts serving students with significant challenges (e.g., Autism Spectrum Disorders, significant cognitive delays, or multiple disabilities) or enrolled in alternative learning programs, like District 75 (D75) and District 79 (D79). Results were available from 2013 through 2019. A mean scale score, the average score of the total students tested, is generated for all grades in a school each year the test is administered. These were standardized for use in this analysis. All elementary and middle schools with scores in ELA and math were.

This data was converted into a panel dataset that indicated the year in which a school first had qualified teacher leaders and the first year a school staffed teacher leaders. I included additional variables to indicate the status of each school for each year, including whether a school had staffed teacher leaders, qualified teacher leaders, or no teacher leaders.

Methods

To study the overall effect on student achievement for the period 2013 to 2019, I employed a quasi-experimental design. First, I identified two levels of treatment – qualified teacher leaders and staffed teacher leaders and devised a number of comparison groups for each treatment level.[2] To examine these pairings and determine treatment groupings for deeper analysis, I started with a difference-in-difference model employing two-way fixed effects (DiD TWFE) that incorporated both school-level fixed effects and year fixed effects, expressed by the following model:

$$Y_{st} = \beta_0 + \beta_1 TCP_{st} + \gamma_s + \lambda_t + \varepsilon_{st}$$

where Y_{st} is a NYSED-tested measure of student achievement in school s in year t. TCP_{st} is a binary indicator that equals one if a school ever had a qualified teacher leader or a staffed teacher leader during the treated time period. β is the coefficient that allows the assessment of the effect of TCP on the outcomes. The inclusion of school fixed effects (γ_s) controls for time invariant characteristics of schools during the analysis window. The year fixed effect (λ_t) controls for factors that change over time across all schools, excluding the presence of TCP, and could influence student achievement, such as changes in district administration, district or state policies, agreements reached with the UFT, state or district educational standards, and changes in the economic and political environment. Additional controls were not added, given the efficiency and high explanatory power of the TWFE. Finally, ε_{st} is a mean-zero random error term. My two outcome variables are standardized mean scale scores in math and ELA for all grades 3 through 8. All standard errors were clustered at the school level, as this is the unit assigned to treatment. I ran all models on a sample that excluded charter schools, which were not eligible to participate in TCP, and those schools that were never tested.

In the above specified model, all years following the introduction of TCP were pooled to calculate the average effect of having a qualified or staffed teacher leader. However, as TCP is implemented over multiple time periods and schools implement the program at various times, the

[2] See Appendix G in Sands (2023b) for the pairings.

use of the conventional DiD model with only two time periods (a pre- and post-treatment period) can lead to biased results (Schueler & Bleiberg, 2021). Specifically, staggered adoption can lead to bias in results due to dynamic treatment effects, wherein present values of the dependent variable are influenced by past values and by unobservable explanatory variables, and treatment effect heterogeneity, wherein the present values vary in magnitude and direction of treatment effects due to nonrandom, explainable variables based on the period in which schools were first treated or the evolution of treatment effects over time (Callaway & Sant'Anna, 2021; Varadhan & Seeger, 2013; Sun & Abraham, 2021).

To minimize heterogeneity bias, I conducted the DiD TWFE once more using a stacked dataset that offered an opportunity for "clean controls" – the creation of a comparison group for each individual treated school comprised of schools that were never or not yet treated in the year the treated school received treatment. Such an approach eliminated the potential of already-treated schools serving as comparisons for newly-treated schools by ensuring only schools that were never or not yet treated served as comparisons for schools implementing in specific years, or "cohorts."

However, with still only two periods to examine – one pre- and one post-treatment period each, I remained limited in my ability to understand how TCP results evolved over time and to learn about treatment effect heterogeneity. To see how treatment effects evolved over time, and to better account for both heterogeneity and dynamic effects, I conducted an event study. For the event study analysis, I focused on two pairings that yielded the most insight in initial tests: comparing schools with only qualified teacher leaders ever to those without any qualified teacher leaders, and schools with staffed teacher leaders ever to those without staffed teacher leaders. The event study model follows a similar framework to the DiD TWFE but indexes the years relative to the event (the year of adoption of TCP) by k, as modeled by Weill et al. (2021). The event study equation takes the following form:

$$Y_{st} = (\sum_{k=-6}^{k=6} \theta_{pk} \times D_{spk}) + \gamma_s + \lambda_t + \varepsilon_{st}$$

The event study coefficient (θ_{pk}) is the estimate of adopting TCP (p) relative to the time to the event (k). D_{spk} is a dummy variable equal to 1 when school s is k years away from being "treated" by TCP (either qualified teacher leaders or staffed teacher leaders). Formally, D_{spk} = $1\{t - TCP_{sp} = k\}$, where TCP_{sp} is the policy p in place for school s, with p being either qualified or staffed teacher leaders, depending on the analysis. I take the standard event study approach of including a dummy for all relative years before the window in which TCP could be implemented (maximum six years pre-implementation), denoted by k = -6, and another for all relative years after, k = 6 (maximum six years post-implementation). The post-treatment indicators enable the assessment of the effect of TCP by the number of years schools are exposed to the program. This means that my panel is unbalanced as each school was not observed every year. As in the standard DiD TWFE analysis, I include school fixed effects (γ_s) to control for unobserved time-invariant differences between schools and year fixed effects (λ_t) to control for unobserved time-varying variables that are common to all schools. Finally, ε_{st} is a mean-zero random error term. Standard errors are clustered at the school level.

I implemented the event study using Borusyak et al. (2021) and Callaway and Sant'Anna's (2021) method, the latter of which relies on aggregating treatment effects of the treated across groups and years.[3] In light of my data limitations, including smaller samples in earlier program

[3] In this approach, causal parameters relevant to each "group" of schools – schools that implement TCP at the same time – are identified, which are then aggregated to identify summary causal effects for each group and for each year. These group-time average treatment effects (ATTGT), which can be defined as the average treatment effect for the

years, I selected Callaway and Sant'Anna's approach in reporting my results, given its ability to examine more post-treatment periods for all cohort years and analyze treatment effects for individual cohort years. It also offered access to additional pretrends testing to satisfy the parallel trends assumption and easily identified and interpreted coefficients to better understand the heterogeneous treatment effects. The staggered adoption of qualified or staffed teacher leaders presented several issues with respect to confirming adherence to parallel trends. In a traditional DiD TWFE model, heterogeneity bias may erroneously suggest that there are pretrends that violate the parallel trends assumption. To correct for the distortion of pretrends resulting from heterogeneity bias, the Callaway and Sant'Anna event study estimator offered testing of pretrends to confirm that there wa no violation of the parallel trends assumption in the pre-treatment period. First, the pretrends were calculated for each pre-treatment year for each group of schools implementing TCP (either qualified or staffed teacher leaders) in a given year (e.g., pretrends for 2014 and 2015 were calculated for schools implementing TCP in 2016). These tests suggested no differences in pretrend effects between schools in pretreatment years that might be influencing post-treatment outcomes. Second, a post-estimation of the pretreatment trends was conducted to determine if all pretreatment average treatment effects on schools implementing TCP were equal to zero. These tests failed to reach statistical significance, meaning that all ATTs were equal to zero and the parallel trends assumption held.

---

group g at time t, where group is the time period when a set of units is first treated, can then facilitate additional estimation and inference. This includes learning about treatment effect heterogeneity and serving as the basis for creating other aggregated causal parameters (Callaway & Sant'Anna, 2021). The advantage of the ATTGT is that the parameters do not restrict heterogeneity with regards to covariates, the first treatment period for schools, or the change in treatment effects over time (Callaway & Sant'Anna, 2021).

Sample

The sample in this paper includes elementary and middle schools only. With sizable samples, I observe small but statistically significant differences across many school characteristics between schools that had only qualified teacher leaders, schools that staffed teacher leaders, and schools that had no teacher leaders ever. In all groups, there are more elementary schools than middle schools in the sample. Schools staffing teacher leaders had much lower student achievement scores than the other two groups. Schools staffing teacher leaders had a higher percentage of students of color, including Black and Hispanic students. They also served a higher percentage of students with disabilities, English language learners, students from lower socioeconomic backgrounds, and male students. Though the mean per pupil funding was higher than schools that never staff teacher leaders, the student population and test scores suggest that schools opting-in to participate in TCP were serving subgroups of students that had a greater need for services and additional supports to improve educational outcomes. The high standard deviations, particularly with respect to test scores, indicate that significant variability is driven from within group differences.

Table 1

Select Characteristics of NYCDOE Schools by Group, 2014 to 2021

	Qualified Teacher Leaders Only (*n*=1,629)		Staffed Teacher Leaders Ever (*n*=8,782)		No Teacher Leaders Ever (*n*=4,019)	
Characteristic	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Total enrollment	738.303	673.231	586.979	518,421	589.599	360.852
Per pupil funding	$8,243.68	$22,124.68	$7,504.27	$4,619.52	$6,954.81	$6,246.30
Pupil to teacher ratio	14.967	2.811	13.900	2.844	14.557	2.816
Elementary schools	0.577	0.494	0.402	0.490	0.807	0.395
Middle schools	0.309	0.462	0.396	0.489	0.253	0.435
Female	0.490	0.082	0.478	0.108	0.482	0.060
Students of color	0.810	0.220	0.910	0.149	0.814	0.219
Black students	0.264	0.259	0.329	0.253	0.251	0.276
Hispanic students	0.362	0.230	0.465	0.251	0.370	0.249
Students with disabilities	0.214	0.170	0.238	0.163	0.217	0.163
English language learners	0.126	0.159	0.149	0.152	0.139	0.131
Free/ reduced price lunch or eligible for Human Resources Administration benefits	0.707	0.250	0.821	0.171	0.745	0.242
ELA, all grades (3-8)	0.394	1.038	-0.265	0.897	0.326	1.015
Math, all grades (3-8)	0.344	1.024	-0.269	0.912	0.350	0.993

Notes: All variables except for school level have missing data. Per pupil funding is calculated using the Fair Student Funding amount from available School Allocation Memorandum divided by total enrollment. Pupil to teacher ratio is missing for all schools in SY 2020-21. All test scores are standardized by subject. No test scores are available from SY 2019-20 and SY 2020-21.

Findings

Student Achievement Effects Using Traditional DiD Approaches

The DiD TWFE estimates using the stacked datasets (Table 2), pooling all years from 2013 to 2019, provide early indicators that formalized roles for teacher leaders may influence student achievement. At first glance, the results appear to suggest that schools with qualified teacher leaders and schools with staffed teacher leaders both experience academic improvement that is statistically significant. However, in the first model, schools with staffed teacher leaders are also considered as having qualified teacher leaders. Looking at schools with only qualified and never any staffed teacher leaders, there are no treatment effects. This suggests statistically significant improvement in schools with qualified teacher leaders in the first instance might be driven by the schools with staffed teacher leaders included in that sample.

At the same time, schools with staffed teacher leaders register a small, positive, statistically significant effect in both subjects across all models. All schools with staffed teacher leaders, regardless of whether they staff once or for multiple years, experience improvement in ELA scores (0.074 SD, $p<0.001$) and math scores (0.077 SD, $p<0.001$). In short, the DiD TWFE performed with a stacked dataset suggests that staffed teacher leaders drive the positive statistically significant changes in student achievement observed in grades 3 through 8. Schools with qualified teacher leaders that are not staffed do not experience a change in student achievement scores.

Table 2

Comparing Effects of TCP Treatment Variation on Grades 3-8 ELA & Math Scores, Stacked Dataset

Treated/comparison group pairings	ELA	Math
Schools with qualified TLs, including schools that may have staffed teacher leaders, compared to schools without any teacher leaders	0.063*** (0.015)	0.060*** (0.015)
Observations (N)	3,212,419	3,212,910
Schools with staffed TLs, compared to schools with no teacher leaders and schools with only qualified teacher leaders	0.074*** (0.018)	0.077*** (0.018)
Observations (N)	2,964,618	2,964,506
Schools with staffed TLs, compared to schools that never had qualified or staffed TLs	0.077*** (0.018)	0.080*** (0.018)
Observations (N)	2,473,639	2,474,042
Schools with qualified but never staffed TLs, compared to never had TLs	0.030 (0.030)	0.005 (0.031)
Observations (N)	2,552,016	2,552,015
Schools that staffed all years they had qualified TLs, compared to schools with qualified teacher leaders but didn't staff every possible year	0.056** (0.018)	0.064*** (0.018)
Observations (N)	1,109,680	1,109,568

Notes. Standard errors clustered by dataset. ****p*<.001, ***p*<.01, **p*<.05

Student Achievement Effects Using an Event Study

The results of the event study provide confirmatory evidence for the results obtained in the DiD TWFE, with the benefit of additional insight into outcomes over time. Table 3 presents the aggregated average treatment effect of the treated by periods before and after treatment for ELA all grades scores; Table 4 does the same for math all grades scores. Each table shows the effects for the separate pre- and post-treatment years, including the year of TCP introduction. These are presented for both treatment levels: schools with qualified teacher leaders only (schools never staffing teacher leaders) and schools with staffed teacher leaders.[4]

[4] As noted previously, only the Callaway and Sant'Anna estimator results are shown here. The results of the event study using the Borusyak et al. (2021) DiD imputation estimator can be found in Appendix L in Sands (2023a).

Schools with staffed teacher leaders show statistically significant gains for both subjects that, overall, improve with each subsequent year, growing from 0.052 SD (*p*<0.001) in the introductory year to 0.290 SD (*p*<0.001) in year 5 of the post-implementation period for ELA and 0.051 (*p*<0.01) in year 1 of the post-implemenation period to 0.192 (*p*<0.001) in year 5 for math. Figure 2 depicts the differences, showing clearly the drop in outcomes in both subjects at schools with only qualified teacher leaders compared to the consistent increases for schools with staffed teacher leaders. Note that the confidence intervals over time widen with each post-implementation period. This is likely because with each post-implementation period, fewer schools would be in the sample, leading to a larger confidence interval with a larger margin of error.

In contrast, schools with only qualified teacher leaders show no statistically significant improvement in ELA or math at any point during the implementation period. Indeed, while the results are not statistically significant, schools with qualified teacher leaders only experience declines. Individual cohort years were as likely to have sustained, compounded declines over multiple post-treatment years as they were to have a mix of gains and declines, but, again, none of these changes were statistically significant.[5]

[5] See Appendix O in Sands (2023).

Table 3

The Effect of Teacher Leadership on ELA Test Scores (All Grades)

	Schools with Qualified Teacher Leaders Only	Schools with Staffed Teacher Leaders
Pre-TCP Year -5	-0.061	-0.015
	(0.043)	(0.028)
Pre-TCP Year -4	-0.034	-0.001
	(0.025)	(0.024)
Pre-TCP Year -3	0.020	0.004
	(0.023)	(0.016)
Pre-TCP Year -2	-0.002	-0.007
	(0.026)	(0.014)
Pre-TCP Year -1	0.025	0.013
	(0.028)	(0.014)
Year of TCP Introduction	0.025	0.052***
	(0.024)	(0.013)
TCP Year 1	0.056	0.076***
	(0.035)	(0.018)
TCP Year 2	0.036	0.104***
	(0.067)	(0.025)
TCP Year 3	-0.177	0.161***
	(0.104)	(0.030)
TCP Year 4	-0.203	0.158***
	(0.152)	(0.037)
TCP Year 5	N/A	0.290***
		(0.050)
Observations (N)	13,506	22,618

Notes. Panels are not balanced; only observations with pair balanced at periods t0 and t1 used. In the "qualified" column, schools with qualified teacher leaders only (never staffing) are compared to schools that never have teacher leaders. All schools that had qualified teacher leaders in 2014 also staffed them; as such, there are no TCP schools on which to report on for TCP Year 5. In the "staffed" column, schools with staffed teacher leaders ever are compared to schools that never staff teacher leaders. Each column and row intersection describes a different set of regressions where the sample is restricted to include a specified cohort (e.g., all schools with staffed teacher leaders a certain number of periods pre- or post-treatment) and all schools that have never or not yet experienced the treatment at that time. Models include all school and year fixed effects. ***p<.001, **p<.01, * p<.05

Table 4

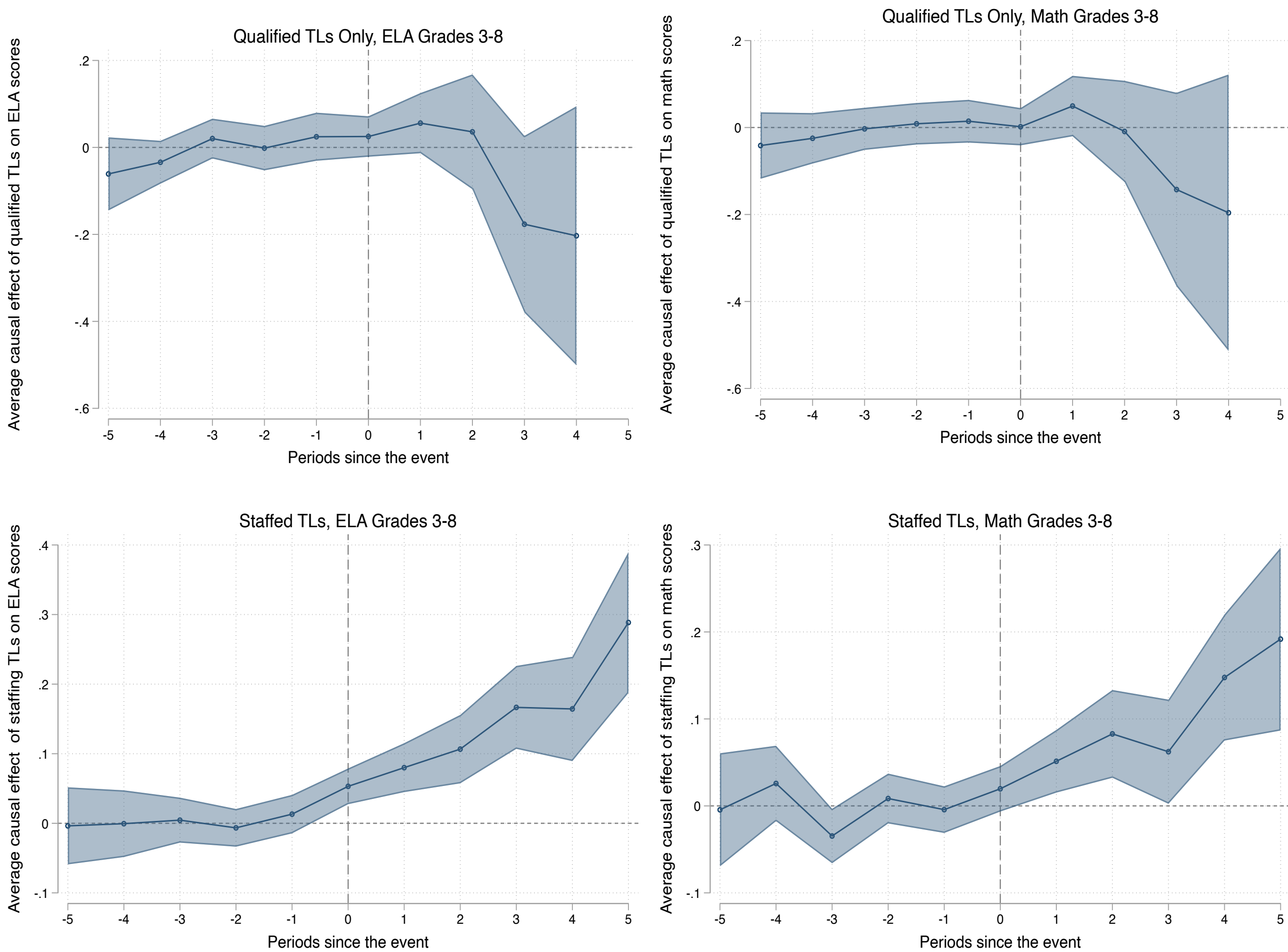
The Effect of Teacher Leadership on Math Test Scores (All Grades)

	Schools with Qualified Teacher Leaders Only	Schools with Staffed Teacher Leaders
Pre-TCP Year -5	-0.041	-0.005
	(0.039)	(0.033)
Pre-TCP Year -4	-0.025	0.026
	(0.030)	(0.022)
Pre-TCP Year -3	-0.003	-0.035*
	(0.025)	(0.016)
Pre-TCP Year -2	0.009	0.008
	(0.024)	(0.015)
Pre-TCP Year -1	0.015	-0.004
	(0.025)	(0.014)
Year of TCP	0.002	0.020
Introduction	(0.022)	(0.013)
TCP Year 1	0.049	0.051**
	(0.035)	(0.018)
TCP Year 2	-0.009	0.083**
	(0.059)	(0.026)
TCP Year 3	-0.143	0.062*
	(0.114)	(0.031)
TCP Year 4	-0.196	0.148***
	(0.162)	(0.037)
TCP Year 5	N/A	0.192***
		(0.054)
Observations (N)	13,505	22,614

Notes. Based on Callaway and Sant'Anna (2021) estimator. Panels are not balanced; only observations with pair balanced at periods t0 and t1 used. In the "qualified" column, schools with qualified teacher leaders only (never staffing) are compared to schools that never have teacher leaders. In the "staffed" column, schools with staffed teacher leaders ever are compared to schools that never staff teacher leaders. Each column and row intersection describes a different set of regressions where the sample is restricted to include a specified cohort (e.g., all schools with staffed teacher leaders a certain number of periods pre- or post-treatment) and all schools that have never or not yet experienced the treatment at that time. Models include all school and year fixed effects. ***p<.001, **p<.01, * p<.05

Figure 2

Pre- and Post-TCP Teacher Leadership Effects on Test Scores on All Grades 3 through 8[6]



Where schools that have qualified teacher leaders but never took the next step of staffing them did not reap any benefits in terms of improved student achievement outcomes, nearly every

[6] An important difference between the CSDID estimator, the command that implements the DID for multiple time periods proposed by Callaway and Sant'Anna (2021), and other event study approaches is that there is no year that is omitted to serve as a baseline (Rios, 2021). To examine periods after treatment, the effect is measured by comparing the last period prior to the first year of treatment to a given year post-treatment. For periods before treatment occurred, the base periods change based on the 2x2 DiD comparison, so that, for example, the pre-treatment coefficient for T-2 (pre-treatment year 2) is derived using T-3 as a base period. Given this difference in comparison groups, the graphs and tables report all years examined.

cohort of schools staffing teacher leaders saw greater improvement in student achievement with each passing year. A look at student achievement outcomes by cohort highlights how staffing teacher leaders into formal roles may have a snowball effect on schoolwide academic performance overtime. Table 5 displays the pre- and post-treatment outcomes for ELA and math by the year in which schools first staffed teacher leaders. In ELA, the 2014 and 2016 cohorts are the main drivers of the statistically significant positive results captured in the aggregated post-treatment outcomes. These cohorts exemplify the compounded nature of academic improvement, with both cohorts showing increasing gains year on year in the post-treatment period. In math, cohorts 2014, 2016, and 2018 are drivers of the statistically significant positive results captured in the aggregated post-treatment outcomes. The 2014 and 2016 cohorts in particular display gains that increase with each subsequent post-treatment year. In short, staffing teacher leaders may have an amplifying impact on academic progress for students – student achievement continuously improves in schools that staff teacher leaders, relative to schools that do not. Figure 3 provides a visual depiction of impact for the 2014 and 2016 cohorts in ELA and math.

The varying results between cohorts are indicative of the heterogeneous treatment effects mentioned earlier. The early cohorts, particularly 2014 and 2016, appear to act as drivers of the aggregated outcomes for schools with staffed teacher leaders, suggesting that there is something distinct about these cohorts that differentiates them from others. This could be attributable to any number of factors, such as the amount of funding, the kinds of implementation support offered, or how the principals implemented TCP in their school communities. Thus, while these results indicate that staffing teachers as teacher leaders in their school communities may lead to improvement in ELA and math scores, the heterogenous treatment effects suggest that there may be some characteristics that differ across cohorts and lead to variability in outcomes.

Finally, it is worth mentioning that for each subject, there is a pre-treatment coefficient that is statistically significant. For ELA, pre-treatment year 1 for the 2015 cohort is statistically significant and, in math, pre-treatment year 2 for the 2018 cohort is statistically significant. For ELA, the statistical significance could potentially indicate that the 2015 cohort is different enough from comparison schools that they were from the start on a different trajectory with respect to student achievement. However, the lack of statistical significance for all post-treatment coefficients for the 2015 cohort eliminates any cause for concern; the results for the 2015 cohort are not driving the aggregated effects in Table 5. In the case of math, the same rationale might explain the statistical significance, but the distance away from the treatment years means that there is no reason to suspect this has an impact on the post-treatment results. Overall, while significant pre-treatment effects could indicate a violation of the parallel trends assumptions, there is no reason to be concerned in this case.

Table 5

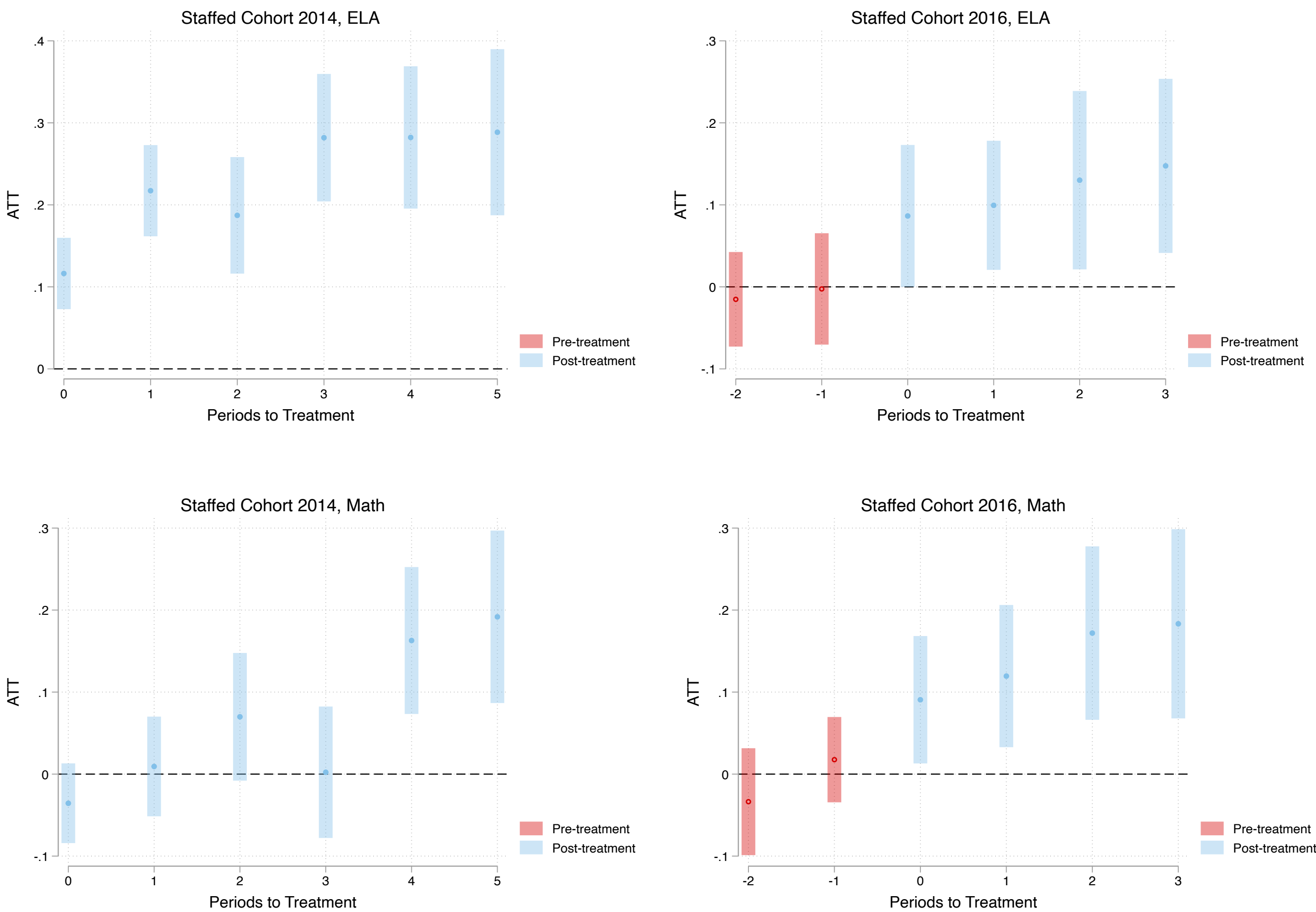
The Effect of Staffed Teacher Leaders on Test Scores by Cohort

	ELA						Math					
Cohort	2014	2015	2016	2017	2018	2019	2014	2015	2016	2017	2018	2019
Pre-treatment year -5						-0.015						-0.005
						(0.028)						(0.033)
Pre- treatment year -4					-0.036	0.023					0.015	0.034
					(0.037)	(0.031)					(0.035)	(0.028)
Pre- treatment year -3				0.029	0.017	-0.024				-0.026	-0.031	-0.043
				(0.029)	(0.030)	(0.024)				(0.030)	(0.029)	(0.025)
Pre- treatment year -2			-0.014	-0.023	0.026	-0.015			-0.035	0.024	0.085**	-0.033
			(0.029)	(0.029)	(0.027)	(0.025)			(0.033)	(0.025)	(0.030)	(0.025)
Pre- treatment year -1		0.060*	-0.008	0.012	0.028	-0.014		-0.022	0.018	-0.009	-0.016	0.004
		(0.026)	(0.034)	(0.029)	(0.033)	(0.029)		(0.032)	(0.026)	(0.027)	(0.033)	(0.030)
Treatment year 1	0.107***	-0.044	0.090*	0.053	-0.010	0.073**	-0.036	-0.011	0.091*	0.007	0.044	0.042
	(0.022)	(0.034)	(0.044)	(0.034)	(0.034)	(0.025)	(0.025)	(0.029)	(0.040)	(0.032)	(0.039)	(0.030)
Treatment year 2	0.205***	-0.088*	0.102*	0.048	0.020		0.009	-0.026	0.120**	0.062	0.105**	
	(0.028)	(0.040)	(0.041)	(0.044)	(0.038)		(0.031)	(0.038)	(0.044)	(0.048)	(0.040)	
Treatment year 3	0.179***	-0.034	0.131*	0.081			0.070	-0.016	0.172**	0.102		
	(0.036)	(0.050)	(0.055)	(0.050)			(0.040)	(0.048)	(0.054)	(0.059)		
Treatment year 4	0.272***	-0.028	0.159**				0.002	0.031	0.183**			
	(0.039)	(0.053)	(0.054)				(0.041)	(0.054)	(0.059)			
Treatment year 5	0.274***	-0.045					0.163***	0.121*				
	(0.043)	(0.055)					(0.046)	(0.056)				
Treatment year 6	0.290***						0.192***					
	(0.050)						(0.054)					
Observations (N)	3,906	3,654	3,722	3,707	3,746	3,883	3,906	3,653	3,721	3,706	3,746	3,882

Notes: Panels are not balanced; only observations with pair balanced at periods t0 and t1 used. Each column and row intersection describes a different set of regressions where the sample is restricted to include a specified cohort (e.g., all schools with staffed teacher leaders starting in 2015) and all schools that have never or not yet experienced the treatment at that time. Models include all school and year fixed effects. ***p<.001, **p<.01, * p<.05

Figure 3

Examples of Treatment Effects by Subject and by Staffed Cohort Year



Discussion & Implications

Proponents of teacher empowerment assert that through formalized leadership roles for teachers, teachers would be able to make sustained contributions to decision-making on areas fundamental to teacher and student success (Hallinger & Heck, 2010; Rice & Schneider, 1994). This should support teachers in maintaining policies and practices intended to bolster student success over time, compounding the policy effects. My findings suggest that this may indeed be happening in schools staffing teacher leaders through TCP: Schools with staffed teacher leaders showed improvement in both ELA and math scores compared to other schools. The student achievement effects identified here build on the findings from the first paper in this series, which conjectured that the positive gains in stakeholder experiences and school quality observed in schools staffing teacher leaders into formal roles should ultimately yield positive gains in student achievement (Sands, 2023a). This suggests that by empowering teachers through distributed leadership and targeting their efforts at improving instructional practices and collaboration between teachers, teacher leaders and school leaders, improvements in distal outcomes like student achievement can be advanced (Sands et al., 2022; Margolis, 2020).

Observing compounded changes in student achievement over time offers more insight into the relationship between teacher empowerment, formal leadership, and outcomes. Recall from earlier Marsh and Hall's (2018) model of stakeholder engagement to understand how teacher leaders are situated as participants in democratic policymaking. On the spectrum of representativeness (who is involved), teacher leadership remains on the selective side as the program is not about expanding decision-making power to all teachers, just to teacher leaders. However, on the spectrum that conceptualizes "what the purpose of decision-making is and how the process should operate," teacher leadership reflects empowerment, just one step below deliberative democracy where community members all consider each other's positions and commit to promoting the common good (Marsh & Hall, 2018, p. 247). My results suggest that empowerment for teacher leaders and, thus, the ability to engage in the process of democratic policymaking within their schools is contingent on being selected to lead and hold positional authority in their school communities. Qualified teacher leaders, who have been identified by the district as having the capability to lead but are not staffed by their school leaders into a role where they have the authority to do so, are not associated with any measurable impact on school change as reflected by School Quality Review results (SQR) (Sands, 2023) and student achievement outcomes. On the other hand, staffed teacher leaders, who have been chosen by their school leaders to serve as teacher leaders in defined roles with broad responsibilities, a mandate to lead, and commensurate compensation, are associated with improvement on valued outcomes.

This finding mirrors much of this theory behind teacher empowerment and democratic policymaking. Reflecting prior research, only teacher leaders that have formal recognition, meaning they are selected by their principal to serve in a teacher leadership role with contracted responsibilities and commensurate pay, hold space within the structure of complex systems to exercise agency (Mayer et al., 2013). In this case, empowerment in the democratic policymaking process is exercised through planned, intentional transfers of power, which in turn allows for teacher leaders to engage in a continuous series of different actions with school leaders and teachers that can combine to produce systemic changes (Cairney, 2012). These actions include exercising autonomy, enacting agency, making decisions, and taking steps that can, with every interaction and over time, transform the school culture and subsequent outcomes (Mayer et al., 2013). Based on this definition, qualified teacher leaders possess none of the enabling features that would make it possible for them to effectuate change within the complex systems in which they operate and, as such, they have no clear schoolwide impact. Meanwhile, staffed teacher leaders have the knowledge of what is happening in their classrooms, amongst their peers, and in the realm of school leaders, and they would have the positional authority to bring this information to bear on decisions that have material consequences on school conditions and instructional improvement (Margolis, 2020; Sands et al., 2022; Marks & Louis, 1997). This would subsequently enable staffed teacher leaders to influence policies that directly impact outcomes such as school quality and, in the long term, student academic achievement (Robertson et al., 1995).

This has implications for the field of teacher leadership. Those studying and practicing teacher leadership have debated the question of formalization in teacher leadership for the better part of two decades. In particular, it has been argued that formalization could lead to reproduction of the power dynamics within school systems, where an elite few (e.g., district administrators, school leaders) have power over teachers and support staff working with students in classrooms (Margolis, 2020). Despite fears that formalized structures could constrain and thwart a seemingly more natural, organic leadership from emerging in schools, I find that a level of formality is necessary and beneficial for teacher leadership to be able to change the organizational structures and practices of schools required for teacher experiences and student learning to improve. Specifically, positional authority, including recognition, compensation, and role definition, appears to be a condition for teacher leaders to influence how and what decisions are made in complex systems with entrenched hierarchies and sway outcomes (Margolis, 2020; Muijs & Harris, 2003).

Indeed, my findings suggest that informal teacher leadership may be too diffuse and ambiguous to influence distal outcomes like student achievement. Qualified teacher leaders may be serving as informal teacher leaders in their school buildings, but this is not associated with the sort of measurable improvement in school quality and student achievement that I find to be associated with staffed teacher leaders. This suggests that while informal teacher leadership undoubtedly has an important place in schools, states, districts, and schools designing teacher leadership programs must give thought to elements of formalization if programs geared at teacher empowerment are to yield measurable change. For schools unable to provide benefits like compensation, consideration for role definitions and how teacher leaders will be incorporated into decision-making structures might help give a level of formalization and recognition that establishes positional and relational authority vis-à-vis both school leaders and teachers in the school community so that the benefits of their expertise might impact the school community.

Assumptions & Limitations

While the DiD TWFE and event study approaches offer many benefits, there are some assumptions that can pose challenges in the case of TCP. The first assumption is that in the case of DiD with staggered treatment adoption, all schools that receive treatment in a year remain treated in the following periods. Translating this assumption to TCP, this suggests that even if a school only has staffed teacher leaders for one year, the experience of ever having been treated with staffed teacher leaders leaves an impression on the school that it doesn't "forget" and so it changes the course of the post-treatment outcomes, even if the school never has staffed teacher leaders again. It is not clear that this is necessarily true, but for the purposes of the event study, I adopt the assumption that any interaction with TCP at any time changes the school's course.

A second assumption is that the assignment of the treatment is to some degree random – eligible teachers can apply at any time to become qualified as teacher leaders and school leaders, so long as they have qualified teacher leaders, can choose to staff them. This may not align perfectly with reality. For example, schools may become eligible for central funding or receive grants from external organizations to fund teacher leadership, and this prompts the school leader to staff existing qualified teacher leaders or recommend teachers apply to become qualified with the goal of staffing them. Once funding is procured, there could be reason to suspect that schools will continue to participate in the program, enticed by future funding possibilities, or because the program becomes enmeshed in the school as an institutional change. Given that this is a quasi-experimental study, strict random assignment of the treatment is not wholly necessary. Thus, I assume that in some cases there is true randomness, where teachers and principals are inclined to opt-in to participate in TCP of their own volition, but for others, this may not be the case.

Finally, as mentioned previously, the parallel trends assumption, that baseline outcomes are consistent for treated and comparison groups is assumed in methods for causal identification. As only one year of pre-treatment data is available for the 2014 cohort, this is not entirely perfect. In light of this, the Callaway and Sant'Anna approach is particularly useful as it allows for more flexibility around this assumption. Specifically, it activates two parallel trends, one for a "never-treated group" and the other for a "not-yet-treated group," and both are conditional on covariates, which allows for the examination of trends that are conditional to covariates over time. Callaway and Sant'Anna allow for the examination of pretrends and post-estimation testing to confirm that the parallel trends assumption holds. As I discussed earlier, my pretrends and post-estimation testing suggest that the parallel trends assumption holds for my study. However, given the limitations of the data, it is worth noting the challenge of establishing pretrends for all years.

Conclusion

In this study, I asked if student achievement scores improved more in schools participating in TCP compared to schools that did not participate in TCP. I explored two levels of treatment, looking at 1) schools staffing teacher leaders, and 2) schools with qualified but not staffed teacher leaders. Schools staffing teacher leaders experienced gains in ELA and math for grades 3 through 8 that were statistically significant from schools that did not staff teacher leaders. Moreover, schools staffing teacher leaders experienced compounded gains in both subjects, with scores increasing year on year for every year in the post-treatment period. Schools with qualified teachers that did not staff them did not improve. Indeed, just as schools with staffed teacher leaders saw compounded gains, schools with qualified teacher leaders only experienced intensifying declines over the post-treatment years.

My results confirm what was found in the case of the SQR results explored in the first part of this study (Sands, 2023), in which I found having qualified teacher leaders alone was insufficient when it came to improving school quality. In this instance, only schools staffing teacher leaders experienced meaningful improvement of student achievement outcomes. In both ELA and math, student achievement improved in schools staffing teacher leaders, with improvement increasing over time. This largely echoes teacher empowerment and democratic policymaking theories – where teacher leaders can assert real authority in determining, interpreting, and implementing policies, desired policy outcomes are more likely to be attained (Marks & Louis, 1997). Accordingly, for teachers to exert meaningful power in schools that changes outcomes like student achievement, teacher leadership positions must be formalized, ideally with recognition, responsibility, agency, and compensation.

References

Balyer, A., Özcan, K., & Yildiz, A. (2017). Teacher empowerment: School administrators' roles. *Eurasian Journal of Educational Research, 70*, 1-18.

Berg, J., Horn, P., Supovitz, J., & Margolis, J. (2019). Typology of Teacher Leadership Programs. CPRE Research Reports. Retrieved from https://repository.upenn.edu/cpre_researchreports/109

Bogler, R. & Nir, A. E. (2012). The importance of teachers' perceived organization support to job satisfaction: What's empowerment got to do with it? *Journal of Educational Administration, 50*(3), 287-306.

Borusyak, K., Jaravel, X., & Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. arXiv:2108.12419.

Cairney, P. (2012). Complexity theory in political science and public policy. *Political Studies Review, 10*, 346-358.

Callaway, B. & Sant'Anna, P. H. C. (2021). Difference-in-difference with multiple time periods. *Journal of Econometrics*, *225*(2), 200-230,

Chrispeels, J. H. (2004). The dynamics of sharing and distributing leadership. In J. H. Chrispeels (Ed.), *Learning to Lead Together: The Promise and Challenge of Sharing Leadership* (p. 3-20). Thousand Oaks, CA: Sage Publications.

Citkowicz, M., Brown-Sims, M., Williams, R., & Gerdman, D. (2017). *Iowa's Teacher Leadership and Compensation Program: Findings from 2016–17*. Downloaded March 19, 2021 from https://www.air.org/resource/iowa-s-teacher-leadership-and-compensation-program-findings-2016-17

Darling-Hammond, L. (1997). *The right to learn: a blueprint for creating schools that work.* 1st ed. San Francisco: Jossey-Bass.

Eckert, J. & Daughtrey, A. (2019). Teacher leadership development: tracking one district's progress over three years. *Education Policy Analysis Archives, 27*(42), 1-22.

Elmore, R. (Ed.) (1990). *Restructuring schools: The next generation of educational reform.* San Francisco: Jossey-Bass Publishers.

Gallagher, C. (2008). Democratic policymaking and the arts of engagement. *The Phi Delta Kappan, 89*(5), 340-346.

Good, A., Barocas, S. F., Chávez-Moreno, L.C., Feldman, R., & Canela, C. (2017). A seat at the table: How the work of teaching impacts teachers as policy agents. *Peabody Journal of Education, 92*(4): 505-520.

Hallinger, P. & Heck, R. H. (2010). Collaborative leadership and school improvement: understanding the impact on school capacity and student learning, *School Leadership and Management, 30*(2), 95-110.

Hargreaves, A., & Shirley, D. (2009). *The fourth way: The inspiring future for educational change*. Corwin Press.

Harris, A. (2005). Teacher leadership: More than just a feel-good factor? *Leadership and Policy in Schools, 4*(3), 201–219.

Hatcher, R. (2005). The distribution of leadership and power in schools. *British journal of sociology of education, 26*(2), 253–267.

Heck, R. H. & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal, 46*(3), 659-689.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16-30. doi:10.3102/0013189X033008016

Hulpia, H., Devos, G., & Rosseel, Y. (2009). The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organizational commitment. *School Effectiveness and School Improvement, 20*(3), 291-317.

Ingersoll, R., Sirinides, P, & Dougherty, P. (2017). School leadership, teachers' roles in school decision-making, and student achievement. CPRE Working Papers.

Katzenmeyer, M. & Moller, G. (2011). Chapter one: Understanding teacher leadership. *Counterpoints*, 408, 3-21.

Leading Educators. (2019). *Student outcome analysis: Early signs of student growth from professional learning in Louisiana and Michigan.* Downloaded March 19, 2021 from https://leadingeducators.org/wp-content/uploads/2020/02/LouisianaMichiganStudentImpactAnalysisResults.pdf

Lipsky, M. (1971). Street-level bureaucracy and the analysis of urban reform. *Urban Affairs Quarterly,* 391-409.

Mangin, M. & Stoelinga, S. R. (Eds.) (2008). *Effective teacher leadership: using research to inform and reform.* Teachers College Press: New York.

Margolis, J. (2012). Hybrid teacher leaders and the new professional development ecology. *Professional Development in Education, 38*(2), 291-315.

Margolis, J. (2020). The semiformality of teacher leadership on the edge of chaos. *Harvard Educational Review, 90*(3): 397-418.

Margolis, J, & Deuel, A. (2009). Teacher leaders in action: motivation, morality, and money. *Leadership and Policy in Schools*, *8*(3): 264-286. d

Marks, H. & Louis, K. S. (1997). Does teacher empowerment affect the classroom? The implications of teacher empowerment for instructional practice and student achievement performance. *Educational Evaluation & Policy Analysis*, *19*(3), 245-275.

Marsh, J. A. & Hall, M. (2018). Challenges and choices: A multidistrict analysis of statewide mandated democratic engagement. *American Educational Research Journal, 55*(2), 243-286.

Mayer, A. P., Donaldson, M. L., LeChasseur, K., Welton, A. D., & Cobb, C. D. (2013). Negotiating site-based management and expanded teacher decision-making: A case study of six urban schools. *Educational Administration Quarterly, 49*(5), 695-731.

Muijs, D. & Harris, A. (2003). Teacher leadership – improvement through empowerment?: An overview of the literature. *Educational Management & Administration, 31*(4), 437-448.

New York City Department of Education. (2022). Test Results. https://infohub.nyced.org/reports/academics/test-results

Rice, E. M. & Schneider, G. T. (1994). A decade of teacher empowerment: An empirical analysis of teacher involvement in decision-making, 1980-1991. *Journal of Educational Administration, 32*(1), 43-58.

Rios, F. (2021, September 2). Cannot really say much about why Peter says that. But, My understanding on this approach is as you say. Avoiding [Comment on the online forum post Diff in Diff: DRDID and CSDID.]. Statalist: The Stata Forum. https://www.statalist.org/forums/forum/general-stata-discussion/general/1622165-diff-in-diff-drdid-and-csdid

Robertson, P., Wohlstetter, P., & Mohrman, S. (1995). Generating curriculum and instructional innovations through school-based management. *Educational Administration Quarterly, 31*(3), 375-404.

Sands, S. (2023). Complexity in action: Evaluating teacher leadership through the lens of complexity theory. SSRN.

Sands, S. (2023). *Leading from the classroom: Teacher leadership in New York City schools.* PhD Thesis, Teachers College, Columbia University, USA.

Sands, S., Margolis, J. & Murphy, M. J. (2022). Understanding teacher leadership in New York City through complexity theory. *International Journal of Complexity in Education, 3*(1), 1-32. Retrieved from: https://drive.google.com/file/d/1jq99CP8C5ZizK5hwUfE8aHcyIxH0BQGW/view

Schueler, B. & Bleiberg, J. (2021). Evaluating education governance: Does state takeover of school districts affect student achievement? *Journal of Policy Analysis and Management, 41*(1), 162-192.

Smith, B. (2019). Leadership for “the all of it”: Formalizing teacher-leader networks. *Teacher College Record*, *121*(030308).

Sun, L. & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175-199, https://doi.org/10.1016/j.jeconom.2020.09.006.

Supovitz, J. A. (2018). Teacher leaders’ work with peers in a quasiformal teacher leadership model, *School Leadership & Management, (38)*1, 53-79.

Supovitz, J. A., & Comstock, M. C (2023). The impact of a formal teacher leadership program on student performance. *Education Policy Analysis Archives, 31*(74).

Sykes, G. (1990). Fostering teacher professionalism in schools. In R. Elmore (Ed.), *Restructuring schools: The next generation of educational reform* (p. 59-96). San Francisco: Jossey-Bass, Inc.

Taylor, M., Goeke, J., Klein, E., Onore, C., & Geist, K. (2011). Changing leadership: teachers lead the way for schools that learn. *Teaching and Teacher Education, 27*, 920-929. doi: 10.1016/j.tate.2011.03.003

Teacher Career Pathways. (2020). "Peer Collaborative Teacher: Responsibilities, benefits, and qualification overview." New York City Department of Education. https://files.constantcontact.com/a7f62c21301/df2c939d-52d1-4cd8-8561-a7ce173d2cec.pdf

Teacher Career Pathways. (2024). "About TCP." https://www.teachercareerpathways.com/about/about-tcp. Accessed June 24, 2024.

Wenner, J. & Campbell, T. (2017). The theoretical and empirical basis of teacher leadership: a review of the literature. *Review of Educational Research*, *87*(1): 134-171.

Varadhan, R. & Seeger, J. D. (2013). Estimation and reporting of heterogeneity of treatment effects. In P. Velentgas, N. A. Dreyer, P. Nourjah (Eds)., *Developing a protocol for observational comparative effectiveness: A user's guide*. Agency for Healthcare Research & Quality. https://www.ncbi.nlm.nih.gov/books/NBK126188/#:~:text=Heterogeneity%20of%20treatment%20effect%20(HTE,for%20individuals%20within%20a%20population.

Weill, J. A., Stigler, M., Deschenes, O., & Springborn, M. R. (2021). Researchers' degree-of-flexibility and the credibility of difference-in-differences estimates: Evidence from the pandemic policy evaluations. *National Bureau of Economic Research Working Paper Series, No. 29550.*

Woods, P. A. & Gronn, P. (2009). Nurturing democracy: The contribution of distributed leadership to a democratic organizational landscape. *Educational Management, Administration, & Leadership, 37*(4): 430-451.

Woods, P. A. (2011). *Transforming education policy: Shaping a democratic future*. Bristol: The Policy Press.

York-Barr, J. & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research, 74*(3), 255-316.

Yost, D. S., Vogel, R., & Liang, L. L. (2009). Embedded teacher leadership: support for a site-based model of professional development. *International Journal of Leadership in Education, 12*(4), 409-433.