This is the final accepted unformatted version of Burns, J.A., **Harbatkin, E.**, Strunk, K.O., Torres, C., McIlwain, A., Waldron, S.F. (2023). The Efficacy and Implementation of Michigan's Partnership Model of School and District Turnaround: Mixed-Methods Evidence from the First Two Years of Reform Implementation. *Educational Evaluation and Policy Analysis,* 45(4), 622 -654. https://doi.org/10.3102/01623737221141415.

The Efficacy and Implementation of Michigan's Partnership Model of School and District Turnaround: Mixed-Methods Evidence from the First Two Years of Reform Implementation

Jason Burns[1], Erica Harbatkin[2], Katharine O. Strunk[3], Chris Torres[4], Aliyah Mcilwain[3], & Sandy Frost Waldron[3]

[1] Seton Hall University and the Education Policy Innovation Collaborative (EPIC), [2] Florida State University and EPIC, 3 Michigan State University and EPIC, [4] University of Michigan and EPIC

Abstract. The recent Every Student Succeeds Act (ESSA) requires states to identify and turn around their lowest performing schools, but breaks somewhat from prior policies by granting states significant autonomy over how they identify and turn around these schools. This mixed-methods study, which draws on administrative, qualitative, and survey data, examines the effectiveness of Michigan's approach to school turnaround under ESSA. We find that students in turnaround schools experienced significant achievement gains in math and to a lesser extent in ELA, with effects concentrated among the lowest achieving students. Analyses of qualitative and survey data suggest that these outcomes were influenced by state-level supports, strategic planning, the threat of accountability for continued low performance, and improved leadership quality in turnaround schools.

We are grateful to the many people who gave their time in support of this work, including research participants, and those who provided technical and research support: Sheila Alles, Gloria Chapman, Dr. Paula Daniels, Dr. Venessa Keesler, Dan LaDue, Dr. William Pearson, Dr. Michael Rice, and the entire staff of the MDE Office of Partnership Districts; Trina Anderson, Rod Bernosky, Tom Howell, and Mike McGroarty from the Center for Educational Performance and Information (CEPI); and EPIC staff including Emily Hatch, Michelle Huhn, Chayse Hurley, Dr. Tara Kilbride, Emily Mohr, and Meg Turner. We would also like to thank the funders of this research: Arnold Ventures, an anonymous foundation, the Michigan Department of Education, and Michigan State University's College of Education. This research used data structured and maintained by the Michigan Education Research Institute (MERI)-Michigan Education Data Center (MEDC). MEDC data are modified for analysis purposes using rules governed by MEDC are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information, and opinions represent the analysis, information, and opinions of the author(s) and are not endorsed by or reflect the views or positions of, grantors, MDE, CEPI, or any employee thereof. All errors are our own.

The Efficacy and Implementation of Michigan's Partnership Model of School and District Turnaround: Mixed-Methods Evidence from the First Two Years of Reform Implementation

Introduction

Over the last two decades, federal education policies such as the No Child Left Behind Act (NCLB), the School Improvement Grant program (SIG), and Race to the Top (RTTT) have centered turnaround as an important strategy for states to improve their lowest performing schools. Similarly, the Every Student Succeeds Act (ESSA), passed in 2015, requires that states continue to identify their lowest performing schools and use evidence-based practices to turn around and improve outcomes in these schools. A growing body of evidence points to positive average effects of turnaround reforms, though there is substantial heterogeneity across contexts and turnaround strategies (Redding & Nguyen, 2020; Schueler et al., 2020). Although the overall effects of turnaround are becoming better understood, relatively few empirical studies on the effects of turnaround have also examined the way(s) in which reforms are implemented (for exceptions, see Dragoset et al., 2015, 2017; Hallgren et al., 2019; Marsh et al., 2013; Papay et al., 2021; and Strunk et al., 2016). Moreover, the turnaround literature has thus far examined school and district reforms under earlier iterations of federal law; as far as we know, no empirical research to date has examined the effect of turnaround under ESSA.

There is reason to believe that turnarounds under ESSA may differ from those enacted under earlier federal policies. Unlike SIG and RTTT, which prescribed specific turnaround models, ESSA allows states broad autonomy in turning around their lowest performing schools. State ESSA plans have outlined varying approaches to school turnaround in terms of district and state roles, identifying schools for intervention, and improving teacher quality (Close et al., 2019; Dunn & Ambroso, 2019; Portz & Beauchamp, 2020). One characteristic all state turnaround strategies must share is a comprehensive planning process; ESSA requires that

districts develop an improvement plan in low-performing schools that is informed by school performance data and a school-level needs assessment.

In this paper, we draw from statewide administrative data, educator surveys, and interviews with leaders of turnaround districts and charter organizations to investigate the implementation and effects of Michigan's school turnaround strategy under ESSA, called the Partnership Model. Michigan first implemented the Partnership Model in 37 chronically low-performing schools in the 2017-18 school year and identified two more rounds of schools that entered turnaround in 2018-19. Although the state used different mechanisms to select each of the three rounds of Partnership schools, each round reflected the state's identification of its lowest performing schools at the time of identification. We examine the effects of the Partnership Model across all three sets of selected schools. Specifically, we ask:1) *What was the impact of the Partnership Model on student and school outcomes?;* 2) *Did the Partnership Model's impacts differ for lower- and higher-achieving students?;* and 3) *In what ways did schools and districts implement the Partnership Model to influence outcomes?*

In what follows, we begin by briefly reviewing the literature on school and district turnaround in the post-NCLB era and then describe the Partnership Model. Next, we describe our quantitative and qualitative data and methods. We then summarize results and conclude with a discussion and policy implications for school turnaround under ESSA.

Literature Review

The Every Student Succeeds Act (ESSA) marks a departure from prior rounds of federal turnaround reform, with four features in particular setting it apart. The first three follow the blueprint laid out by Elementary and Secondary Education Act (ESEA) waivers, which gave states expanded flexibility to approach turnaround in the period after the RTTT and SIG

interventions and before the adoption of ESSA. First, ESSA does not mandate specific turnaround models. Second, it devolves the responsibility of turnaround design from the federal government to state and local education agencies, requiring states to identify Comprehensive Support and Improvement (CSI) schools and then calling for local education agencies to work with stakeholders to implement an improvement plan. This policy design implicitly recognizes the importance of the district in leading and promoting school improvement, and several states have responded by elevating the district role (Dunn & Ambroso, 2019; Fullan, 2016; Hopkins et al., 2014; Karcher & Knight, 2021; Meyers, 2020). Importantly, this approach to turnaround differs from state takeovers of districts, which have shown some positive but on average null effects across the country (Schueler et al., 2017; Schueler & Bleiberg, 2022). Instead, turnaround under ESSA entrusts more autonomy to district leaders rather than assigning a third party to oversee the district more closely. Third, ESSA does not allocate substantial additional funding for school turnaround. Finally, ESSA requires states to identify low-performing schools using a multidimensional measure of school performance rather than through proficiency rates alone. By contrast, prior waves of federal accountability policy have either explicitly focused on proficiency targets (i.e., NCLB) or states have implemented systems centered around proficiency rates (i.e., RTTT, SIG, and ESEA waiver Priority schools). This expanded definition of school performance may change the ways states and districts approach turnaround because it changes the nature of the outcomes they are trying to improve.

Effects of Prior Reforms

The shift in school turnaround and accountability under ESSA comes on the heels of a string of reforms that produced positive average effects, but with substantial heterogeneity across contexts (Redding & Nguyen, 2020; Schueler et al., 2021). Here, we discuss prior reforms

beginning with NCLB and culminating with ESEA waivers. There is some evidence that NCLB interventions improved student achievement and closed achievement gaps (Dee & Jacob, 2011), though effects were uneven across grade levels, subject areas, and state contexts (Dee & Jacob, 2011; Ladd, 2017; Lee & Reeves, 2012; Strunk et al., 2014a; Strunk & McEachin, 2014). However, NCLB's emphasis on proficiency rates also created perverse incentives for schools to focus on so-called "bubble" students to the detriment of students throughout the achievement distribution, or to push out students expected to score below proficient. In particular, improvements in student achievement and decreased achievement gaps were concentrated among students near the cusp of the test proficiency threshold—suggesting districts may have triaged supports to the students most likely to increase proficiency rates (Balfanz et al., 2007; Booher-Jennings, 2005; Cullen & Reback, 2006; Darling-Hammond, 2006; Ho, 2008; Jacob, 2005; Jennings & Bearak, 2014; Krieg, 2008; Reback, 2008; Strunk et al., 2014).

NCLB interventions gave way to RTTT and SIG whole-school reform interventions, which introduced school and district turnaround to the federal policy parlance around low-performing schools. Unlike prior iterations of whole-school reform such as those taking place under NCLB (and previously the Comprehensive School Reform program), turnaround aimed to impel rapid and dramatic changes in low-performing schools (Aladjem et al., 2010; Herman et al., 2008; Peurach & Neumerski, 2015). Additionally, federal turnaround policy prescribed specific strategies for states to follow—providing considerably less discretion to state and local education leaders than prior federal policy. Turnaround schools in Massachusetts, California, Ohio, and Tennessee made student achievement gains under RTTT and SIG (Carlson & Lavertu, 2018; LiCalsi et al., 2015; LiCalsi & Píriz, 2016; Papay & Hannon, 2018; Sun et al., 2017; Zimmer et al., 2017). Some reforms also produced positive effects on proximate outcomes such

as increased teacher retention and quality, and reduced student absenteeism (Papay & Hannon, 2018; Pham et al., 2020; Sun et al., 2017). Across multiple states, the positive effects of SIG were largely sustained after the reforms concluded (Sun et al., 2020).

However, positive average treatment effects mask heterogeneity both across and within interventions. Turnaround in North Carolina, Tennessee, and Texas produced mixed effects (Dickey-Griffith, 2013; Heissel & Ladd, 2018; Henry & Guthrie, 2019; Zimmer et al., 2017), and some of the interventions yielding positive overall effects also produced null or negative effects in particular contexts or cohorts (Carlson & Lavertu, 2018; Strunk et al., 2016; Zimmer et al., 2017). The heterogeneity of effects highlights variation both in the features and implementation of state turnaround interventions. Even under SIG, which prescribed specific turnaround models and practices, there was variation in the practices states chose to implement (Dragoset et al., 2015).

As RTTT and SIG grants concluded, states replaced their tightly mandated turnaround interventions with more flexible policies under ESEA waivers, which required states to identify and support their low-performing schools but—like ESSA—did not mandate how states should do so. These waiver-based reforms yielded few positive effects, though the reform models themselves and their implementation again varied widely across states. Of the states with evaluations of waiver reforms, one—Kentucky—produced positive effects on student achievement, which the authors attributed in part to clearly articulated state guidance for the comprehensive school improvement planning process (Bonilla & Dee, 2020). Waiver-based reforms in New York, Michigan, Rhode Island, and Louisiana produced either null or negative effects on student achievement (Atchison, 2020; Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2017, 2018). A similar state-initiated reform in North Carolina that

focused on tailoring turnaround services to school needs also led to negative effects on student achievement, early grade outcomes, and teacher retention (Henry et al., 2022; Henry & Harbatkin, 2020). By contrast, an Ohio reform that involved supporting district and school leaders in developing comprehensive plans produced positive effects on school-level student achievement (Player & Katz, 2016). In Washington state, strategies delineated in school improvement plans under SIG and ESEA waivers predicted improvements in student achievement and absenteeism (Sun et al., 2019). These findings highlight the importance of the comprehensive planning process in more loosely mandated turnaround reforms like ESEA waivers and the new wave of reforms under ESSA.

As in other states, Michigan's reforms under these earlier federal policies produced mixed effects. Students in Priority schools, which comprised the bottom 5% of schools under the state's ESEA waiver reform, fared worse than their counterparts in other low-performing schools in math and reading in both the first and second years of the intervention (Hemelt & Jacob, 2017). In Focus schools (schools with the widest achievement gaps between high and low achieving students), a short-run decrease in the math achievement gap was driven by stagnant performance among high achieving students rather than growth among lower achievers. There were no effects on the reading achievement gap, or math or reading performance (Hemelt & Jacob, 2018).

School Turnaround under ESSA

It is unclear how evidence stemming from prior reforms will generalize to ESSA turnaround interventions. While there were some positive effects of prior turnaround efforts, several studies suggest schools that were subject to more disruptive reform models under RTTT and SIG fared better than schools under less disruptive models (Carlson & Lavertu, 2018; Dee,

2012; Strunk et al., 2014; Sun et al., 2017). It is thus far unclear whether and to what extent a flexible and tailored turnaround model in combination with a more nuanced view of school performance will translate into improved student outcomes. As no peer-reviewed studies have yet examined school turnaround policies under ESSA, little is known about whether these new features are conducive to improved student outcomes.

However, prior literature points to four mechanisms through which less disruptive ESSA reforms like the Partnership Model could impact student achievement: the comprehensive planning process; external supports; the threat of accountability; and improvements to human capital. First, the quality of comprehensive plans can vary and their content plays an important role in reform implementation (Anfara et al., 2006; Mintrop & MacLellan, 2002; Redding & Searby, 2020; Strunk et al., 2016). This is important because a comprehensive planning process that generates clearly articulated plans has the potential to induce meaningful improvements in low-performing schools and districts—while a pro forma approach may not (Bonilla & Dee, 2020; Huber & Conway, 2015; Player & Katz, 2016; Strunk et al., 2016; Sun et al., 2019).

Second, the scope and quality of external supports and resources made available to turnaround schools and districts matter. Schools may be low performing because they lack the resources needed to build the systems that undergird improvement (Karcher & Knight, 2021). While early attempts at school reform focused largely on accountability, RTTT and SIG paired accountability with an influx of federal funding for states to support low-performing schools. Positive effects under RTTT and SIG were unique in both their prevalence and magnitude, suggesting that these new resources may have played an important role (see Carlson & Lavertu, 2018; LiCalsi et al., 2015; LiCalsi & Píriz, 2016; and Papay & Hannon, 2018). Third, there is evidence from the NCLB era that the possibility of high-stakes consequences for failing to make

reasonable progress can induce improvement (Carnoy & Loeb, 2002; Chiang, 2009; Dee & Jacob, 2011; Hanushek & Raymond, 2005; Jacob, 2005). Under NCLB in particular, the threat of sanctions appeared to increase student achievement in some contexts (Reback et al., 2014; Springer, 2008).

Finally, an abundance of research highlights the importance of human capital to school effectiveness and improvement. A stable and highly effective teacher and principal workforce is critical to successful turnaround (Harbatkin, 2022; Henry et al., 2020; Pham et al., 2020). When teachers and principals do turn over, schools need a robust local educator labor market from which to draw replacements (Malen & Rice, 2016; Strunk et al., 2016). Principals in particular are responsible for building the framework for school improvement, for example by setting school climate and culture, coalescing staff around a shared mission, and acting as instructional leaders (Burkhauser, 2017; Duke, 2004; Duke & Salmonowicz, 2010; Hitt et al., 2018; Leithwood & Jantzi, 1990; Weiner, 2016; Woulfin & Weiner, 2019).

While prior research points to these mechanisms as important for successful school improvement under earlier reforms, there is thus far no evidence on how they may contribute to turnaround under ESSA. This paper adds to the turnaround literature in three ways. First, we examine implementation in addition to impact to unpack not just *whether* Partnership turnaround affected school outcomes but also *how* it affected them. There are, to date, few studies that combine both impact and implementation analyses (see Henry & Harbatkin, 2020; Papay et al., 2021; and Strunk et al., 2016 for notable exceptions). Second, we provide early evidence on the implementation and effects of a generalizable school turnaround reform under ESSA. Though its implementation began a year before states were required to begin intervening in low-performing schools under ESSA, the model is the state's intervention under ESSA, meets ESSA

requirements for identifying and supporting low-performing schools, and is similar to ESSA turnaround models throughout the country in its focus on customizing supports to school and district-specific needs. Thus, results are likely generalizable to ESSA turnaround reforms across the country. Finally, we examine the effects of ESSA turnaround on low-achieving students, who tended not to improve under prior accountability interventions that classified schools using proficiency rates. While previous research has estimated the effects of accountability interventions on the lowest achieving students, this study builds on prior work by examining the effects of a turnaround reform that defines progress using a multidimensional measure of school effectiveness rather than a singular focus on proficiency rates.

Michigan's Partnership Model of School and District Turnaround

Michigan's Partnership Model of School and District Turnaround began in the spring of 2017. Under Michigan law in effect at the time, the state could force schools that consistently performed in the bottom 5% of schools statewide to close. In January of 2017, the state identified schools across nine districts that were in the bottom 5% for three consecutive years and announced that they may face closure.[1] This move provoked public outcry, with several districts filing lawsuits challenging the State's authority to close their schools. At the time, officials in the Michigan Department of Education (MDE) were working on Michigan's ESSA plan that was to be implemented that fall. In light of the political and legal challenges around closing persistently low-performing schools, the governor and state superintendent of public instruction reached an agreement that moved up the timeline for implementing MDE's plan to support low-performing schools and districts and the schools slated for possible closure were instead provided with supports and accountability under the new Partnership Model. As of 2020, a total of 119 schools across 36 districts have become Partnership schools through three rounds of identification, each

employing a slightly different method of selecting schools for Partnership, that took place in spring 2017, fall 2017, and spring 2018.[2]

Figure 1 shows the theory of change guiding the Partnership Model. The primary goal of the Partnership Model is to improve student outcomes in low-performing schools. At the same time, the Partnership Model conceives of low performance at the school level as symptomatic of issues at the district level and so has a secondary goal of improving capacity in the districts that operate low-performing schools. As such, the Partnership Model aims to improve student outcomes in identified schools by supporting the local educational agencies (LEAs) that operate them. This approach to turnaround follows the intended theory of action under ESSA, which relegates the responsibility of designing turnaround strategy from the federal government to states and districts (Dunn & Ambroso, 2019).

After a school is identified as low performing, it is designated as a Partnership school and the LEA that operates it is designated as a Partnership district to enact and oversee the turnaround of its Partnership school(s). Both traditional public schools (TPSs) and charter schools (called public school academies in Michigan) can be identified for turnaround. In the case of TPS Partnership schools, their parent district is charged with leading their turnaround. In the case of charters, different entities may serve as the Partnership district, often their central office or educational service provider – sometimes referred to as an educational management organization or management company.

After a district is identified for Partnership, a series of supports become available to guide turnaround work in identified schools. State-level supports include a Partnership Agreement liaison from MDE to provide individualized supports to district leadership along with access to grants available only to Partnership districts. At the regional level, districts receive greater

support from their intermediate school district (ISD), entities which generally coincide with counties and are intended to provide local districts with services that MDE does not. District-ISD relationships vary considerably, but districts may contract with ISDs around special education services, professional development, data services, and a range of other supports. Through the Partnership Model, MDE directs additional funding to ISDs that serve Partnership districts to provide extra supports for identified Partnership schools. In most cases, this funding was used to provide Partnership districts with additional professional development and various forms of coaching related to the needs identified by the district. At the local level, Partnership districts are encouraged to reach out to community organizations, such as civic groups, the local business community, and community health agencies, for additional supports. The original theory of change conceived of local partners as key players in Partnership districts' turnaround efforts, but their role was de-emphasized during implementation.

Drawing on these supports, Partnership districts develop a Partnership Agreement that serves as a contract with MDE to improve student outcomes in each of its Partnership schools over a 36-month period. As with other ESSA-based turnaround models, this process begins with the district conducting a comprehensive needs assessment to identify its own, and its Partnership schools', strengths and weaknesses. That needs assessment is used to identify, for each Partnership school: academic and non-academic improvement goals to be met over 18- and 36-month time periods; strategies aligned with those improvement goals; supports the district will receive from state, regional, and local partners; and accountability measures to be implemented if improvement goals are not met. After the Agreement is drafted, it must be formally approved by the district, its ISD, MDE, and community partners. Once approved, the Agreement is implemented beginning in the following school year. To monitor progress toward the goals

outlined in their Partnership Agreements, MDE conducts an interim evaluation 18 months into implementation, during which MDE gauges each Partnership school's progress toward its improvement goals and then uses its assessment to rate schools' progress. Partnership districts that are found to be "off-track" are required to implement additional improvement strategies as well as undergo another evaluation at 24 months of implementation. Partnership districts' final evaluations are conducted at 36 months, at which point schools that meet their goals are released from Partnership status and the accountability measures spelled out in the Agreement are administered in schools that fail to meet their goals.

As the bottom of Figure 1 shows, a secondary aim of the Partnership Model is to foster improvement in district-level systems that in turn supports sustained improvement at the school level, particularly in identified Partnership schools. This should then lead to intermediate outcomes such as increased educator retention and higher quality instruction and eventually in improved academic and whole-child outcomes for students. Ideally, this improvement will then continue beyond the time period covered by the Partnership Agreement.

Sample, Data, and Methods

To evaluate the Partnership Model, we use a mixed-methods triangulation design (Creswell & Plano Clark, 2017) that includes multiple types of data and methods of analysis. A triangulation design is well-suited to an evaluation of an intervention as complex as the Partnership Model because it allows us to assess results through multiple sources of data and methodological strategies. By integrating analyses of varied sources of qualitative and quantitative data we are able to paint a rich picture of how this reform is being implemented across districts and schools, and the effectiveness of the reform along its intended outcomes. We employ this triangulation design by first using a series of event study models to answer research

questions 1 and 2, estimating the effect of the Partnership Model on student outcomes overall and for subgroups of students by prior achievement. Next, to answer research question 3, we draw on qualitative and survey data to examine how implementation of the Model at the school and district levels shaped these effects.

Administrative Data

Sample. The event study draws from statewide administrative data on students and teachers from 2013-14 through 2018-19 provided by the Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI). This includes approximately 9 million student-year observations representing 2,244,145 students. To examine the impact of Partnership on student outcomes, we limit the student sample to just those who were enrolled in a Partnership school during these years and identify students in Partnership schools as the treated group and students enrolled in near-selected schools for each cohort as the comparison group. This analytic sample contains 971,656 student-year observations representing 306,307 unique students. Our main models further restrict this sample by identifying students as being either in the treatment or comparison group based only on the school they attended in the year of Partnership identification (see Table 1 for counts of students and schools by treatment condition in the identification year). By classifying students as treatment or comparison based on their school assignment at the time of Partnership identification, we are estimating an "intent-to-treat" (ITT) effect of Partnership based on school assignment prior to the intervention. This time-invariant treatment assignment provides the most conservative effect estimates by avoiding selection bias from families choosing to transfer students in or out of Partnership schools in response to the intervention.[3]

Thus far, there have been three rounds of Partnership identification and the analytic sample therefore contains three rounds of treated schools. We provide detail on how each round of Partnership schools was selected in online appendix B. Because the state identified Rounds 2 and 3 in the same school year and treated them on the same implementation and evaluation schedules, we combine them into a single treatment group we call Cohort 2. We therefore separately examine each cohort of schools—Cohort 1 (comprising 37 Round 1 schools) and Cohort 2 (comprising 39 Round 2 and 43 Round 3 schools for a total of 82 schools). Because Rounds 2 and 3 schools were identified using different metrics and six months apart, we also conduct supplemental analyses to unpack differences in effects between these two rounds.

For each cohort, we construct a comparison group of near-selected low-performing schools that were never identified for Partnership. The Cohort 1 comparison group comprises the 80 Priority schools from 2015-16 that were not selected as Partnership schools in Rounds 1, 2, or 3. The Cohort 2 comparison group includes the 156 schools in the bottom 10% on Michigan's ESSA index system in 2017-18 that were not identified for treatment in any round of Partnership.[4]

Treated schools in both Cohort 1 and Cohort 2 are, on average, lower performing than the schools in their respective comparison groups, which is not surprising given that schools were selected for Partnership on the basis of low performance. Online appendix Table A-2 shows that treated and comparison schools have similar shares of students who are economically disadvantaged and that treated schools have larger shares of students who are Black, slightly larger shares of students who receive special education services, and lower shares of students who are White, Hispanic or Latinx, and English Language Learners. The final two columns of Table A-2 make clear that comparison schools are more observably similar to treated schools

than other schools throughout the state. As we describe later, the causal interpretation of the estimates in our event study do not rely on baseline equivalence but rather on parallel pre-treatment trajectories on the outcome and evidence that there was no anticipatory response to treatment assignment. Collectively, data used for analyses of Cohort 1 include 414,456 student-year observations (149,135 unique students). Cohort 2 data include 689,407 student-year observations (230,240 unique students).[5]

While the decision to assign treatment status based on students' location in the identification year most cleanly addresses selection issues that are endogenous to Partnership identification, the estimates from these models may be attenuated for two reasons. First, students in a Partnership school during the identification year could, during implementation, move to either a comparison school in our sample or to a non-comparison school that we do not include in our sample. Similarly, students in comparison schools during the identification year may transfer to Partnership schools (or non-comparison, non-treated schools) after identification. To that end, not all students classified as treated will have received the full dose—or in some cases, any dose—of treatment, and some students classified as comparison will have received some dose of treatment. Second, some schools in the comparison group may receive some element of the treatment being that Partnership is implemented through districts with a particular focus on targeted schools within them. To the extent that Partnership contributes to improved district-level systems and functioning, comparison schools in Partnership districts may improve alongside treated schools.

We estimate models on three additional populations to examine the extent to which each of these potential sources of attenuation bias may contribute to our preferred estimates. First, we employ Sun et al.'s (2017) "all starters" approach by assigning students to treatment or

comparison based on the school they attended during the first implementation year rather than the identification year. For Cohort 1, this approach creates a sample based on school assignment in the 2017-18 school year and follows the same cohort of students into the next year. For Cohort 2, where we have only one year of outcome data, this approach produces a treatment-on-the-treated (TOT) estimate because it estimates the effect of treatment based on a single year of implementation.[6] To the extent that our preferred estimates are attenuated by student mobility, the all starters models will find larger effects.

Second, we return to the original student assignment approach (i.e., treatment assignment based on school in the identification year) but drop non-Partnership schools in Partnership districts from our comparison group. This approach drops 16,515 and 22,943 students in 29 and 42 schools from the Cohort 1 and 2 comparison groups, respectively. To the extent that the Partnership Model improved district-level systems that led to increases in student achievement for students in low-performing, non-Partnership schools in Partnership districts, these secondary effects would cause estimated effects of the intervention to be larger than in our preferred models. Thus, both specification tests should confirm that our preferred approach returns the most conservative estimate of Partnership Model impact.

Finally, we estimate models in which we move students in non-Partnership schools in Partnership districts (i.e., the students we dropped out in the prior specification) from the comparison group to the treated group to more directly estimate the effects of Partnership on untreated low-performing schools in Partnership districts. This approach allows us to estimate the effect of Partnership on students in all low-performing schools in Partnership districts, including schools identified for Partnership and students in near-selected schools.

Outcome measures. The primary outcome of interest is student performance on state standardized tests in mathematics and English language arts (ELA) given to all fourth- through eighth- and eleventh-grade students each year. Additionally, we examine SAT achievement, high school graduation, and high school dropout. We provide additional details for these measures in online appendix C.

Lagged achievement. We classify students as low or higher achievers within their school based on their performance on the Michigan Student Test of Educational Progress (M-STEP), the suite of assessments administered to Michigan students in grades 3-8, in the identification year for their cohort. Specifically, we classify low achievers as those in the bottom quartile of their school in the identification year and higher achievers as the top three quartiles. While other research has focused on bubble students and examined heterogeneity in lagged performance based on distance from the proficiency threshold (e.g., Booher-Jennings, 2005; Strunk et al., 2014), we choose to examine the difference between low- and high-achievers within their own schools for two reasons. First, the vast majority of students in Partnership schools score below proficient, making the "bubble" students-type analyses less relevant in this instance.[7] Second, we want to understand how treated schools and districts may or may not provide greater services to, or focus on, the lowest performing students within their own schools and districts. Unlike NCLB, which targeted schools based on proficiency rates alone, the more expansive view of achievement under ESSA and Partnership may lead to less focus on moving students over proficiency thresholds.

Estimation Strategy

We estimate event study models separately for the two cohorts of treated and comparison schools. The model for student outcomes (i.e., grades 4-8 test scores) in Cohort 1 takes the form

$$y_{igst} = \alpha_0 + \sum_{r=-3}^{2} I_{2016+r} + \sum_{r=-3}^{2} I_{2016+r} \times Partnership_{ist} + X_{ist}\theta + \delta_g + \psi_i + \varepsilon_{igst}$$

estimating outcome *y* (i.e., test score level or test score gains) for student *i* in grade *g* in school *s* in year *t*. I_{2016+r} represents a series of year indicators from 2013-14 through 2018-19, with 2016-17 (the year of Partnership identification for Cohort 1) as the omitted reference year. We interact these year indicators with a binary indicator of treatment status, $Partnership_{ist}$ that takes a value of 1 for students who were assigned to Partnership schools at the time of identification. The coefficients on these interactions provide the estimated difference in outcomes between students in Partnership and comparison schools relative to the omitted 2016-17 school year. The interactions for 2017-18 and 2018-19 provide the estimated effects of the Partnership reform for Cohort 1 in the first two years of implementation. $\mathbf{X}_{ist}$ is a vector of time-varying student and school covariates. Student covariates include indicators denoting whether the student was economically disadvantaged,[8] an English learner, and had a special education designation. School-level covariates include the school-level proportions of each race/ethnicity, economically disadvantaged, English learner, and special education, as well as logged school enrollment. We also include grade-level fixed effects (δ_g) with fourth grade as the reference category for test score outcomes,[9] and student fixed effects (ψ_i), which allow us to estimate each student's shift from their own pretreatment performance. ε is an idiosyncratic error term clustered at the school level. For fourth- through eighth-grade achievement, we estimate two separate models—one predicting the effect of Partnership on achievement levels and one predicting the effect on one-year achievement gains. We interpret these estimates as bounding the true effect of Partnership (Imberman, 2010), though neither model should be interpreted as always providing the upper or lower bound. The model predicting student outcomes for Cohort 2 follows the same format as

the model for Cohort 1 except it omits 2017-18 (the Cohort 2 identification year) as the reference year and therefore provides only one year of effect estimates. In this case, the I_{2016+r} indicators become I_{2017+r} and *r* indexes to 1 instead of 2.[10]

To estimate heterogeneous effects by lagged achievement, we estimate separate models for low and higher achievers based on their test scores in a particular subject in the identification year. Estimates of school-level outcomes (i.e., SAT scores, graduation rate, dropout rate) largely follow the model above, but exclude student covariates and substitute school fixed effects for student fixed effects. The school-level models use treatment assignment based on the school's status (treatment or comparison) in the relevant identification year. In other words, a school identified for Partnership in year *t* would be assigned to the treatment group and retain that assignment throughout the panel and the graduation rate or other school-level outcome observed for that school in year $t+1$ is based on all students in that school in that year regardless of the school those students attended in year *t.*

There are two key identifying assumptions in our event study design. The first is that in the absence of Partnership identification, outcomes in Partnership schools would have followed the same pattern as those observed in our comparison schools, conditional on covariates. We evaluate this assumption by interpreting the coefficients and p-values for the $\sum_{r=-3}^{2} I_{2016+r} \times Partnership_{ist}$ terms in the years leading up to Partnership identification. Here, coefficients that are small and not statistically significant suggest that outcomes in Partnership and comparison schools were similar leading up to Partnership identification (see, for example, the rows of Table 2 that correspond to the years prior to Partnership identification, indicated by "ref."). The second is the conditional no anticipation assumption, which assumes that students in treated and comparison groups did not respond differentially in anticipation of the treatment, again

conditional on covariates. To address concerns related to this assumption, we also estimate models with the year prior to identification as the excluded reference year. Again, we do not observe significant differences between treated and comparison schools in the pretreatment years (these estimates are provided in online appendix Table A-7). These models can provide further evidence that the estimates are not being driven by mean regression following significant identification year dips.

We conduct several additional analyses to examine the robustness of our results. First, estimates from the two-way fixed effects approach can be sensitive to violations of strict exogeneity of covariates (Baker et al., 2022; Wooldridge, 2021). We therefore re-estimate our models using covariates measured at baseline, interacted with a linear year trend, rather than time-varying covariates. Second, while our estimates will not be biased by staggered intervention timing, our main analyses do combine two identification rounds (Rounds 2 and 3) into a single implementation cohort (Cohort 2) because they implemented on the same timeline. Due to differences in how these schools were identified, there may be concerns that a common treatment effect could obscure meaningful differences in outcomes. We therefore re-estimate our Cohort 2 models allowing for flexible effects in Rounds 2 and 3.

Three additional analyses are focused on the district-level nature of the intervention. Specifically, while treatment is assigned at the school level, it is implemented through the district and could therefore be considered as a district-level treatment. First, we estimate three different models to assess the extent to which the district-level nature of the treatment may be affecting (a) outcomes in low-performing non-Partnership schools in Partnership districts, (b) estimates for treated schools, and (c) inference on the coefficient estimates in our main models. To estimate the effects of the Partnership Model on low-performing non-Partnership schools in Partnership

districts, we run a parallel set of event studies in which we move all students in non-Partnership schools in Partnership districts from the comparison to treated group. Specifically, we count a student as being treated if they are in a near-selected school in a Partnership district, regardless of that school's Partnership status. In the post-treatment years, we then estimate two separate treatments—one for Partnership schools as in our main models, and another for near-selected schools in Partnership districts. These models provide the estimated effects of being assigned to a Partnership school and being assigned to another low-performing school in a Partnership district. If Partnership improved district processes and systems or allocated resources across Partnership *and* other low-performing schools, these estimates would be positive. Second, we estimate a set of models that drop non-Partnership schools in Partnership districts from the comparison group. To the extent that Partnership affects outcomes in these schools, excluding them from the comparison creates a cleaner counterfactual. Finally, to assess the robustness of our inferences, we re-estimate our main models with standard errors clustered at the district level.

Survey Data and Analysis

To understand the elements of Partnership that made it more or less effective, we draw on surveys of teachers in Partnership districts and interviews with Partnership district leaders. Analyses of survey data provide a window into how Partnership impacted classrooms as teachers worked to implement their district's Partnership Agreement. We surveyed all teachers in Partnership districts (in Partnership and non-Partnership schools) in fall 2018 and fall 2019. The survey garnered response rates of 38% (N=2,718) in 2018 and 49% (N=3,324) in 2019. Panel A of online appendix Table A-3 shows response rates by cohort, year, and treatment status. As expected, response rates were higher for educators in Partnership schools. Panel B compares the

characteristics of teachers who did and did not participate in this survey, by Partnership status. In Partnership schools, Black and male teachers responded at a slightly lower rate and White teachers responded at a slightly higher rate. In non-Partnership schools, male teachers responded at a lower rate and elementary-certified teachers responded at a slightly higher rate. These differences are descriptively small and fall below the threshold for baseline equivalence of .25 standard deviation units recommended by the What Works Clearinghouse (What Works Clearinghouse, 2022). However, we repeat our analyses using inverse probability weights generated using teacher demographics, certification, and school fixed effects.

Survey items focused on a range of constructs related to Partnership implementation and teachers' perceptions of conditions within their school. In this study, we focus on constructs related to the mechanisms of improvement outlined in the Partnership Model's theory of change. To generate constructs to measure implementation, we conducted factor analyses within sets of items from the larger instrument that had been asked in both years of survey administration. We conducted separate factor analyses on all items within a single construct by first determining the number of factors using parallel analysis (Horn, 1965) and then creating principal components factors. Where the parallel analysis pointed to multiple factors within a set of items, we applied a varimax rotation to identify orthogonal factors, removing items that did not fit statistically with other items in the factor. This yielded three factors: 1) *Staffing*, teachers' perceptions of teacher and staff retention and teacher attendance; 2) *School accountability*, teachers' perceptions that consequences will follow if their school fails to meet its improvement goals; and 3) *Perception of school leadership*, the extent to which teachers found their school leader to be effective at communicating school mission, strategies, and goals, working with staff to meet curriculum standards, engaging family and community members, and establishing discipline policies. Scale

reliability coefficients range from 0.822 (school accountability) to .957 (school leadership). Online appendix Table A-4 provides descriptive statistics for the items used to create these factors, along with the text of items and their response scales. Factor loadings are presented in online appendix Table A-5. We conduct *t*-tests on these factor scores to compare the perceptions of teachers in Partnership schools with teachers in non-Partnership schools within Partnership districts. We focus on the differences between these sets of schools to help us understand what Partnership districts may be doing differently in their Partnership schools relative to their other schools that are not specifically identified for turnaround.

Qualitative Data

A key characteristic of the Partnership Model is that it provided districts with significant flexibility around the improvement goals set for Partnership schools and the strategies for achieving those goals. Additionally, the Partnership Model provided a range of supports from which districts could draw to improve their identified schools. This flexibility allowed variation in how Partnership was implemented in practice across districts that could not be captured with administrative data or surveys. To understand how districts approached the development and implementation of their Partnership Agreements, we recruited leaders in all Partnership districts/charters who were primarily responsible for the creation or implementation of their Partnership Agreement. We interviewed Partnership district leaders in 2018-2019 (n=22/36 Partnership districts, or 61%), and in 2019-2020 (n=22/29 Partnership districts, or 76%). In 2018-19 we conducted these interviews when Cohort 1 schools were generally near their 18-month interim evaluation and in 2019-20 when Cohort 2 schools were at the midpoint of implementing their Partnership Agreements.

In these 60-minute semi-structured interviews, we sought to understand how leaders perceived the Partnership Agreement and its design and implementation as well as overall perceptions of the successes, challenges, and intent of the Partnership Model. We began with open-ended questions asking them to describe what they put in their Partnership Agreement, what they did to implement the Agreement, and why. We also asked about the successes and challenges they experienced with design and implementation. Depending on the depth of responses, we asked probing questions aligned to the literature on turnaround and policy implementation such as whether or not turnover, relationships between stakeholders, and district capacity influenced implementation. Starting with open-ended questions allowed themes to first emerge primarily from the participants' *experience* rather than from the interviewers signaling the importance of specific factors. Following up with non-leading, probing questions allowed us to check to make sure the participant did not accidentally omit or overlook a theme that might have been important. To help maintain confidentiality, we assigned hockey team name pseudonyms to interview districts, and we describe interviewee roles with general categories, such as "superintendent" or "leader."

Team members coded the interviews using a deductive coding scheme in Dedoose software to condense data in alignment with the theory of change and key themes related to implementation, including codes such as "role of [district leader]," "district capacity," "human capital initiatives," and "role of MDE." The coding scheme was developed by three members of the qualitative team, with Year 2's coding scheme building from the original Year 1 coding scheme. For Year 1, coders independently coded three interviews to check for reliability and in Year 2, coders re-established reliability by coding one shared interview and finding agreement. Coders also met regularly to share findings and ensure reliability in coding. In Year 1, after an

initial round of coding, we organized key quotes and observations in a matrix to compare findings within and across interviews. Matrices facilitate comparisons and allow trends to emerge within and across cases as researchers group findings together (Bush-Mecenas & Marsh, 2018). Columns of the matrix included response to the reform, perceived benefits, perceived challenges, and conditions impacting coherence, and each row contained one interview's data. The matrices allowed us to categorize and better understand the variety of responses within and across Partnership districts. In Year 2, we used a similar process to identify key themes related to implementation as districts continued to refine their practices and plans to meet Partnership goals. We organized quotes by theme across interviews to better understand the scope and variation within the data. Finally, when possible, we counted mentions of key themes that emerged from district leader responses to the structured elements of our interviews. When appropriate, we present these counts to show similarities and differences across Partnership districts. Examples include mentions of specific ways districts utilized partners, positive impressions of working with the state education agency and intermediate school districts around teacher recruitment and retention, and/or using funding associated with the reform.

In the following section, we first discuss the results of our event study analyses that estimate changes in student outcomes attributable to the Partnership Model. We then turn to our qualitative and survey data to examine how Partnership implementation shaped these outcomes. By triangulating across these findings, we are able to explore how the ways in which Partnership schools and districts approached their reforms have shaped the reform's effectiveness.

Results

The Partnership Model's Estimated Effects on Student Outcomes

We show that the Partnership Model improved student outcomes in Partnership schools overall, though these effects varied by cohort, subject, and year. We also find evidence that Partnership had positive effects on other low-performing schools in Partnership districts, as suggested by the Partnership Model's theory of change. In both cohorts, these effects are larger when excluding students in non-Partnership schools in Partnership districts from the comparison group, pointing to possible positive effects of the district-level intervention components. We also find smaller but positive effects for students in near-selected schools in Partnership districts, providing additional evidence for district-level effects across lower performing schools in Partnership districts. In addition, Partnership effects are larger for lower achieving students.

Full Sample

Cohort 1. Table 2 provides the effect estimates from our main event study models. For Cohort 1, we find evidence that Partnership led to increased student achievement in both math and ELA in the first year of implementation, with the ELA effects continuing into the second year. We provide results from the Cohort 1 event study models in Panel A of Table 2 and the top row of Figure 2. Columns 1-8 of Table 2 show results for each outcome and the rows provide coefficients for the *Partnership x year* interaction term in our event study model, which estimates the difference between our treatment and comparison groups, relative to the year of Partnership identification, in each year. We interpret the coefficients for the years after Partnership identification as the causal effect of Partnership on the relevant outcome and use the coefficients for the years leading up to Partnership identification to test the assumption of

parallel trends between these groups prior to their selection for Partnership. For Cohort 1, we observe outcomes for two years of treatment: the 2017-18 and 2018-19 school years.

Columns 1-2 and 3-4 provide the levels and gains models for grades 4-8 math and ELA, respectively, which should bound the true effect of Partnership in the years after identification (Imberman, 2010). Here, significant estimates in either the gains or levels specification is suggestive of a significant effect, while significant estimates for both specifications provide stronger evidence.[11] Coefficients for the years leading up to Partnership identification, which are also illustrated in Figure 2, are generally small and none are statistically significant, which supports the assumption of conditional parallel trends. Columns 1-2 suggest math improvements in grades 4-8 of .047 to .121 standard deviations in 2017-18, though the estimate on the lower bound is not statistically significant. In 2018-19, we see descriptively positive but insignificant estimates of .071 to .075 standard deviations. Columns 3-4 show that fourth- through eighth-grade students in Partnership schools experienced significant and sustained achievement gains in ELA in the first two years of Partnership. In the first year of implementation (2017-18), the ELA achievement of students in Partnership schools improved by between .066 and .112 standard deviations. Improvements continued into 2018-19, when students assigned to Partnership performed by between .075 and .103 standard deviations higher relative to comparison students. We do not find evidence in our school-level models of significant improvements in SAT scores (columns 5-6), graduation (column 7), or dropout (column 8).

Cohort 2. We find little evidence in our main models, shown in Panel B of Table 2 and the second row of Figure 2, that Partnership impacted the Cohort 2 student and school outcomes under study. Only one year of post-Partnership outcomes are available for Cohort 2 as this group of schools was identified for treatment in the 2017-18 school year and began implementing their

respective Partnership Agreement in the 2018-19 school year. Here, too, estimates for the years prior to Partnership identification are, overall, small and none are significant.

Columns 1-4 show that the coefficient estimates for grades 4-8 math and reading are positive, but none are significant at conventional levels. Estimates for school-level SAT scores, graduation rate, and dropout rate, shown in Columns 5-8, are highly imprecise. Together, these results suggest that the Partnership Model was more effective at improving grades 4-8 student achievement in Cohort 1 than Cohort 2, and had no effect on SAT achievement, on-time high school graduation, and dropout. Supplemental analyses provide some evidence of small positive effects in grades 4-8 math and ELA in Round 2 but not Round 3 schools (online appendix Table A-6).[12]

Robustness tests. We test the robustness of these findings in several ways. For both cohorts, our estimates are qualitatively similar, though less precise, when we move the reference year back to the year prior to identification (see online appendix Table A-7). Results are also similar using the "all starters" approach—though somewhat larger and generally more precise. These results (online appendix Table A-8), provide some evidence that Partnership did not induce transfers out by the highest performing students. We find similar effects using baseline rather than time-varying covariates (online appendix Table A-9) and can make similar inferences when we cluster standard errors at the district rather than school level (online appendix Table A-10).

The Effect of Partnership on Other Low-Performing Schools in Partnership Districts

Using the full sample of Partnership and comparison schools may lead to downward-biased estimates if the intervention leads to improvements in Partnership districts' operations that enable them to more effectively support *all* of their schools and not just those identified for

Partnership. We therefore present a set of estimates in which the treatment group includes all students in Partnership schools and the comparison group includes only students in schools within districts never identified for Partnership. Using this sample, estimates for grades 4-8 math and ELA are both larger in magnitude and more precise for both cohorts of Partnership schools.

The left panel of Table 3 presents these results for Cohort 1 Partnership schools. We find that student math achievement increased by .087 to .174 standard deviations after the first year of Partnership implementation and by .123 to .124 standard deviations after the second, with each estimate significant at the $p<.05$ level. Estimates for ELA range from .092 to .137 standard deviations after the first year of Partnership and .10 to .133 standard deviations after the second, all significant at the $p < .05$ level. These estimates are meaningfully larger in magnitude, especially in math, and more significant when compared to our estimates from the full sample.

Results for Cohort 2 Partnership schools are shown in the right panel of Table 3. Here, too, estimates for M-STEP ELA and math are larger than the corresponding estimates when using the full sample. In math, student achievement improved by .051 to .088 standard deviations, both significant at conventional levels. ELA estimates are likewise larger in magnitude using this sample, at .021 to .059 standard deviations, but imprecise. For the other outcomes of SAT math and ELA achievement, graduation rate, and dropout (not shown here), estimates are similar to the main specification and not significant. These results show that the Partnership model has stronger effects when students in identified Partnership schools are compared with students in schools that were neither directly nor indirectly impacted by Partnership. Consistent with the Partnership Model's theory of change, Partnership may have district-level effects that benefit students in other low-performing schools, thus attenuating estimates in our main models.

Next, we answer this question more directly by estimating the effects of being in a low-performing school in a Partnership district in addition to the effects of being a Partnership school. In both cohorts, we find effects in grade 4-8 math in near-selected Partnership district schools that are statistically significant and similar in magnitude to the effects on Partnership schools (online appendix Table A-11). While we find positive and significant effects in ELA for Cohort 1 Partnership schools, we do not detect ELA effects for near-selected Partnership district schools in either cohort. Thus, consistent with the Partnership Model's theory of change, enhanced district-level systems may have driven improvements not just in Partnership schools but also in other low-performing schools in Partnership districts. It is also possible that Partnership districts allocated resources and/or directed efforts to Partnership schools as well as other lower-performing schools. We explore these possibilities in the section below where we discuss how Partnership district leaders approached the strategic planning element of the Partnership Model.

Partnership Effects by Prior Achievement

Table 4 provides estimates for grades 4-8 math and ELA achievement by prior achievement, with Cohort 1 in Panel A and Cohort 2 in Panel B. Columns 1-2 of Panel A show that Cohort 1 students in the bottom quartile of achievement improved math scores by .092 to .183 standard deviations in the first year, with both estimates being significant, and .096 to .126 standard deviations in the second year, though only the upper bound is significant. Estimates for students in the top three quartiles (columns 3-4) are descriptively positive but insignificant. Estimates for bottom quartile students in ELA are .102 to .195 standard deviations in Year 1 and .079 to .093 standard deviations in the second year (columns 5 and 6). However, the *Partnership x 2015-16* term for this sample (also in columns 5 and 6) is also positive and significant,

indicating that the positive estimates for lower achieving students in ELA may in part stem from mean regression following an identification year dip for these students—though the estimates are descriptively larger in the first year of implementation and sustain into the second year. In ELA, we also find some evidence that comparatively higher-achieving students experienced achievement gains following Partnership identification—though these estimates are less consistent. As shown in columns 7 and 8, estimates range from .051 to .087 standard deviations in year 1 and .07 to .102 standard deviations in year 2, with the upper bounds of those estimates significant at the $p<.10$ and $p<.05$ levels, respectively.

Panel B provides these estimates for Cohort 2. The significant positive estimates on the *Partnership x 2016-17* indicator (i.e., the year prior to identification) point to an identification year dip for the lowest achievers in math. We thus interpret these findings with caution. Following that dip, we find that estimates for lower-achieving students in math, shown in columns 1 and 2, rebounded to pre-identification levels (column 1) and were descriptively higher than pre-identification gains (column 2). Columns 5 and 6 of Table 4 show that ELA achievement for lower-achieving students increased by .054 to .073 standard deviations. Collectively, these results suggest that Partnership was particularly beneficial for the lowest-achieving students in Partnership schools, with gains in math and ELA evident in both cohorts, and that the overall effects of Partnership are driven by improvement in this group.

How Did Partnership Implementation Contribute to Outcomes?

In this section, we draw on our interview and survey data to explore how Partnership districts' implementation of the Partnership Model contributed to the policy's positive effects on ELA and math achievement in grades 4-8. The Partnership Model's theory of change posits several mechanisms through which the policy facilitates improvement in low-performing

schools. Our study of Partnership implementation identifies four such mechanisms that shaped these positive outcomes: strategic planning, external supports, the threat of accountability, and improving human capital.

Strategic Planning

Consistent with the theory of change held by the Partnership Model and similar ESSA-based turnaround reforms, we find that the strategic planning component of Partnership helped leaders align their ongoing and new improvement initiatives. For the most part, Partnership districts used the required strategic planning as an opportunity either to connect their previous reform efforts into their Partnership-driven reforms or to adopt new interventions they deemed necessary for improvement. For those who used the Agreement as a way to focus on intervention strategies already in place, many drew on already-existing strategic plans. However, this does not mean that these districts' approach toward their Partnership Agreement was purely performative. A number of leaders framed the development of their Agreement as a transition in their ongoing work around school improvement. For instance, the superintendent of Avalanche recalled that their Partnership Agreement tied together three strands of ongoing work in the district:

> That was really a culmination of – or a combination of what the ISD had already put in place, what we were already developing in terms of our own capacity, and then what we've realized we needed additional capacity and based on the MI Excel Blueprint [an improvement model focused on building school- and district-level systems]. I would say as far as the writing of that [the Partnership Agreement], it really came from those three places.

Speaking to the coherence that came about from aligning ongoing work with newly realized needs, they later said that it "really has given us more of a language for work we needed to get

done here." This example shows how Partnership district leaders used the flexible nature of the Partnership Model to design an improvement plan around their vision of reform.

In executing their strategic plan, some leaders focused primarily on improving their Partnership schools while others sought to make changes that were broader in scope, but that still centered their Partnership schools. Some district leaders shared how they differentiated efforts between Partnership and non-Partnership schools. For example, the leader of Canadiens said

> I mean, the work with the Partnership schools, I would say—on the surface, one would say, "Well, a lot of the schools face the same issues as the Partnership schools." I would generally agree with that. […] There are similarities in challenges, but there—in this past year, I've even seen that the Partnership schools have more pronounced issues. Some of the issues that are common across schools are more complex and deeper problems in our Partnership schools...

As a result, the superintendent of Canadiens focused their efforts more intently on their Partnership schools. Describing a broader approach to their reform, Bruins' superintendent indicated that several of their Partnership Agreement process goals included district-wide initiatives, but that they focused first on implementing these new programs in Partnership schools "more intensely" than other schools. For example, their new writing curriculum was phased in across grade levels over time across the district, but in their Partnership schools "we did it all- all grade levels right away." These examples show how districts approached Partnership reforms with a concentration on Partnership schools, but without necessarily excluding non-Partnership schools in their efforts, which may contribute to the positive effects we observe in low-performing, non-Partnership schools (see online appendix Table A-11).

External Supports: Technical Assistance and Funding

Providing supports to build capacity for improvement is a central theme of the Partnership Model's theory of change. External supports provided through the Partnership Model in the form of technical assistance and increased financial resources helped Partnership district leaders develop and implement their Partnership Agreement. Technical assistance was most often provided by the Office of Partnership Districts (OPD) at the Michigan Department of Education (MDE) to assist districts with the planning and implementation of their Partnership Agreement. The Partnership district's intermediate school district, or ISD, also provided technical assistance, with a focus on supporting reforms and initiatives at the school level. Financial support for Partnership districts came in the form of 21H, a grant administered by OPD to provide funding for Partnership districts' turnaround efforts. Our qualitative data suggest that both the technical assistance and monetary supports were important for Partnership districts in the design and implementation of their turnaround efforts.

Technical support from OPD is primarily provided by Partnership Agreement Liaisons (PALs), who were assigned to work with specific Partnership districts to serve as a link between the districts and MDE, smoothing the way for districts to gain access to needed information and supports. We find that Partnership leaders believed OPD and their PALs to be a helpful thought partner in their reform. For the most part, Partnership district leaders gave positive reports of the liaisons. For instance, the superintendent of Oilers described the support they received from their liaison as helping them to refine their plans around instruction:

> …our first liaison really helped us dig into, again, what M-STEP is because that's what our index score is based on […] We're able to look at those standards and

> then see with our own curriculum map, "Are we spending a lot of time on things that are measured on that assessment or not?"

In other cases, early on in Partnership implementation, the liaison was not particularly helpful. One district leader told us in the first year of interviews that the liaison "…wasn't proactive. It was more a reactive thing." Over time OPD made efforts to standardize the role of liaisons in their work with Partnership districts, positioning them as brokers between the MDE/OPD and local stakeholders. Partnership district leaders responded positively to this shift, with 21 of the 22 we interviewed in the second year of the study perceiving their liaison as helpful in this new role.

ISDs supported Partnership Agreement implementation by providing services such as professional development, curriculum support, and coaching, among others, to educators in Partnership schools. In interviews, approximately two-thirds of Partnership superintendents spoke favorably of the supports they received from their ISD. For example, the superintendent of Flames noted how their ISD provided direct support for their teachers and leaders:

> …they've been very supportive in providing professional development, even funding and some supplies and everything to our teachers and teaching staff and even to the administrators at the school to assist them throughout the process. They've had PDs [professional developments] on learning and for the principal and vice principals. It's really getting to those core issues that were lacking before. It's really helping them.

In another instance, the leader of Black Hawks described the support they received around curriculum from their ISD: "From [ISD], the partners who have been allocated for our building have done a phenomenal job at aligning the curriculum, [so that it] also helps us to meet the social and emotional needs of our children."

Additional financial support for Partnership schools comes from the 21H grant, which is appropriated by the state legislature and administered by OPD to support school turnaround. Monies from 21H can be used for a range of purposes to help Partnership districts implement the initiatives they feel are best aligned with the needs of their Partnership schools. Many Partnership superintendents spoke positively of the 21H funding; in the second year of the study, 19 of 22 superintendents interviewed reflected this view. Moreover, Partnership leaders noted that it was not just the money itself – which for many larger districts was only "a drop in the bucket" – but the flexibility given to them in how they could spend these relatively unrestricted dollars. On this theme, the superintendent of Flyers said, "We've been able to execute some of the strategies because of funding that we've received through 21H." They also reported:

> …it has definitely been a resource that has allowed us to move forward with some strategies that again, we've had our sights set on, but didn't have necessarily a way to make it come to fruition because there just wasn't funding available behind it in the school's budget.

Across both years of data collection, superintendents mentioned using 21H funding to provide trainings for their staff, purchase new curricula, and to create staff positions to support their turnaround work. For instance, through this funding the superintendent of Hurricanes "created three positions that were a direct correlation to the issues that we identified in our plan… a reading interventionist, a math interventionist, and an attendance liaison" and the superintendent of Black Hawks elected to provide trainings as a way of building capacity in the district, saying "Our hope is that those practices that the training allowed will continue." These examples show that the latitude given to districts in how this grant funding is spent has allowed them to be strategic by investing in programs and practices aligned with their Partnership Agreement.

Though superintendents overwhelmingly perceived grant funding from MDE to be helpful, a common refrain among Partnership superintendents in both years was that this additional funding was not sufficient to fully implement their vision for turnaround or to meet the needs of their district. The superintendent of Islanders explained that their grant was "…nothing when you're talking about doing programming. That's a year's programming. You're talking about substantial growth for kids who have five to six years of deficit. It doesn't work that way." Superintendents often framed these unmet needs in terms of the drastic reforms needed to turn around their Partnership schools, such as the superintendent of Stars, who said that their funding was merely "a drop in the bucket." Put simply, Partnership district leaders appreciated the opportunities afforded them by grant support but noted that more was needed if they were to bring about dramatic reform and improvement.

The Threat of Accountability

Although the focus of the Partnership Model is on providing supports to improve student outcomes in low-performing schools and districts, it creates pressure for schools to improve by requiring Partnership Agreements to include next-level accountability (NLA) measures to be implemented if a school's 36-month improvement goals are not met. We find that the threat of accountability is salient for educators in Partnership schools and that district leaders have used the threat of NLA to build momentum for their reform agenda.

Table 5 compares teachers' perceptions of school accountability if their school's improvement goals are not met between teachers who work in Partnership and non-Partnership schools. Comparisons that use the inverse probability weights are shown in online appendix Table A-12. Teachers working in Partnership schools have a much stronger sense of accountability than teachers who work in non-Partnership schools and this difference is

significant at the $p < .001$ level. Some items that make up this school accountability composite score illuminate teachers' perceptions around accountability for failing to improve. For instance, teachers in Partnership schools, compared to teachers in non-Partnership schools, see it as much more likely that their school leader or other staff will be removed if their school does not meet its improvement goals (see online appendix Figure A1). This suggests that teachers in Partnership schools see the threat of NLA as real, which may motivate them to more fully support their school's improvement efforts.

Partnership teachers' heightened perception of accountability likely comes at least in part from the messaging they received around Partnership from their school and district leadership. In interviews, multiple leaders reported using their Partnership status to build a sense of urgency for their reform agenda and used this to enact changes that otherwise would have been more challenging. Describing a debate within the district over when to implement a new reading program, the superintendent of Blues used their Partnership status to push for more immediate action, noting "How do we help the teachers feel this urgency? That's where I think the power of the Partnership can come in if you capitalize on it, like, look [we've only got so much time]." Some leaders identified specific actions they were able to take using the urgency they built around their Partnership status, such as the superintendent of Penguins, who at one of their Partnership schools changed the grade configuration, replaced the principal, and also replaced a large share of the teachers. They noted that such drastic changes would normally be quite difficult but that Partnership created a space to implement this change. Another Partnership district leader explained this strategy as "I would just say a leader knows how to use extra accountability as a lever for change and creating a sense of urgency." They further elaborated:

> Having come in knowing that these were the lowest performing schools, I've been able to, more over the last couple of years, use the Partnership Agreement as a lever to get a greater sense of urgency among the principal and teachers to say "We have to do things differently here. Why? Because we're a Partnership school."

These accounts illustrate how some Partnership district leaders used Partnership, and in particular the threat of high-stakes accountability, as a warrant to make changes that aligned with their vision for improvement in identified schools.

Improving Human Capital

Consistent with earlier research on school turnaround (Heissel & Ladd, 2018; Harbatkin, in press; Henry et al., 2020; Strunk et al., 2016; Zimmer et al., 2017), the Partnership Model's theory of change recognizes human capital as a key driver of improvement in turnaround schools. There are two important elements of human capital to consider – effective school leaders (Aladjem et al., 2010; Hitt et al., 2018) and high-quality teachers (Henry & Harbatkin, 2020; Henry et al., 2020). We find that Partnership districts had mixed success in improving human capital in their Partnership schools across these two categories. Overall, it appears that Partnership districts were able to bring in more effective principals to lead Partnership schools. However, while some leaders reported improvements in their ability to attract and retain teachers in Partnership schools compared to prior years, they still cited maintaining a supply of high-quality teachers as among their most significant challenges.

In our interviews, Partnership district leaders noted some areas of concern around their leaders but identified principals as key players in their reforms. Speaking about some of the

specific challenges in their Partnership schools, the superintendent of Canadiens lamented that many of their leaders were not well-equipped to drive instructional change, saying:

> I think a lot of our Partnership principals are not instructional leaders, defined as individuals that know how to recruit, retain teachers, know how to give constant feedback to teachers, can problem-solve with teachers, and create a culture of instructional focus where the principal is truly the instructional leader. Can go in as a generalist in specific content areas, provide feedback on how to improve instruction, can lead professional development, can analyze data, and really have credibility with teachers to improve practice.

Acknowledging issues with leadership, several Partnership district leaders replaced principals to create an environment more conducive to growth and improvement. The superintendent of Penguins, for example, strategically selected a principal from another school to lead a Partnership school and anchored their reforms there around this new leadership, saying:

> That school has been a pain in the whatever of anybody who's been sitting in this chair, because it just has been a mess. I mean a total mess. No buy in or whatever. [...] We turn it over to this new principal – not new but [principal name], and we say, "Go for it." [...] and it's like the sun has risen on that place. It's a new environment, it's a new place, different staff, kids are happy, not a peep out them, community is happy.

Teachers in Partnership schools seemed to understand district leaders' efforts to improve the quality of leadership in their schools. Table 5 shows that, on average, teachers in Partnership schools gave their principals significantly higher ratings than teachers in non-Partnership schools, though this difference is not significant when using weighted responses (see online

appendix Table A-12). Within this construct, Partnership teachers gave their principals significantly higher ratings for facilitating and encouraging professional development (see online appendix Figure A1). This suggests that the types of efforts by district leaders noted above improved leadership quality in their Partnership schools, which in turn served as a catalyst for positive change.

Some district leaders cited improvements in their ability to recruit and retain teachers relative to prior years, which could help explain some of the academic growth in Partnership schools. For example, the leader of Senators noted they had “increased the number of certified teachers [in Partnership schools].” The Avalanche leader said they were able to use funds from Partnership to improve their hiring process, which increased the fit between candidates and Partnership schools. The Ducks superintendent said, “we’ve seen some improvement in recruitment and retention.” Others were able to use modest financial incentives to help stabilize staffing, and the leader of Flyers explained that they were able to change "job descriptions to attract a different level of talent.” Despite these perceived improvements, these and other leaders still felt greater staffing challenges in their Partnership schools compared to other schools.

These improvements in staffing are best characterized as making the problem somewhat less acute; retaining and developing high-quality teachers in Partnership schools remained a significant challenge for most. Partnership schools experience much greater teacher turnover than other schools and struggled to replace them with the high-quality teachers they believe are necessary for their reforms, which hampered their efforts to improve instruction. In fact, 23% to 29% of teachers exit Partnership schools each year, roughly twice the statewide average, a trend that predated Partnership and persisted into the years of Partnership implementation (authors’ calculation, 2021). It is perhaps unsurprising then that, as Table 5 shows, teachers in Partnership

schools give their school lower ratings for staffing (robust to using survey weights, see online appendix Table A-12). Within this construct, teachers in Partnership schools gave their schools particularly low marks for teacher retention (online appendix Figure A1). Moreover, leaders reported significant challenges in attracting the high-quality educators they need, and sometimes challenges even in attracting educators with basic credentials. The superintendent of Devils noted:

> For every opening, you used to get 40 applicants, ten of whom were very hirable, five of whom were probably great. I'm talking up 'till five about five years ago. Now, you are lucky to get one (applicant), extremely lucky to get two or three, and incredibly lucky if one of them is really a hirable candidate.

Of the 22 Partnership district leaders interviewed in the second year of data collection, 20 raised the supply of teachers as a concern in their turnaround work, which complicated Partnership districts' efforts to reform the instructional core of their Partnership schools.

Issues around teacher recruitment and retention made it more challenging for Partnership schools to implement their Partnership Agreement by making it difficult to build momentum around instructional reform. Recalling that turnover forced their district to keep "starting over," the superintendent of Flyers described:

> We've been unable to have one program model implemented more than one year due to turnover in either leadership and/or teachers. High-quality certified teachers is another very large obstacle that we continue to face. In a building with 19 total teachers, four of them are certified teachers, so we have a serious shortage in the availability of certified teachers. […] There's the consistency issue, which

> prevents the model from being able to really take hold, and then there's the lack of talent.

In addition to the challenge of sustaining reform in the face of turnover, this quote refers to a reliance on teachers who are less qualified, and in some cases not certified. Though it was not typical for a majority of teachers in Partnership schools to be non-certified, in the first year of our Partnership superintendent interviews, one-third described using long-term substitute teachers to fill teaching positions in their Partnership school(s). These examples highlight the challenges of successfully implementing a turnaround plan in the face of instability amongst the teaching staff and suggest that Partnership may have been more effective if Partnership schools had greater access to certified and effective teachers.

Discussion and Implications

This study is the first to examine a statewide school turnaround policy, Michigan's Partnership Model, under the Every Student Succeeds Act. Overall, we find that students in both cohorts of Partnership schools experienced achievement gains in grades 4-8 math and that students in the first cohort of Partnership schools also saw gains in ELA. We also find evidence of district-level effects in other low-performing Partnership district schools, consistent with the Partnership Model's theory of change that aims to improve outcomes in Partnership districts more broadly. Moreover, the Model's positive achievement effects are driven by gains amongst the lowest-achieving students in Partnership schools. In this way, Partnership compares favorably with recent turnaround efforts, which find positive, but smaller effects (Redding & Nguyen, 2020; Schueler et al., 2020), and with earlier turnaround efforts in Michigan under NCLB waivers, which found null or negative effects (Hemelt & Jacob, 2017, 2018). In addition, Partnership produced much more positive effects than the generally negative impacts of

Michigan's earlier school closures (Brummet, 2014) - the initially proposed treatment for Michigan's lowest performing schools. These findings may somewhat alleviate early concerns that the reduced federal footprint and flexibility within ESSA would allow states to be less ambitious in turning around their low-performing schools (for a summary of concerns, see Egalite et al., 2017). Unfortunately, due to the COVID-19 pandemic, we are not able to observe outcomes for students in Partnership schools several years into implementation, when a growing body of recent evidence that suggests that effects are most evident (e.g., Peurach & Neumerski, 2015; Redding & Nguyen, 2020; Sun et al., 2017).

At the same time, we do not find evidence that Partnership improved outcomes for students in Partnership high schools in terms of SAT achievement, graduation, or dropout. While this may be because the Partnership Model had no effect on high schools, other factors may contribute to these null results. One is that the comparatively small number of high schools in our sample means that our school-level analyses lack sufficient statistical power to detect an effect. Second, since the majority of Partnership schools served students in the elementary and middle school grades, Partnership districts may have placed most of their focus on those schools, leaving fewer resources to support improvement in their high schools. Lastly, because SAT achievement and graduation/dropout rely on experiences students have throughout high school, it may take more than one or two years to observe improvement in these outcomes.

Given that evaluations of recent turnaround efforts find variation both across and within contexts, it is vital to understand the elements of turnaround policies that are most conducive to improvement in student outcomes so that successes can be replicated. This is especially so given that ESSA provides states and districts significant leeway in how they turn around their low-performing schools, which may allow them to adapt to emerging evidence on effective

turnaround. Drawing on our qualitative and survey data, this study contributes to the literature on school turnaround by identifying factors that helped Partnership schools and districts move the needle on student outcomes. First, the supports provided through the Partnership Model helped districts tailor their reforms and strategies to the needs of their identified Partnership schools. Partnership district leaders gave positive reports of the assistance they received from their liaison at MDE/OPD, citing examples where their liaison provided support for data analysis and connected leaders with state-level resources. Partnership districts also linked support from their ISD, in areas such as professional development and curriculum consulting, with successful implementation of their Partnership Agreements. This adds to the existing evidence that technical assistance for low-performing schools supports achievement gains, especially for disadvantaged students (see, for example, Strunk & McEachin, 2014). Financial support also helped Partnership districts implement the strategies they felt were important for successful turnaround. However, district leaders simultaneously noted that far more money was needed to help Partnership schools and districts implement the kinds of reforms that might make the largest difference in outcomes – in particular improving teacher and leader recruitment and retention.

Second, Partnership district leaders leveraged the planning and accountability elements of the Partnership Model to bring about the change they felt was needed to move their Partnership schools forward. For instance, we found that, as in other studies of turnaround reforms (e.g. Meyers & Hitt, 2018; Strunk et al., 2016; Sun et al., 2017), the requirement to develop an improvement plan and engage in strategic planning was critical to reform success as leaders used the planning process to align existing initiatives as well as integrate new reforms into their work. Given the centrality of planning in the ESSA model, this bodes well for the potential success of ESSA-induced whole school reforms.

Third, staffing Partnership schools with skilled school leaders who were a good fit for their district's reform agenda was important for successful turnaround. Partnership district leaders were intentional about the principals they selected to lead their Partnership schools and teachers in Partnership schools reported higher-quality leadership than teachers in other schools within Partnership districts. However, the recruitment and retention of teachers presented a significant challenge for Partnership districts, which is troubling given the documented importance of teachers for successful turnaround (see Redding & Nguyen, 2020; Strunk et al., 2016; Zimmer et al., 2017). Partnership schools experience high rates of teacher exit and leaders report resorting to less qualified teachers to fill teaching positions in them. Though we find that some district leaders were able to make relative improvements in teacher recruitment and retention, Partnership schools and districts overall struggled to substantially move the needle on staffing schools with qualified and high-quality teachers, which may help to explain why achievement gains were not larger or more consistent across cohorts.

Last, some Partnership leaders drew on the threat of next-level accountability included in Partnership Agreements to bring about change by using it to build a sense of urgency for their reforms, which allowed them to take actions that may have been difficult in another context. This is interesting, especially as in the later years of the reform the Michigan Department of Education worked to shift the focus from accountability to support, which – as we note above – was found to be central to the success of the model. This suggests that retaining some version of next level accountability may still be important as a possibility should support and intervention be insufficient to help schools and districts improve in the longer-term.

Together, the results from this study lead to several implications for policymakers and practitioners, both in Michigan and more broadly across the country in light of ESSA's focus on

states taking the lead in school turnaround. First, ESSA requires that turnaround interventions be "evidence-based," and the findings of this study add to the small but growing body of research on not just the effectiveness of school turnaround, but the factors that *promote* successful turnaround. The positive results we find for Cohort 1 schools indicates that policymakers should continue to support turnaround efforts in Partnership schools and districts, letting the reform play out and improve rather than turning whole-sale to a new improvement strategy. Partnership district leaders found the technical assistance and grant funding through the Partnership Model to be helpful in their turnaround efforts. Additionally, leaders leveraged these resources in different ways that aligned with their vision of reform. This suggests that states should incorporate some degree of flexibility for schools and districts in how they approach reform.

Second, we argue that the stronger effects we find for Cohort 1, relative to Cohort 2, provides yet another data point that large-scale reforms require investments – both in terms of financial and in-kind resources. Partnership leaders reported that additional resources, such as grant funding to support Partnership districts' implementation of their reforms and support from ISDs, helped their Partnership schools make progress toward their improvement goals. At the same time, the level of these supports remained relatively flat as new schools were identified. Cohort 2 consists of more than twice as many Partnership schools (82) as Cohort 1 (37). Moreover, more than half of Cohort 2 Partnership schools are in districts that operate Cohort 1 Partnership schools, which could have strained their capacity as they worked to implement improvement efforts across a larger number of schools. These factors may have limited the impact of Partnership and explain why we find smaller effects for Cohort 1 in the second year of implementation (which coincided with the first implementation year for Cohort 2) and smaller effects for Cohort 2 than Cohort 1. Importantly, providing additional funding for turnaround

schools is one area where Michigan's Partnership Model stands out from the requirements of ESSA, which does not provide or require additional resources for turnaround schools. It is therefore critical that states fund and support improvement efforts, ideally targeting additional resources to the areas identified in schools' and districts' needs assessments and strategic plans.

Third, the Partnership example highlights the need for human capital reforms in addition to instructional and operational changes. Without a sufficient supply of high-quality educators, it will be difficult for even the best laid plans to come to fruition, and schools will face greater challenges achieving their long-term improvement goals. Here, too, states have a crucial role since they are well positioned to enact policies and programs that 1) bolster the supply of qualified and high-quality educators, and 2) create incentives for educators to work in turnaround schools and districts. Lastly, while accountability receives much less emphasis under ESSA than previous generations of turnaround, states may wish to retain some element of accountability in their turnaround policies since savvy leaders can use the threat of accountability to bring about more rapid change.

Endnotes

[1] The state identified 38 schools for possible closure. After this announcement, but before the Partnership Model was adopted, one charter school closed and so was not included.

[2] A fourth round of Partnership schools and districts was scheduled to be identified in the fall of 2020. Due to COVID-19 pandemic delays, the fourth round will be selected in fall 2022.

[3]A descriptive analysis shows that students who transfer are systematically different from those who stay. Students who transfer are lower achieving and more likely to be economically disadvantaged and Black, and less likely to be Hispanic or Latinx than stayers. Online appendix Table A-1 provides results of *t*-tests comparing the two groups for each analytic sample.

[4] These two comparison groups are not mutually exclusive. For example, a 2016 Priority School that was also in the bottom 10% of schools in 2016-17 but was never identified as Partnership would be part of both comparison groups. This approach is analogous to sampling with replacement and allows us to construct comparison groups for each treatment wave that are most similar to the treated schools at baseline on school performance. We also considered using the

bottom 7% of schools as our comparison group, but this led to a small comparison group because the bottom 5% were identified for Partnership. To obtain adequate statistical power, we selected the bottom 10% as our comparison group. In results not shown here, we ran our analyses using the bottom 7% as the comparison group, which yielded estimates that were similar in direction and magnitude but less precise.

[5] Some schools and students are included in the comparison groups for both cohorts of Partnership schools if they attended a school in the comparison group for Cohort 1 in 2016-2017 and then a school in the comparison group for Cohort 2 in 2017-2018. Because of this, adding the number of unique students or the number of student-years from the two cohort analyses appears greater than the number of unique students and student-years in the analytical sample. There is no overlap of students in the treated group as in cases where a student attended a Cohort 1 school in 2016-2017 and then a Cohort 2 school in 2017-2018, the student was assigned to the Cohort 1 treated group because that was when they were first treated under Partnership.

[6] We also estimate a treatment-on-the-treated (TOT) effect for Cohort 1 year 2 and find similar effects to the all starters estimate.

[7] 80% of Cohort 1 and 78% of Cohort 2 students were in the lowest achievement category (1 on a 1-4 scale where 3 and above is proficient) in math and 78% and 74%, respectively, were in this category in ELA.

[8] In Michigan, students are considered economically disadvantaged if they are eligible for free or reduced-price lunch, are in households that receive assistance through SNAP or TANF, are homeless, migrant, or in foster care.

[9] Using statewide rather than sample means and standard deviations to standardize the outcomes yields a more conservative estimate because the state-level means are larger (our sample of Partnership and comparison schools are low performing) and the distribution of z-scores among students in Partnership and comparison schools standardized on the statewide sample is tighter than the distribution of z-scores standardized only on the analytic sample would be.

[10] Due to the COVID-19 pandemic, Michigan (like all other states) did not administer standardized tests in the spring of 2020 and not all students took the tests in the spring of 2021. For this reason, we cannot estimate effects past the second year for Cohort 1 and the first year for Cohort 2.

[11] The negative adjusted R^2 on the gains models should not be construed as a signal about the model's ability to estimate the effects of Partnership on the gains outcomes. In this case, it simply means that the regression line, on average, provides a poor fit overall (i.e., a poorer fit than a horizontal line), which is unsurprising given the variance of the delta outcomes and the relationship between the delta in particular and the variables in the equation, on average. However, these models include student fixed effects, and the within-student R2 are all positive because the model does provide improved goodness of fit over a horizontal line within students. Finally, it is important to note that the extent to which the independent variables of interest explain the outcome is measured by the test statistic on the relevant regression coefficient and

not the model R^2. Here, the standard errors on the interactions between the Partnership and year indicators suggest that our estimates of interest are reasonably precisely estimated.

[12] The consistently positive estimates on the levels and gains models in Round 2 provide evidence of a positive effect, though we note that only the estimates on the levels models are statistically significant at conventional levels. For Round 3, the levels and gains models together show that the true effect is between about -0.02 and +0.06 on both outcomes. Because this range includes zero, we consider the estimates to represent null effects. While we cannot tell from our analyses why we see more positive effects in Round 2 than 3, there are several possible mechanisms that may be driving these differences. First, Round 3 schools were identified as the state's first cohort of CSI schools while Round 2 schools were identified based on state education leader decisions. It is possible that Round 3 schools were hindered by the initial stigma of the federal low-performing designation while Round 2 schools were not. It is also possible that because Round 2 schools were identified largely based on test scores while Round 3 schools were identified based on the state's multidimensional school quality index under ESSA, Round 2 schools were more laser-focused on raising test scores and therefore able to make more immediate impacts on test score outcomes. Finally, because Round 2 schools were identified in fall 2017, they had more time for improvement planning than Round 3 schools, which were identified in spring 2017. It may be the case that by the first implementation year, Round 2 schools were better prepared to implement their improvement plans than Round 3 schools.

References

Aladjem, D. K., Birman, B. F., Orland, M., Harr-Robins, J., Heredia, A., Parrish, T. B., & Ruffini, S. J. (2010). *Achieving Dramatic School Improvement: An Exploratory Study*. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. https://eric.ed.gov/?id=ED526783

Anfara, V. A., Patterson, F., Buehler, A., & Gearity, B. (2006). School Improvement Planning in East Tennessee Middle Schools: A Content Analysis and Perceptions Study. *NASSP Bulletin*, *90*(4), 277–300. https://doi.org/10.1177/0192636506294848

Atchison, D. (2020). The Impact of Priority School Designation Under ESEA Flexibility in New York State. *Journal of Research on Educational Effectiveness*, *13*(1), 121–146. https://doi.org/10.1080/19345747.2019.1679930

Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395. https://doi.org/10.1016/j.jfineco.2022.01.004

Bonilla, S., & Dee, T. (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

Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, *42*(2), 231–268. https://doi.org/10.3102/00028312042002231

Burkhauser, S. (2017). How Much Do School Principals Matter When It Comes to Teacher Working Conditions? *Educational Evaluation and Policy Analysis*, *39*(1), 126–145. https://doi.org/10.3102/0162373716668028

Bush-Mecenas, S., & Marsh, J. A. (2018). The DIVE Approach: Using Case-Ordered Meta-Matrices and Theory-Based Data Displays to Analyze Multiple Case Study Data. In C. R. Lochmiller (Ed.), *Complementary Research Methods for Educational Leadership and Policy Studies* (pp. 33–56). Springer International Publishing. https://doi.org/10.1007/978-3-319-93539-3_3

Carlson, D., & Lavertu, S. (2018). School Improvement Grants in Ohio: Effects on Student Achievement and School Administration. *Educational Evaluation and Policy Analysis*, *40*(3), 287–315. https://doi.org/10.3102/0162373718760218

Carnoy, M., & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331. https://doi.org/10.3102/01623737024004305

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9), 1045–1057. https://doi.org/10.1016/j.jpubeco.2009.06.002

Close, K., Amrein-Beardsley, A., & Collins, C. (2019). Mapping America's teacher evaluation plans under ESSA. *Phi Delta Kappan*, *101*(2), 22–26. https://doi.org/10.1177/0031721719879150

Dee, T. (2012). *School Turnarounds: Evidence from the 2009 Stimulus* (Working Paper No. 17990). National Bureau of Economic Research. https://doi.org/10.3386/w17990

Dee, T., & Dizon-Ross, E. (2019). School Performance, Accountability, and Waiver Reforms: Evidence From Louisiana. *Educational Evaluation and Policy Analysis*, 0162373719849944. https://doi.org/10.3102/0162373719849944

Dee, T., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446. https://doi.org/10.1002/pam.20586

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in Texas. *The Georgetown Public Policy Review*, *19*(1), 21–39.

Dougherty, S. M., & Weiner, J. M. (2017). The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance. *Educational Policy*, *33*(4), 555–586. https://doi.org/10.1177/0895904817719520

Dragoset, L., James-Burdumy, S., Hallgren, K., Perez-Johnson, I., Herrmann, M., Tuttle, C., Angus, M. H., Herman, R., Murray, M., Tanenbaum, C., & Graczewski, C. (2015). *Usage of Practices Promoted by School Improvement Grants* (NCEE 2015-4019). National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED559928

Duke, D. L. (2004). The Turnaround Principal: High-Stakes Leadership. *Principal*, *84*(1), 12–23.

Duke, D. L., & Salmonowicz, M. (2010). Key Decisions of a First-year 'Turnaround' Principal. *Educational Management Administration & Leadership*, *38*(1), 33–58. https://doi.org/10.1177/1741143209345450

Dunn, L., & Ambroso, E. (2019). Balancing Act: State and District Roles in School Improvement under ESSA. In *Center on School Turnaround at WestEd*. Center on School Turnaround at WestEd. https://eric.ed.gov/?id=ED602966

Fullan, M. (2016). The elusive nature of whole system improvement in education. *Journal of Educational Change*, *17*(4), 539–544. https://doi.org/10.1007/s10833-016-9289-1

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327. https://doi.org/10.1002/pam.20091

Harbatkin, E. (2022). Staffing for School Turnaround in Rural Settings. *Leadership and Policy in Schools*, 1–23. https://doi.org/10.1080/15700763.2022.2058963

Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, *62*, 302–320. https://doi.org/10.1016/j.econedurev.2017.08.001

Hemelt, S. W., & Jacob, B. (2017). *Differentiated Accountability and Education Production: Evidence from NCLB Waivers* (Working Paper No. 23461). National Bureau of Economic Research. https://doi.org/10.3386/w23461

Hemelt, S. W., & Jacob, B. A. (2018). How Does an Accountability Program that Targets Achievement Gaps Affect Student Performance? *Education Finance and Policy*, *15*(1), 45–74. https://doi.org/10.1162/edfp_a_00276

Henry, G. T., & Guthrie, J. E. (2019). The Effects of Race to the Top School Turnaround in North Carolina. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://edworkingpapers.com/ai19-107

Henry, G. T., & Harbatkin, E. (2020). The Next Generation of State Reforms to Improve their Lowest Performing Schools: An Evaluation of North Carolina's School Transformation Intervention. *Journal of Research on Educational Effectiveness*. https://doi.org/10.1080/19345747.2020.1814464

Henry, G. T., McNeill, S. M., & Harbatkin, E. (2022). Accountability-driven school reform: Are there unintended effects on younger children in untested grades? *Early Childhood Research Quarterly*, *61*, 190–208. https://doi.org/10.1016/j.ecresq.2022.07.005

Henry, G. T., Pham, L. D., Kho, A., & Zimmer, R. (2020). Peeking Into the Black Box of School Turnaround: A Formal Test of Mediators and Suppressors. *Educational Evaluation and Policy Analysis*, *42*(2), 232–256. https://doi.org/10.3102/0162373720908600

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *Turning Around Chronically Low-Performing Schools. IES Practice Guide* (NCEE 2008-4020). National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED501241

Hitt, D. H., Woodruff, D., Meyers, C. V., & Zhu, G. (2018). Principal Competencies that Make a Difference: Identifying a Model for Leaders of School Turnaround. *Journal of School Leadership*, *28*(1), 56–81. https://doi.org/10.1177/105268461802800103

Hopkins, D., Stringfield, S., Harris, A., Stoll, L., & Mackay, T. (2014). School and system improvement: A narrative state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 257–281. https://doi.org/10.1080/09243453.2014.885452

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Huber, D. J., & Conway, J. M. (2015). The Effect of School Improvement Planning on Student Achievement. *Planning & Changing*, *46*(1/2), 56–70.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(5), 761–796. https://doi.org/10.1016/j.jpubeco.2004.08.004

Jennings, J., & Bearak, J. M. (2014). "Teaching to the Test" in the NCLB Era: How Test Predictability Affects Our Understanding of Student Performance. *Educational Researcher*, *43*(8), 381–389. https://doi.org/10.3102/0013189X14554449

Karcher, H., & Knight, D. S. (2021). Funding Structures and State Capacity for School Improvement under the Every Student Succeeds Act: Case Studies of Five States. *Journal of Education Human Resources*, *39*(4), 397–429. https://doi.org/10.3138/jehr-2021-0005

Ladd, H. F. (2017). No Child Left Behind: A Deeply Flawed Federal Policy. *Journal of Policy Analysis and Management*, *36*(2), 461–469. https://doi.org/10.1002/pam.21978

Lee, J., & Reeves, T. (2012). Revisiting the Impact of NCLB High-Stakes School Accountability, Capacity, and Resources: State NAEP 1990–2009 Reading and Math Achievement Gaps and Trends. *Educational Evaluation and Policy Analysis*, *34*(2), 209–231. https://doi.org/10.3102/0162373711431604

Leithwood, K., & Jantzi, D. (1990). Transformational Leadership: How Principals Can Help Reform School Cultures. *School Effectiveness and School Improvement*, *1*(4), 249–280. https://doi.org/10.1080/0924345900010402

LiCalsi, C., Citkowicz, M., Friedman, L. B., & Brown, M. (2015). *Evaluation of Massachusetts Office of District and School Turnaround Assistance to Commissioner's Districts and Schools: Impact of School Redesign Grants*. American Institutes of Research. https://eric.ed.gov/?id=ED583091

LiCalsi, C., & Píriz, D. G. (2016). *Evaluation of level 4 school turnaround efforts in Massachusetts. Part 2: Impact of school redesign grants*. American Institutes of Research. http://www.doe.mass.edu/research/reports/2016/09L4TurnaroundImpact.docx

Malen, B., & Rice, J. K. (2016). School reconstitution as a turnaround strategy: An analysis of the evidence. In W. Matthis & T. Trujillo (Eds.), *Learning from the federal market-based*

reforms: Lessons for the Every Student Succeeds Act (ESSA) (pp. 99–125). Information Age Publishing.

Meyers, C. V. (2020). District-Led School Turnaround: A Case Study Of One U.S. District's Turnaround Launch For Multiple Schools. *Leadership and Policy in Schools*, *19*(4), 710–729. https://doi.org/10.1080/15700763.2019.1637902

Mintrop, H., & MacLellan, A. M. (2002). School Improvement Plans in Elementary and Middle Schools on Probation. *The Elementary School Journal*, *102*(4), 275–300.

Papay, J., & Hannon, M. (2018, November 8). *The Effects of School Turnaround Strategies in Massachusetts*. 2018 APPAM Fall Research Conference: *Evidence for Action: Encouraging Innovation and Improvement*, Washington, D.C. https://appam.confex.com/appam/2018/webprogram/Paper26237.html

Papay, J. P., Kraft, M. A., & James, J. K. (2021). Operator Versus Partner: A Case Study of Blueprint School Network's Model for School Turnaround. *Journal of Research on Educational Effectiveness*, *0*(0), 1–27. https://doi.org/10.1080/19345747.2021.1998737

Peurach, D. J., & Neumerski, C. M. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379–420. https://doi.org/10.1007/s10833-015-9259-z

Peurach, D., & Neumerski, C. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379–420. https://doi.org/10.1007/s10833-015-9259-z

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 Local Innovation Zones. *AERA Open*, *6*(2). https://doi.org/10.1177/2332858420922841

Player, D., & Katz, V. (2016). Assessing School Turnaround: Evidence from Ohio. *The Elementary School Journal*, *116*(4), 675–698. https://doi.org/10.1086/686467

Portz, J., & Beauchamp, N. (2020). Educational Accountability and State ESSA Plans. *Educational Policy*, 0895904820917364. https://doi.org/10.1177/0895904820917364

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207–241. https://doi.org/10.1257/pol.6.3.207

Redding, C., & Nguyen, T. D. (2020). The Relationship Between School Turnaround and Student Outcomes: A Meta-Analysis. *Educational Evaluation and Policy Analysis*, *41*(4), 0162373720949513. https://doi.org/10.3102/0162373720949513

Redding, C., & Searby, L. (2020). The Map Is Not the Territory: Considering the Role of School Improvement Plans in Turnaround Schools. *Journal of Cases in Educational Leadership*, *23*(3), 63–75. https://doi.org/10.1177/1555458920938854

Schueler, B. E., Asher, C. A., Larned, K. E., Mehrotra, S., & Pollard, C. (2021). Improving Low-Performing Schools: A Meta-Analysis of Impact Evaluation Studies. *American Educational Research Journal*, *Online first*. https://doi.org/10.3102/00028312211060855

Schueler, B. E., & Bleiberg, J. F. (2022). Evaluating Education Governance: Does State Takeover of School Districts Affect Student Achievement? *Journal of Policy Analysis and Management*, *41*(1), 162–192. https://doi.org/10.1002/pam.22338

Schueler, B. E., Goodman, J. S., & Deming, D. J. (2017). Can States Take Over and Turn Around School Districts? Evidence From Lawrence, Massachusetts. *Educational*

Evaluation and Policy Analysis, *39*(2), 311–332. https://doi.org/10.3102/0162373716685824

Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*(5), 556–563. https://doi.org/10.1016/j.econedurev.2007.06.004

Strunk, K. O., Marsh, J. A., Bush-Mecenas, S. C., & Duque, M. R. (2016). The Best Laid Plans: An Examination of School Plan Quality and Implementation in a School Improvement Initiative. *Educational Administration Quarterly*, *52*(2), 259–309. https://doi.org/10.1177/0013161X15616864

Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, S., & Weinstein, T. (2016). The Impact of Turnaround Reform on Student Outcomes: Evidence and Insights from the Los Angeles Unified School District. *Education Finance and Policy*, *11*(3), 251–282. https://doi.org/10.1162/EDFP_a_00188

Strunk, K. O., & McEachin, A. (2014). More Than Sanctions: Closing Achievement Gaps Through California's Use of Intensive Technical Assistance. *Educational Evaluation and Policy Analysis*, *36*(3), 281–306. https://doi.org/10.3102/0162373713510967

Strunk, K. O., McEachin, A., & Westover, T. N. (2014a). The Use and Efficacy of Capacity-Building Assistance for Low-Performing Districts: The Case of California's District Assistance and Intervention Teams. *Journal of Policy Analysis and Management*, *33*(3), 719–751. https://doi.org/10.1002/pam.21658

Strunk, K. O., McEachin, A., & Westover, T. N. (2014b). The Use and Efficacy of Capacity-Building Assistance for Low-Performing Districts: The Case of California's District Assistance and Intervention Teams. *Journal of Policy Analysis and Management*, *33*(3), 719–751. https://doi.org/10.1002/pam.21658

Sun, M., Kennedy, A., & Loeb, S. (2020). *The Longitudinal Effects of School Improvement Grants* (EdWorkingPaper No. 20-177). Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai20-177

Sun, M., Liu, J., Zhu, J., & LeClair, Z. (2019). Using a Text-as-Data Approach to Understand Reform Processes: A Deep Exploration of School Improvement Strategies. *Educational Evaluation and Policy Analysis*, *41*(4), 510–536. https://doi.org/10.3102/0162373719869318

Sun, M., Penner, E. K., & Loeb, S. (2017). Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants. *American Educational Research Journal*, *54*(4), 607–643. https://doi.org/10.3102/0002831217695790

Weiner, J. (2016). Possibilities or paradoxes? How aspiring turnaround principals conceptualise turnaround and their place within it. *School Leadership & Management*, *36*(5), 471–492. https://doi.org/10.1080/13632434.2016.1247037

Wooldridge, J. M. (2021). *Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators* (SSRN Scholarly Paper ID 3906345). Social Science Research Network. https://doi.org/10.2139/ssrn.3906345

Woulfin, S. L., & Weiner, J. (2019). Triggering Change: An Investigation of the Logics of Turnaround Leadership. *Education and Urban Society*, *51*(2), 222–246. https://doi.org/10.1177/0013124517714865

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

Tables

Table 1. Number of Unique Schools and Students by Cohort In Their Identification Year

	Treatment		Comparison	
	Schools	ITT students	Schools	ITT students
Cohort 1 (Round 1)	37	16,424	80	35,156
Cohort 2 (Rounds 2 and 3)	82	33,327	156	57,786

Table 2. Partnership Effects on Student Outcomes

Panel A. Cohort 1

	(1) M-STEP Math Levels	(2) M-STEP Math Gains	(3) M-STEP ELA Levels	(4) M-STEP ELA Gains	(5) SAT Math	(6) SAT ELA	(7) Grad Rate	(8) Dropout Rate
Partnership School 2013- 2014	0.070	0.053	0.063	0.020	0.044	-0.001	0.045	0.028
	(0.050)	(0.047)	(0.053)	(0.046)	(0.040)	(0.035)	(0.032)	(0.031)
Partnership School 2014- 2015	0.020	-0.023	0.004	-0.055	0.117*	0.020	0.061	-0.014
	(0.048)	(0.046)	(0.050)	(0.043)	(0.057)	(0.045)	(0.039)	(0.028)
Partnership School 2015- 2016	0.037	0.027	0.032	0.058	-0.023	0.060	-0.001	0.034
	(0.033)	(0.037)	(0.035)	(0.052)	(0.052)	(0.053)	(0.032)	(0.021)
Partnership School 2016- 2017	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2017- 2018	0.047	0.121*	0.066*	0.112*	0.000	0.038	0.055	-0.021
	(0.036)	(0.054)	(0.030)	(0.045)	(0.057)	(0.047)	(0.039)	(0.038)
Partnership School 2018- 2019	0.071	0.075	0.103*	0.075+	-0.002	0.064	0.058	-0.009
	(0.048)	(0.049)	(0.040)	(0.042)	(0.058)	(0.050)	(0.043)	(0.036)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	81,830	81,830	81,990	81,990	211	211	215	215
Adjusted R^2	0.678	-0.223	0.706	-0.214	0.803	0.870	0.884	0.784
Within R^2	0.026	0.020	0.027	0.017	0.254	0.078	0.218	0.087

Panel B. Cohort 2

	(1) M-STEP Math Levels	(2) M-STEP Math Gains	(3) M-STEP ELA Levels	(4) M-STEP ELA Gains	(5) SAT Math	(6) SAT ELA	(7) Grad Rate	(8) Dropout Rate
Partnership School 2013- 2014	0.042	0.043	0.008	0.004	0.039	-0.047	0.001	-0.002
	(0.047)	(0.043)	(0.045)	(0.042)	(0.052)	(0.051)	(0.034)	(0.030)
Partnership School 2014- 2015	-0.025	-0.067	-0.047	-0.074	0.055	-0.033	-0.017	0.004
	(0.047)	(0.043)	(0.046)	(0.048)	(0.049)	(0.050)	(0.033)	(0.024)
Partnership School 2015- 2016	-0.008	0.023	-0.009	0.046	0.048	0.074	0.005	0.000
	(0.035)	(0.035)	(0.039)	(0.031)	(0.041)	(0.045)	(0.023)	(0.025)
Partnership School 2016- 2017	0.029	0.049	0.025	0.040	0.033	0.055	0.010	-0.005
	(0.028)	(0.035)	(0.031)	(0.037)	(0.049)	(0.047)	(0.023)	(0.021)
Partnership School 2017- 2018	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2018- 2019	0.034+	0.055	0.018	0.035	-0.050	-0.072	-0.016	0.005
	(0.020)	(0.036)	(0.016)	(0.038)	(0.055)	(0.052)	(0.031)	(0.033)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	154,866	154,866	155,015	155,015	496	496	505	505
Adjusted R^2	0.642	-0.216	0.667	-0.212	0.636	0.709	0.899	0.750
Within R^2	0.016	0.014	0.016	0.010	0.205	0.052	0.143	0.080

Note: Student-level models (i.e., columns 1–4) include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year; robust standard errors clustered at the time-invariant school level in parentheses. School-level models (i.e., columns 5–8) include school fixed effects and time-variant treatment assignment; robust standard errors clustered at the time-variant school level in parentheses. SAT ELA represents the evidence-based reading and writing section of this assessment. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table 3. Student Outcomes, Non-Partnership Schools in Partnership Districts Removed from the Comparison Group

	Cohort 1				Cohort 2			
	(1) M-STEP Math Level	(2) M-STEP Math Gains	(3) M-STEP ELA Level	(4) M-STEP ELA Gains	(5) M-STEP Math Level	(6) M-STEP Math Gains	(7) M-STEP ELA Level	(8) M-STEP ELA Gains
Partnership School 2013-2014	0.092 (0.057)	0.057 (0.050)	0.055 (0.063)	-0.011 (0.048)	0.049 (0.050)	0.019 (0.047)	0.021 (0.049)	-0.013 (0.044)
Partnership School 2014-2015	0.021 (0.053)	-0.034 (0.050)	-0.022 (0.055)	-0.083+ (0.046)	-0.016 (0.051)	-0.086+ (0.048)	-0.034 (0.049)	-0.088+ (0.049)
Partnership School 2015-2016	0.052 (0.036)	0.057 (0.047)	0.033 (0.035)	0.082 (0.052)	-0.014 (0.040)	-0.010 (0.040)	-0.000 (0.043)	0.040 (0.035)
Partnership School 2016-2017	ref.	ref.	ref.	ref.	0.030 (0.032)	0.049 (0.039)	0.042 (0.033)	0.061 (0.042)
Partnership School 2017-2018	0.087* (0.038)	0.174** (0.057)	0.092** (0.029)	0.137** (0.046)	ref.	ref.	ref.	ref.
Partnership School 2018-2019	0.123* (0.050)	0.124* (0.052)	0.133*** (0.039)	0.100* (0.044)	0.051* (0.021)	0.088* (0.039)	0.021 (0.019)	0.059 (0.041)
Constant	0.061 (0.193)	1.076*** (0.214)	0.085 (0.107)	1.001*** (0.152)	-0.149 (0.125)	0.799*** (0.122)	-0.087 (0.120)	0.843*** (0.130)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X	X	X	X	X
Grade FE	X	X	X	X	X	X	X	X
N	59,708	59,708	60,003	60,003	116,065	116,065	116,379	116,379
Adjusted R^2	0.692	-0.224	0.721	-0.214	0.633	-0.218	0.663	-0.214
Within R^2	0.026	0.020	0.028	0.017	0.016	0.015	0.017	0.010

Note: All models include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year; robust standard errors clustered at the time-invariant school level in parentheses.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table 4. Grades 4-8 Math and ELA Achievement by Lagged Performance

Panel A. Cohort 1

	M-STEP Math				M-STEP ELA			
	Bottom quartile		Top 3 quartiles		Bottom quartile		Top 3 quartiles	
	(1) Levels	(2) Gains	(3) Levels	(4) Gains	(5) Levels	(6) Gains	(7) Levels	(8) Gains
Partnership School 2013-2014	0.157*	0.095	0.095	0.113*	0.102+	0.084	0.102	0.026
	(0.071)	(0.084)	(0.057)	(0.052)	(0.058)	(0.068)	(0.065)	(0.054)
Partnership School 2014-2015	0.054	-0.007	0.021	-0.046	-0.010	-0.019	-0.033	-0.136**
	(0.057)	(0.061)	(0.051)	(0.054)	(0.049)	(0.061)	(0.055)	(0.045)
Partnership School 2015-2016	0.057	0.022	0.015	0.015	0.068+	0.129*	0.021	0.061
	(0.046)	(0.058)	(0.034)	(0.043)	(0.039)	(0.062)	(0.036)	(0.057)
Partnership School 2016- 2017	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2017-2018	0.092*	0.183*	0.028	0.089+	0.102**	0.195**	0.051	0.087+
	(0.045)	(0.076)	(0.037)	(0.053)	(0.036)	(0.062)	(0.033)	(0.045)
Partnership School 2018-2019	0.126*	0.096	0.052	0.074	0.093*	0.079	0.102*	0.070
	(0.055)	(0.066)	(0.049)	(0.051)	(0.045)	(0.050)	(0.042)	(0.045)
Constant	-1.233***	0.363+	0.391+	1.282***	-1.296***	0.397+	0.417**	1.156***
	(0.192)	(0.213)	(0.223)	(0.208)	(0.174)	(0.232)	(0.135)	(0.171)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X	X	X	X	X
Grade FE	X	X	X	X	X	X	X	X
N	16,095	16,095	52,642	52,642	15,903	15,903	53,052	53,052
Adjusted R^2	0.411	-0.063	0.684	-0.185	0.418	-0.030	0.703	-0.191
Within R^2	0.189	0.148	0.032	0.037	0.213	0.166	0.035	0.030

Panel B. Cohort 2

	M-STEP Math				M-STEP ELA			
	Bottom quartile		Top 3 quartiles		Bottom quartile		Top 3 quartiles	
	(1) Level	(2) Gains	(3) Level	(4) Gains	(5) Level	(6) Gains	(7) Level	(8) Gains
Partnership School 2013-2014	0.113*	0.067	0.049	0.038	0.007	-0.024	0.040	0.026
	(0.055)	(0.064)	(0.050)	(0.043)	(0.060)	(0.061)	(0.049)	(0.048)
Partnership School 2014-2015	-0.005	-0.062	-0.031	-0.080	-0.033	-0.028	-0.056	-0.107+
	(0.053)	(0.059)	(0.050)	(0.051)	(0.051)	(0.059)	(0.050)	(0.059)
Partnership School 2015-2016	0.038	0.079	-0.025	0.007	-0.012	0.037	-0.003	0.052
	(0.038)	(0.048)	(0.036)	(0.038)	(0.036)	(0.037)	(0.040)	(0.035)
Partnership School 2016-2017	0.073*	0.081	0.023	0.047	0.022	0.035	0.029	0.042
	(0.033)	(0.053)	(0.029)	(0.036)	(0.032)	(0.044)	(0.032)	(0.040)
Partnership School 2017- 2018	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2018-2019	0.075**	0.123*	0.021	0.040	0.054**	0.073+	0.005	0.026
	(0.026)	(0.048)	(0.021)	(0.036)	(0.021)	(0.042)	(0.018)	(0.039)
Constant	-1.244***	0.240	0.242+	0.776***	-1.255***	0.056	0.279+	0.865***
	(0.158)	(0.172)	(0.137)	(0.128)	(0.148)	(0.146)	(0.147)	(0.146)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X	X	X	X	X
Grade FE	X	X	X	X	X	X	X	X
N	27,060	27,060	90,992	90,992	26,308	26,308	91,984	91,984
Adjusted R^2	0.334	-0.047	0.643	-0.170	0.326	-0.041	0.666	-0.182
Within R^2	0.173	0.165	0.037	0.030	0.166	0.169	0.036	0.021

Note: Prior achievement measured in identification year. Bottom quartile (Cols 1-2 and 5-6) are those in the bottom quartile of their school in the identification year. Top 3 (Cols 3-4 and 7-8) quartiles are those above the 25th percentile of their school in the identification year. Note the sample is smaller than the main sample because to be in this sample, students need to have an M-STEP score in the identification year. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table 5. Teacher Perceptions in Partnership Schools and Non-Partnership Schools in Partnership Districts

	Partnership	Non-Partnership	Difference (Partnership - Non-Partnership)	p-value
Staffing	-0.182	0.128	-0.311***	0.000
School leadership	0.064	-0.043	0.107**	0.001
School accountability	0.320	-0.236	0.556***	0.000
Observations	5474			

Note: Factor scores are based on responses from when teachers were in the first year of their school's Partnership reform.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Figures

Figure 1. Partnership Model Theory of Change

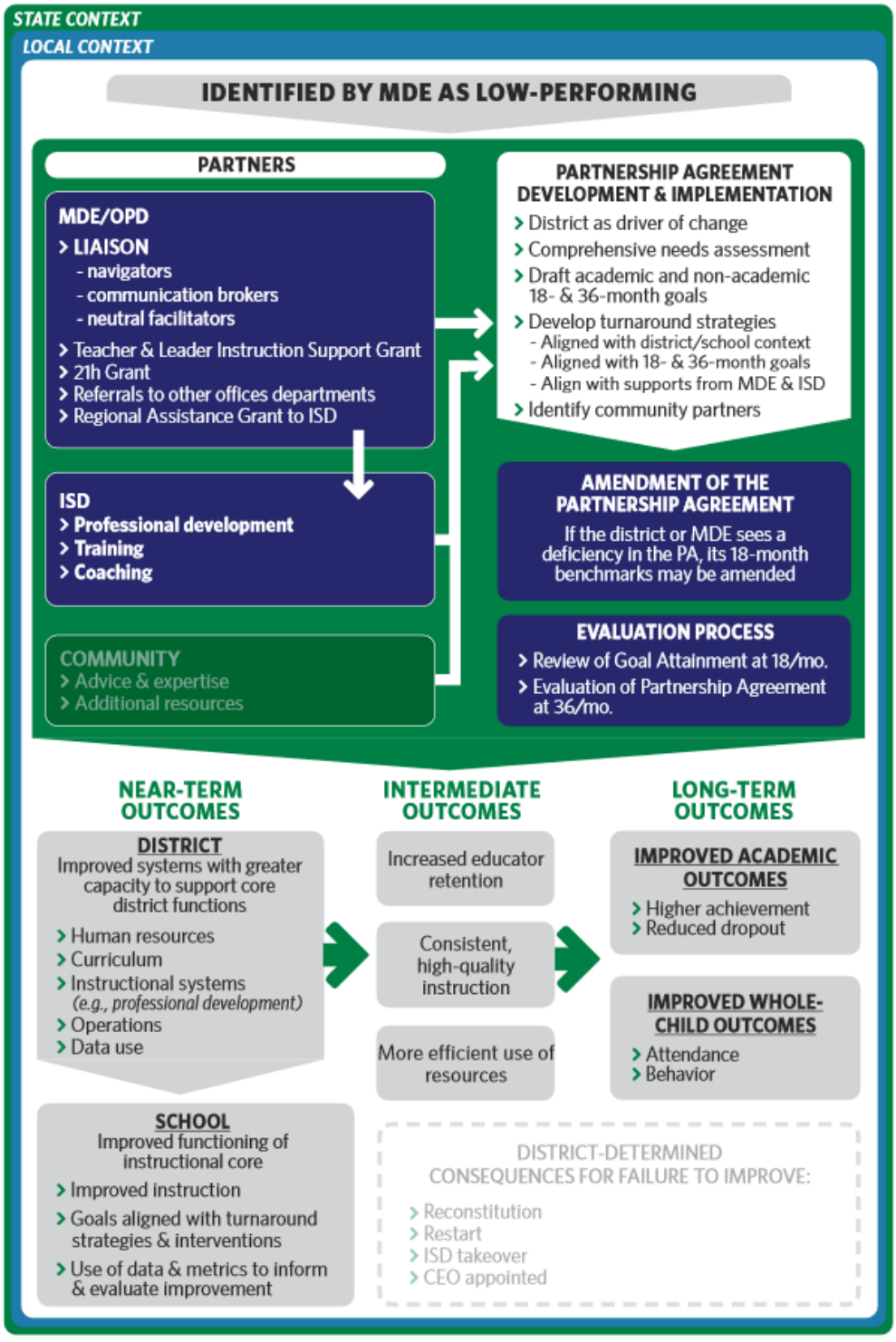
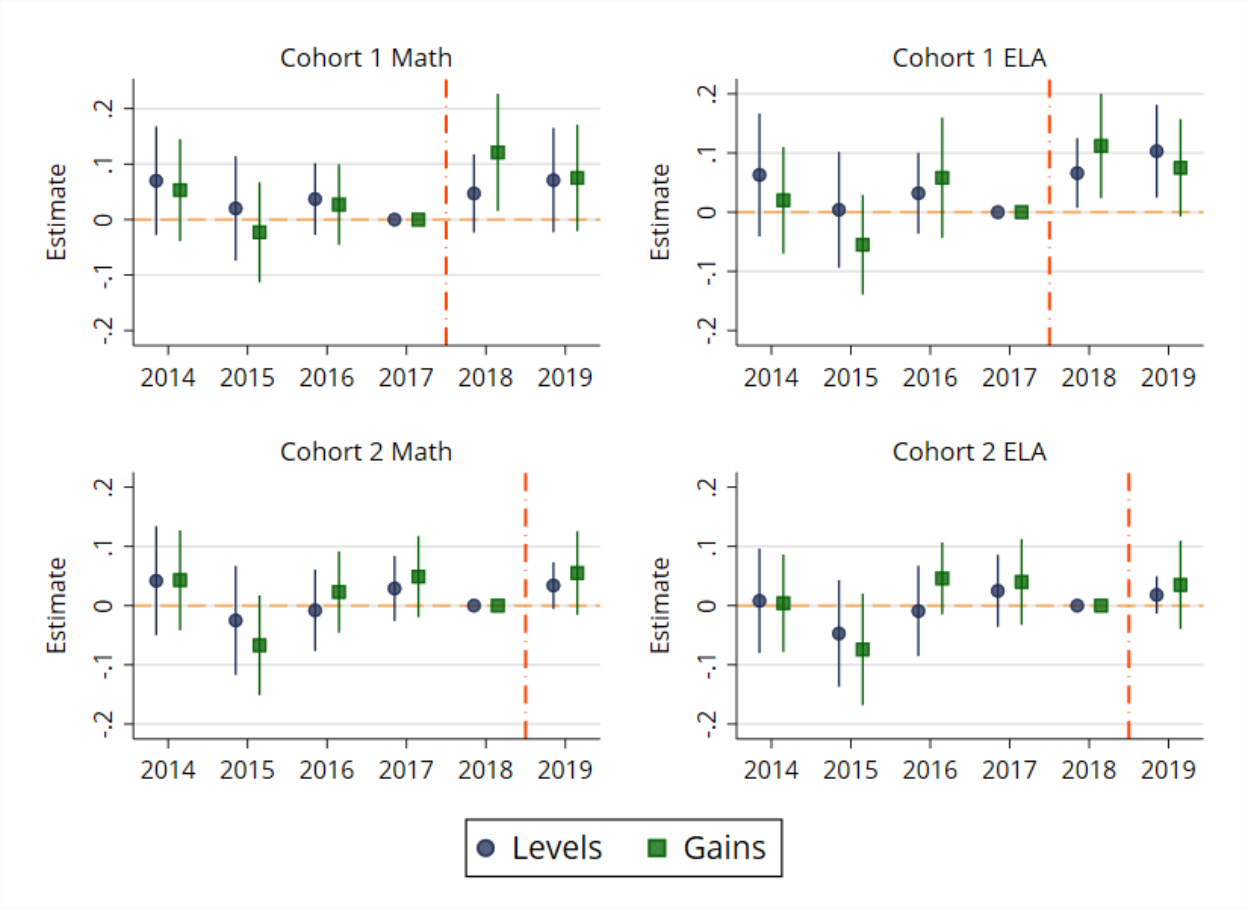


Figure 2. Event Study Estimates by Subject and Cohort



Note: Graphs illustrate event study point estimates on the *Partnership x year* indicators, in standard deviation units, with 95% confidence intervals for student achievement in math and ELA in grades 4-8 on the M-STEP, with the reference year represented by the dotted vertical line.

Online Appendix A. Supplemental Tables and Figures

Table A-1. Differences Between Stayers and Leavers, Treatment Year 1

Panel A. Cohort 1

	Treated			Comparison		
	Transfers	Stayers	Difference	Transfers	Stayers	Difference
Math	-1.249	-1.102	-0.147***	-0.881	-0.740	-0.141***
ELA	-1.132	-1.037	-0.095***	-0.770	-0.675	-0.095***
Economic disadvantage	0.933	0.904	0.028***	0.904	0.854	0.051***
English Learner	0.020	0.039	-0.019***	0.075	0.160	-0.085***
Special education	0.178	0.186	-0.008	0.153	0.147	0.006
Black	0.905	0.885	0.021***	0.631	0.554	0.076***
Hispanic or Latinx	0.036	0.062	-0.026***	0.094	0.184	-0.090***

Panel B. Cohort 2

	Treated			Comparison		
	Transfers	Stayers	Difference	Transfers	Stayers	Difference
Math	-1.128	-1.004	-0.125***	-0.927	-0.764	-0.163***
ELA	-1.062	-0.947	-0.115***	-0.845	-0.708	-0.137***
Economic disadvantage	0.938	0.916	0.021***	0.910	0.884	0.025***
English Learner	0.031	0.059	-0.029***	0.077	0.128	-0.051***
Special education	0.174	0.178	-0.004	0.136	0.141	-0.004
Black	0.853	0.831	0.022***	0.736	0.659	0.077***
Hispanic or Latinx	0.050	0.078	-0.028***	0.068	0.107	-0.039***

Note: Coefficients from *t*-tests on school-level differences between student transfers and stayers (transfers minus stayers).

*** p<0.001, ** p<0.01, * p<0.05

Table A-2. School-level Student Descriptives for Both Cohorts In Their Identification Year

	Cohort 1		Cohort 2		All other Schools	
	Partnership	Comparison	Partnership	Comparison	2016	2017
Economically disadvantaged	85.8 (6.7)	82.1 (15.8)	91.5 (7.7)	86.2 (12.1)	51.3 (24.5)	56.1 (24.4)
English Language Learner	2.4 (6.2)	12.0 (23.1)	6.3 (14.0)	9.9 (20.3)	6.0 (12.3)	6.1 (12.4)
Black	90.1 (14.6)	55.6 (35.1)	80.3 (24.1)	56.8 (35.9)	13.6 (22.8)	13.9 (23.2)
Hispanic or Latinx	4.0 (8.4)	12.9 (21.7)	7.0 (14.5)	9.2 (16.2)	7.7 (10.8)	7.8 (10.8)
Other Non-White[1]	2.7 (3.7)	5.9 (5.5)	4.3 (5.7)	5.5 (5.6)	7.7 (8.6)	8.0 (8.7)
White	3.1 (5.2)	25.5 (29.4)	8.4 (13.6)	28.5 (30.8)	70.9 (26.3)	70.4 (26.7)
Special education	19.8 (8.6)	14.5 (6.2)	16.8 (5.4)	14.6 (8.5)	18.1 (20.3)	18.6 (21.0)
Observations	37	80	82	156	3164	3144

NOTE: School-level means from identification year for each cohort, i.e., 2016-17 for Cohort 1 and 2017-18 for Cohort 2 with standard deviations in parentheses. All other schools category includes schools not in either treatment or comparison group.

[1]Due to small *N*s, Asian, Native American of Alaska Native, Native Hawaiian or Other Pacific Islander, and Two or More Races were combined into the group "Other Non-White."

Table A-3. Teacher Survey Response Rates and Respondent Characteristics

Panel A. Response rates by treatment condition and year

	2018	2019	Total
Partnership Schools	2,641 (42.3%)	2,424 (57.1%)	5,065 (49.4%)
Non-Partnership Schools	4,462 (35.9%)	4,459 (44.8%)	8,921 (40.3%)
Total	7,103 (38.3%)	6,883 (49.1%)	13,986 (43.6%)

Panel B. Differences in respondents and nonrespondents by treatment condition

	Partnership			Non-Partnership		
	Respondents	Non-respondents	Diff	Respondents	Non-respondents	Diff
Black	0.490	0.527	0.037**	0.231	0.223	-0.007
Hispanic or Latinx	0.014	0.013	-0.001	0.038	0.038	0.000
White	0.443	0.402	-0.041**	0.683	0.693	0.010
Other Non-White[1]	0.053	0.058	0.005	0.049	0.046	-0.003
Male	0.169	0.232	0.063***	0.193	0.242	0.049***
Years of experience	13.728	13.894	0.166	13.749	13.548	-0.202
Elementary certified	0.637	0.612	-0.024	0.610	0.581	-0.030**
Secondary certified	0.323	0.308	-0.014	0.350	0.370	0.020
Observations	2,441	2,564		3,501	5,380	

Note: Panel A shows the number of teachers who were invited to participate in the survey by Partnership status and year, with group response rate shown in parentheses. Panel B shows proportions by group, with the exception of years of experience, which is presented in years.

[1]Due to small *N*s, Asian, Native American of Alaska Native, Native Hawaiian or Other Pacific Islander, and Two or More Races were combined into the group "Other Non-White."

Table A-4. Item-level Survey Descriptives by Partnership School Status

	Partnership	Non-Partnership	Total
Staffing			
There is a high rate of staff turnover[1]	2.81 (1.15)	3.11 (1.25)	2.99 (1.22)
Teacher retention[2]	3.11 (1.20)	3.44 (1.27)	3.30 (1.25)
Staff retention[2]	3.07 (1.18)	3.34 (1.25)	3.23 (1.23)
Teacher attendance[2]	3.68 (0.96)	3.92 (0.91)	3.82 (0.94)
School leadership			
Working with staff to meet curriculum standards[3]	3.32 (1.14)	3.22 (1.14)	3.26 (1.14)
Communicating central mission of the school[3]	3.57 (1.14)	3.44 (1.12)	3.49 (1.13)
Using evidence to make data-driven decisions[3]	3.54 (1.14)	3.42 (1.11)	3.47 (1.12)
Establishing clear discipline policies[3]	3.08 (1.27)	2.98 (1.26)	3.02 (1.27)
Working with community partners[3]	3.32 (1.18)	3.26 (1.17)	3.28 (1.18)
Facilitating and encouraging professional development[3]	3.49 (1.17)	3.31 (1.17)	3.38 (1.17)
Encouraging parental engagement[3]	3.35 (1.17)	3.29 (1.17)	3.32 (1.17)
Communicating improvement strategy and goals[3]	3.55 (1.14)	3.35 (1.16)	3.43 (1.16)
Accountability			
The school will receive a low accountability[4]	3.77 (1.12)	3.31 (1.26)	3.51 (1.22)
The school will lose students[4]	3.14 (1.27)	2.81 (1.34)	2.95 (1.32)
The school will face staff and leader removal[4]	3.27 (1.25)	2.57 (1.31)	2.87 (1.33)
The school will be closed[4]	2.72 (1.32)	1.88 (1.15)	2.23 (1.29)
Total observations	5,072		

Note: Item-level means with standard deviations in parentheses.

[1]Teachers were asked "Please indicate the extent to which you agree or disagree with the following statements about your school… The response scale was: (1) Strongly disagree (2) Disagree (3) Neither agree nor disagree (4) Agree (5) Strongly agree. This scale was reverse coded for analysis so that a higher value indicates that the teacher perceived there to be a lower rate of staff turnover.

[2]Teachers were asked "We are interested in how well you believe your school is implementing activities in the following areas. Please give your school a grade, from A(high) to F (low) in each of the following areas…" The

response scale was (1) F (2) D (3) C (4) B (5) A

[3]Teachers were asked "Indicate how effectively your principal or school leader performed each of the following…" The response scale was (1) Not at all effectively (2) Slightly ineffectively (3) Somewhat effectively (4) Very effectively (5) Extremely effectively

[4]Teachers were asked "If your school's improvement goals are not met, to what extent do you believe that your school will face the following consequences…" The response scale was (1) Very unlikely (2) Somewhat unlikely (3) Neither unlikely nor likely (4) Somewhat likely (5) Very likely

Table A-5. Factor loadings

Panel A. Staffing

	Staffing climate	Psi (Uniqueness)
High rate of staff turnover[1]	0.774	0.401
Teacher retention	0.924	0.145
Staff retention	0.916	0.161
Teacher attendance	0.617	0.619
N	4,127	

Note: Scale reliability coefficient=0.833

[1] Reverse-coded from the original item so that a higher value indicates that the teacher perceived there to be a lower rate of staff turnover.

Panel B. School leadership

	School leadership	Psi (Uniqueness)
Leader works with staff to meet curriculum standards	0.897	0.195
Leader communicates central mission of the school	0.891	0.207
Leader uses evidence to make data-driven decisions	0.895	0.199
Leader establishes clear discipline policies	0.828	0.314
Leader works with community partners	0.866	0.251
Leader facilitates and encourages PD	0.881	0.224
Leader encourages parental engagement	0.861	0.259
Leader communicates school strategies/goals with teachers [2]	0.909	0.174
Observations	3908	

Note: Scale reliability coefficient=0.957

[2] The 2018 survey asked these items separately (how well the school leader communicates goals, and how well the school leader communicates strategies). Due to an almost perfect correlation between responses, the items were combined in the 2019 survey (how well the school leader communicates goals and strategies). In order to retain the item in the factor, we took the mean of the two items in 2018 and used the mean in the factor analysis.

Panel C. School Accountability

	Accountability Perceptions	Psi (Uniqueness)
How likely school receives low accountability score if goals not met?	0.783	0.386
How likely school loses students if goals not met?	0.816	0.334
How likely school faces staff and leader removal if goals not met?	0.865	0.253
How likely school be closed if goals not met?	0.764	0.416
Observations	4541	

Note: Scale reliability coefficient=0.822

Table A-6. Event Study Estimates for Rounds 2 and 3

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership 2013- 2014	0.041	0.043	0.007	0.004	0.016	-0.077	0.001	-0.003
	(0.047)	(0.043)	(0.045)	(0.042)	(0.057)	(0.054)	(0.035)	(0.029)
Partnership 2014- 2015	-0.025	-0.067	-0.047	-0.074	0.050	-0.060	-0.017	0.004
	(0.047)	(0.043)	(0.046)	(0.049)	(0.052)	(0.053)	(0.033)	(0.024)
Partnership 2015- 2016	-0.009	0.023	-0.009	0.047	-0.064	-0.120*	0.005	0.000
	(0.035)	(0.035)	(0.039)	(0.031)	(0.044)	(0.051)	(0.023)	(0.025)
Partnership 2016- 2017	0.028	0.049	0.025	0.040	-0.041	-0.075	0.010	-0.005
	(0.028)	(0.035)	(0.031)	(0.037)	(0.051)	(0.048)	(0.023)	(0.021)
Partnership 2017- 2018	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Round 2 x 2018-2019	0.073**	0.050	0.049*	0.019	-0.064	-0.081	-0.038	-0.019
	(0.026)	(0.037)	(0.020)	(0.038)	(0.065)	(0.064)	(0.049)	(0.038)
Round 3 x 2018-2019	-0.017	0.062	-0.022	0.056	-0.070	-0.096	-0.000	0.021
	(0.029)	(0.045)	(0.022)	(0.048)	(0.073)	(0.071)	(0.041)	(0.046)
Constant	-0.024	0.840***	0.060	0.894***	-1.406**	-1.056**	0.495**	0.395**
	(0.120)	(0.111)	(0.126)	(0.117)	(0.420)	(0.364)	(0.173)	(0.140)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	154,866	154,866	155,015	155,015	496	496	505	505
Adjusted R^2	0.642	-0.216	0.667	-0.212	0.616	0.691	0.899	0.750
Within R^2	0.017	0.014	0.017	0.010	0.184	0.054	0.145	0.082

Note: Student-level models (i.e. columns 1-4) include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year. Robust standard errors clustered at the time-invariant school level in parentheses. School-level models (i.e. columns 5-8) include school fixed effects and time-variant treatment assignment; robust standard errors clustered at the time-variant school level in parentheses. SAT ELA represents the evidence-based reading and writing section of this assessment. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A-7. Partnership Student Outcomes with the Year Prior to Identification as the Reference Year

Panel A. Cohort 1 (2015-16 as reference year)

	(1) M-STEP Math Levels	(2) M-STEP Math Gains	(3) M-STEP ELA Levels	(4) M-STEP ELA Gains	(5) SAT Math	(6) SAT ELA	(8) Grad Rate	(9) Dropout Rate
Partnership School 2013- 2014	0.033	0.025	0.036	-0.034	0.048	0.039	0.005	0.043
	(0.036)	(0.046)	(0.038)	(0.045)	(0.057)	(0.053)	(0.064)	(0.058)
Partnership School 2014- 2015	-0.019	-0.054	-0.026	-0.111*	0.115+	0.059	0.035	0.022
	(0.028)	(0.040)	(0.036)	(0.054)	(0.058)	(0.066)	(0.046)	(0.034)
Partnership School 2015- 2016	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2016- 2017	-0.039	-0.029	-0.033	-0.060	0.013	0.043	-0.025	0.008
	(0.032)	(0.037)	(0.035)	(0.051)	(0.052)	(0.056)	(0.040)	(0.033)
Partnership School 2017- 2018	0.010	0.092+	0.035	0.055	0.005	0.082	0.027	0.027
	(0.039)	(0.051)	(0.038)	(0.043)	(0.080)	(0.056)	(0.061)	(0.052)
Partnership School 2018- 2019	0.032	0.045	0.071+	0.018	0.002	0.133+	0.034	0.018
	(0.047)	(0.046)	(0.041)	(0.046)	(0.080)	(0.072)	(0.063)	(0.048)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	82,174	82,174	82,340	82,340	211	211	217	217

Panel B. Cohort 2 (2016-17 as reference year)

	(1) M-STEP Math Levels	(2) M-STEP Math Gains	(3) M-STEP ELA Levels	(4) M-STEP ELA Gains	(5) SAT Math	(6) SAT ELA	(7) Grad Rate	(8) Dropout Rate
Partnership School 2013-2014	0.015	-0.009	-0.014	-0.032	0.057	-0.002	-0.042	0.056
	(0.032)	(0.031)	(0.034)	(0.031)	(0.058)	(0.045)	(0.036)	(0.044)
Partnership School 2014-2015	-0.053+	-0.119**	-0.071*	-0.115**	0.091*	0.015	0.076	0.020
	(0.032)	(0.037)	(0.028)	(0.038)	(0.045)	(0.041)	(0.059)	(0.027)
Partnership School 2015-2016	-0.036+	-0.027	-0.033	0.008	-0.023	-0.044	0.046	0.050
	(0.020)	(0.029)	(0.022)	(0.027)	(0.041)	(0.042)	(0.043)	(0.050)
Partnership School 2016-2017	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2017-2018	-0.028	-0.048	-0.025	-0.039	0.041	0.075	0.004	0.007
	(0.028)	(0.035)	(0.031)	(0.037)	(0.051)	(0.048)	(0.023)	(0.037)
Partnership School 2018-2019	0.007	0.007	-0.006	-0.004	0.026	0.016	0.019	0.027
	(0.032)	(0.029)	(0.029)	(0.030)	(0.057)	(0.057)	(0.039)	(0.043)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	155,683	155,683	155,862	155,862	496	496	520	520

Note: Student-level models (i.e., columns 1–4) include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year; robust standard errors clustered at the time-invariant school level in parentheses. School-level models (i.e., columns 5–8) include school fixed effects and time-variant treatment assignment; robust standard errors clustered at the time-variant school level in parentheses. SAT ELA represents the evidence-based reading and writing section of this assessment.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table A-8. Student Outcomes, All-Starters Assignment to Treatment and Comparison

	Cohort 1				Cohort 2			
	M-STEP Math		M-STEP ELA		M-STEP Math		M-STEP ELA	
	(1) Level	(2) Gains	(3) Level	(4) Gains	(5) Level	(6) Gains	(7) Level	(8) Gains
Partnership School 2013-2014	0.032	0.027	0.009	-0.000	0.006	-0.013	-0.013	-0.058+
	(0.042)	(0.045)	(0.039)	(0.037)	(0.042)	(0.037)	(0.038)	(0.035)
Partnership School 2014-2015	-0.003	-0.023	0.006	-0.009	-0.066+	-0.118***	-0.072*	-0.127**
	(0.041)	(0.047)	(0.038)	(0.050)	(0.038)	(0.033)	(0.036)	(0.039)
Partnership School 2015-2016	0.051+	0.072+	0.028	0.049	-0.034	-0.005	-0.030	0.007
	(0.029)	(0.040)	(0.027)	(0.035)	(0.030)	(0.028)	(0.029)	(0.024)
Partnership School 2016-2017	ref.	ref.	ref.	ref.	-0.006	0.010	-0.004	0.010
					(0.021)	(0.027)	(0.023)	(0.027)
Partnership School 2017-2018	0.050	0.144**	0.038	0.092*	ref.	ref.	ref.	ref.
	(0.037)	(0.044)	(0.036)	(0.045)				
Partnership School 2018-2019	0.053	0.086+	0.056	0.077*	0.032	0.031	0.030	0.032
	(0.047)	(0.046)	(0.042)	(0.038)	(0.024)	(0.028)	(0.021)	(0.030)
Constant	0.211	1.019***	0.057	0.936***	0.110	0.958***	0.228+	1.047***
	(0.174)	(0.148)	(0.105)	(0.095)	(0.116)	(0.109)	(0.126)	(0.103)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X	X	X	X	X
Grade FE	X	X	X	X	X	X	X	X
N	79,453	79,453	79,627	79,627	150,474	150,474	150,853	150,853
Adjusted R2	0.666	-0.226	0.699	-0.219	0.646	-0.225	0.671	-0.216
Within R2	0.024	0.017	0.023	0.015	0.015	0.012	0.016	0.010

Note: All models contain year indicators, *Partnership x year* indicators, time-variant student characteristics (economic disadvantaged status, disability status, English learner status, grade level), school-level student demographics, and student fixed effects, with robust standard errors clustered by school in parentheses.
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A-9. Event Study Estimates with Covariates Measured at Baseline and Interacted with a Linear Time Trend

Panel A. Cohort 1

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership School 2013- 2014	0.047	0.081	-0.010	0.017	0.033	0.011	-0.057	0.053
	(0.063)	(0.051)	(0.058)	(0.048)	(0.069)	(0.054)	(0.054)	(0.043)
Partnership School 2014- 2015	0.004	-0.010	-0.040	-0.058	0.098	0.022	0.012	0.003
	(0.053)	(0.049)	(0.057)	(0.047)	(0.065)	(0.053)	(0.058)	(0.036)
Partnership School 2015- 2016	0.036	0.035	0.020	0.058	-0.012	-0.039	-0.021	0.036
	(0.035)	(0.037)	(0.036)	(0.052)	(0.051)	(0.058)	(0.041)	(0.028)
Partnership School 2016- 2017	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2017- 2018	0.058	0.113*	0.094**	0.120*	-0.006	0.041	0.054	-0.021
	(0.035)	(0.055)	(0.028)	(0.048)	(0.062)	(0.054)	(0.049)	(0.039)
Partnership School 2018- 2019	0.083+	0.057	0.146***	0.086+	-0.006	0.086+	0.101+	-0.026
	(0.047)	(0.050)	(0.037)	(0.045)	(0.067)	(0.047)	(0.051)	(0.033)
Constant	-0.459***	0.353***	-0.328***	0.480***	-0.931***	-0.932***	0.587***	0.208***
	(0.046)	(0.076)	(0.041)	(0.050)	(0.014)	(0.015)	(0.010)	(0.010)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	81,830	81,830	81,990	81,990	211	211	215	215
Adjusted R^2	0.678	-0.225	0.706	-0.216	0.802	0.857	0.871	0.779
Within R^2	0.026	0.018	0.028	0.015	0.282	0.077	0.129	0.069

Panel B. Cohort 2

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership School 2013-2014	0.062 (0.043)	0.063 (0.043)	0.028 (0.038)	0.026 (0.040)	0.016 (0.055)	-0.024 (0.056)	0.030 (0.037)	-0.037 (0.032)
Partnership School 2014-2015	-0.013 (0.039)	-0.051 (0.040)	-0.037 (0.037)	-0.058 (0.046)	0.041 (0.049)	-0.032 (0.053)	0.007 (0.033)	-0.028 (0.026)
Partnership School 2015-2016	-0.005 (0.030)	0.031 (0.035)	-0.007 (0.033)	0.055+ (0.030)	-0.070+ (0.041)	-0.101* (0.048)	0.019 (0.026)	-0.019 (0.024)
Partnership School 2016-2017	0.027 (0.026)	0.052 (0.035)	0.023 (0.029)	0.043 (0.037)	-0.039 (0.049)	-0.063 (0.047)	0.015 (0.022)	-0.012 (0.022)
Partnership School 2017- 18	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2018-2019	0.034 (0.021)	0.053 (0.036)	0.017 (0.018)	0.031 (0.039)	-0.058 (0.059)	-0.089+ (0.052)	-0.025 (0.031)	0.012 (0.034)
Constant	-0.329*** (0.054)	0.460*** (0.053)	-0.172** (0.054)	0.533*** (0.052)	-0.931*** (0.016)	-0.899*** (0.016)	0.500*** (0.008)	0.245*** (0.007)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	154,866	154,866	155,015	155,015	496	496	505	505
Adjusted R^2	0.643	-0.217	0.669	-0.213	0.623	0.697	0.897	0.745
Within R^2	0.020	0.013	0.023	0.009	0.196	0.070	0.124	0.058

Note: Student-level models (i.e., columns 1–4) include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year; robust standard errors clustered at the time-invariant school level in parentheses. School-level models (i.e., columns 5–8) include school fixed effects and time-variant treatment assignment; robust standard errors clustered at the time-variant school level in parentheses. SAT ELA represents the evidence-based reading and writing section of this assessment.

*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table A-10. Event Study Estimates with Standard Errors Clustered at the District Level

Panel A. Cohort 1

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership School 2013- 2014	0.070	0.053	0.063	0.020	0.035	-0.004	0.045+	0.028
	(0.044)	(0.058)	(0.059)	(0.051)	(0.037)	(0.036)	(0.023)	(0.029)
Partnership School 2014- 2015	0.020	-0.023	0.004	-0.055	0.102+	0.016	0.061+	-0.014
	(0.042)	(0.046)	(0.060)	(0.042)	(0.056)	(0.042)	(0.034)	(0.027)
Partnership School 2015- 2016	0.037	0.027	0.032	0.058	-0.013	-0.043	-0.001	0.034
	(0.036)	(0.042)	(0.039)	(0.051)	(0.052)	(0.060)	(0.031)	(0.023)
Partnership School 2016-2017	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2017- 2018	0.047	0.121+	0.066*	0.112+	-0.008	0.039	0.055	-0.021
	(0.030)	(0.062)	(0.031)	(0.060)	(0.061)	(0.054)	(0.037)	(0.045)
Partnership School 2018- 2019	0.071	0.075	0.103*	0.075+	-0.011	0.090	0.058	-0.009
	(0.053)	(0.063)	(0.043)	(0.042)	(0.073)	(0.054)	(0.049)	(0.038)
Constant	0.031	1.064***	0.051	1.032***	-1.427*	-1.288+	0.695	0.440
	(0.222)	(0.215)	(0.165)	(0.147)	(0.625)	(0.684)	(0.528)	(0.389)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	81,830	81,830	81,990	81,990	211	211	215	215
Adjusted R^2	0.678	-0.223	0.706	-0.214	0.794	0.855	0.884	0.784
Within R^2	0.026	0.020	0.027	0.017	0.254	0.067	0.218	0.087

Panel B. Cohort 2

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership School 2013-2014	0.042	0.043	0.008	0.004	0.016	-0.077	0.001	-0.002
	(0.064)	(0.045)	(0.051)	(0.063)	(0.063)	(0.059)	(0.026)	(0.023)
Partnership School 2014-2015	-0.025	-0.067+	-0.047	-0.074	0.050	-0.060	-0.017	0.004
	(0.060)	(0.039)	(0.057)	(0.070)	(0.062)	(0.057)	(0.025)	(0.020)
Partnership School 2015-2016	-0.008	0.023	-0.009	0.046	-0.064	-0.120*	0.005	0.000
	(0.036)	(0.032)	(0.045)	(0.036)	(0.052)	(0.049)	(0.020)	(0.028)
Partnership School 2016-2017	0.029	0.049	0.025	0.040	-0.041	-0.075+	0.010	-0.005
	(0.025)	(0.033)	(0.042)	(0.048)	(0.049)	(0.040)	(0.017)	(0.018)
Partnership School 2017-2018	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2018- 2019	0.034	0.055	0.018	0.035	-0.067	-0.091+	-0.016	0.005
	(0.026)	(0.036)	(0.015)	(0.053)	(0.054)	(0.047)	(0.026)	(0.036)
Constant	-0.031	0.841***	0.054	0.897***	-1.404**	-1.053**	0.488**	0.388**
	(0.126)	(0.128)	(0.144)	(0.112)	(0.413)	(0.353)	(0.178)	(0.133)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	154,866	154,866	155,015	155,015	496	496	505	505
Adjusted R^2	0.642	-0.216	0.667	-0.212	0.617	0.692	0.899	0.750
Within R^2	0.016	0.014	0.016	0.010	0.184	0.054	0.143	0.080

Note: Student-level models (i.e., columns 1–4) include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year; robust standard errors clustered at the time-invariant district level in parentheses. School-level models (i.e., columns 5–8) include school fixed effects and time-variant treatment assignment; robust standard errors clustered at the time-variant district level in parentheses. SAT ELA represents the evidence-based reading and writing section of this assessment.*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table A-11. Event Study Estimates Including Low-performing Non-Partnership Schools in Partnership Districts

Panel A. Cohort 1

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership 2013- 2014	0.049	0.025	-0.008	-0.057	0.052	-0.002	0.006	0.064+
	(0.058)	(0.045)	(0.061)	(0.049)	(0.046)	(0.042)	(0.031)	(0.033)
Partnership 2014- 2015	-0.006	-0.039	-0.062	-0.089+	0.126*	-0.027	0.020	0.013
	(0.055)	(0.051)	(0.051)	(0.048)	(0.060)	(0.049)	(0.037)	(0.028)
Partnership 2015- 2016	0.030	0.056	0.003	0.060	0.016	-0.057	-0.015	0.037+
	(0.039)	(0.046)	(0.037)	(0.049)	(0.058)	(0.063)	(0.031)	(0.022)
Partnership 2017- 2018	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2017 - 2018	0.070+	0.169**	0.069*	0.122**	0.008	0.044	0.045	-0.007
	(0.039)	(0.056)	(0.033)	(0.046)	(0.065)	(0.062)	(0.046)	(0.043)
Partnership School 2018 - 2019	0.105*	0.118*	0.109**	0.084+	0.002	0.088+	0.031	0.004
	(0.050)	(0.051)	(0.040)	(0.044)	(0.063)	(0.051)	(0.046)	(0.038)
Near-selected Partnership district school 2017-2018	0.093**	0.133**	0.060	0.052	0.050	0.067	0.059	0.027
	(0.035)	(0.047)	(0.040)	(0.059)	(0.068)	(0.075)	(0.056)	(0.052)
Near-selected Partnership district school 2018-2019	0.125**	0.117**	0.065+	0.047	0.031	0.026	-0.022	0.021
	(0.045)	(0.041)	(0.037)	(0.048)	(0.075)	(0.070)	(0.042)	(0.049)
Constant	0.047	1.079***	0.051	1.020***	-1.406*	-1.356*	0.592	0.495
	(0.195)	(0.197)	(0.116)	(0.147)	(0.670)	(0.604)	(0.693)	(0.472)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	81,830	81,830	81,990	81,990	211	211	215	215
Adjusted R^2	0.678	-0.221	0.706	-0.213	0.793	0.855	0.882	0.783
Within R^2	0.027	0.021	0.028	0.018	0.259	0.077	0.214	0.096

Panel B. Cohort 2

	M-STEP Math Levels	M-STEP Math Gains	M-STEP ELA Levels	M-STEP ELA Gains	SAT Math	SAT ELA	Grad Rate	Dropout Rate
Partnership 2013- 2014	0.023	-0.018	0.014	-0.032	0.014	-0.032	-0.021	0.018
	(0.044)	(0.041)	(0.046)	(0.037)	(0.051)	(0.050)	(0.027)	(0.025)
Partnership 2014- 2015	-0.008	-0.071	-0.016	-0.068	0.081	-0.001	-0.024	0.014
	(0.045)	(0.044)	(0.047)	(0.042)	(0.050)	(0.050)	(0.027)	(0.025)
Partnership 2015- 2016	-0.018	-0.042	0.003	0.013	-0.071+	-0.098*	0.002	-0.005
	(0.037)	(0.038)	(0.040)	(0.031)	(0.040)	(0.048)	(0.023)	(0.020)
Partnership 2016- 2017	0.016	0.026	0.037	0.055	-0.026	-0.046	0.013	-0.004
	(0.028)	(0.033)	(0.030)	(0.036)	(0.043)	(0.045)	(0.020)	(0.019)
Partnership 2017- 2018	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Partnership School 2018- 2019	0.050*	0.081*	0.024	0.057	-0.069	-0.071	-0.022	0.003
	(0.022)	(0.037)	(0.020)	(0.039)	(0.057)	(0.055)	(0.033)	(0.033)
Near-selected Partnership district school 2018- 2019	0.051+	0.104*	0.003	0.072	-0.067	-0.094	-0.019	-0.031
	(0.028)	(0.046)	(0.022)	(0.047)	(0.053)	(0.058)	(0.022)	(0.026)
Constant	-0.010	0.905***	0.056	0.934***	-1.395***	-1.019**	0.480**	0.391**
	(0.119)	(0.112)	(0.120)	(0.116)	(0.404)	(0.360)	(0.168)	(0.140)
School covariates	X	X	X	X	X	X	X	X
Student covariates	X	X	X	X				
Grade FE	X	X	X	X				
N	154,866	154,866	155,015	155,015	496	496	505	505
Adjusted R^2	0.641	-0.216	0.667	-0.212	0.622	0.693	0.900	0.751
Within R^2	0.016	0.014	0.016	0.010	0.197	0.058	0.147	0.086

Note: Student-level models (i.e. columns 1-4) include student fixed effects and time-invariant treatment assignment based on school assignment in the identification year' robust standard errors clustered at the time-invariant school level in parentheses. School-level models (i.e. columns 5-8) include school fixed effects and time-variant treatment assignment; robust standard errors clustered at the time-variant school level in parentheses. SAT ELA represents the evidence based reading and writing section of this assessment. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A-12. Teacher Perceptions in Partnership Schools and Non-Partnership Schools In Partnership Districts (Weighted)

	Partnership	Non-Partnership	Difference (Partnership - Non-Partnership)	p-value
Staffing	-0.287	0.144	-0.432	0.000***
School leadership	-0.021	-0.057	0.035	0.487
School accountability	0.421	-0.213	0.633	0.000***
Total observations	5,474			

Note: Factor scores are based on responses from when teachers were in the first year of their school's Partnership reform.
*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Figure A-1. Comparison of Selected Survey Responses From Teachers in Partnership and Non-Partnership Schools

Notes: For principal effectiveness, teachers were asked to rate their principal's effectiveness along several dimensions.

For accountability for failure, teachers were asked the extent to which they believed their school would face staff/leader removal if their improvement goals are not met.

For teachers' grade of their school, teachers were asked to assign grades to their school along several dimensions.

Online Appendix B. Identification Mechanisms for Each Round of Partnership Schools

Round 1 Partnership schools, which were identified in spring 2017 and began Partnership implementation in the 2017-18 school year, were schools that had been identified as Priority schools, meaning they were in the bottom 5% of schools on Michigan's Top-to-Bottom index, for three consecutive years. Michigan's school accountability lists, such as the Priority list or CSI list, were typically released in fall/winter using performance data from the prior school year. For instance, the 2016 Priority list that was used to identify the first cohort of Partnership schools was developed using data from the 2015-16 school year and released in the fall of the 2016-17 school year. Michigan's Top-to-Bottom list, which was used to annually identify low-performing schools through the 2016-17 school year, ranked schools on a metric that incorporated their proficiency rate in that year, the average proficiency rate from the present and prior year, change in average student performance over time, the gap between the highest- and lowest-performing 30% of the student body, and graduation rate (if the school was a high school).

Round 2 Partnership schools were identified in fall 2017 and began implementation in 2018-19. MDE selected schools for Round 2 if they were low performing in 2015-16 and experienced continued low achievement in 2016-17. Originally, schools could have been considered for Partnership identification in Round 2 if they were: a Priority school in 2016-17 and were not identified in Round 1, in a Round 1 Partnership district, or in a district that operated a school on the 2016-17 Priority school list. From there, schools could be identified for Partnership if they experienced a decrease in ELA or math proficiency, or if proficiency in ELA or math was less than 10%. In practice, only 2016-17 Priority schools were identified for Partnership in Round 2 and the proficiency criteria do not account for why some schools were selected for Partnership and others not. In conversations with officials at MDE, we learned that

"executive decision-making" was used when Round 2 Partnership schools were selected in fall 2017. This research began in the fall of 2018, by which time the officials who identified Round 2 schools were no longer at MDE to clarify how Round 2 schools were selected. We therefore use the same comparison group for Round 2 and 3 Partnership schools because they were selected in the same year and were on the same implementation timeline.

Round 3 schools were identified in the spring of 2018 and also started implementation in 2018-19. These schools were the bottom 5% of schools on the state's ESSA index system that was first released for the 2017-18 school year. If a school was identified in the bottom 5% of schools in spring 2018 but had already been identified for Partnership in an earlier round, the school proceeded on the initial implementation timeline for its Partnership Agreement. Michigan's School Index, which was first used in the 2017-18 school year to identify CSI schools, the lowest-performing 5% of schools, ranked schools on a metric that incorporated the school's proficiency rate, student growth, graduation rate (if the school was a high school), a school quality measure (which included the share of students not chronically absent for all schools; and advanced course participation and postsecondary enrollment for high schools; and for elementary schools access to the arts, physical education, and a librarian or media specialist), English Language Learner proficiency, and participation in statewide testing. While both the Top-to-Bottom index, which was used to identify Round 1 Partnership schools, and the ESSA index, which was used to identify Round 3 Partnership schools, use student proficiency as a main component and include graduation rates for high schools, they consider different measures of growth and the Top-to-Bottom system considered the gap between the top and bottom 30% of students while the Performance Index that began in 2017-18 includes the above described school quality measure, a focus on English Language Learners, and considers assessment participation.

Online Appendix C. Detailed Descriptions of Outcome Measures

Our primary outcome of interest is English Language Arts (ELA) and math achievement on the suite of standardized assessments administered to all third- through eight- and eleventh-grade students in Michigan each year. Several changes in the assessments administered to Michigan's students occurred during the time period we examine. In the 2014-2015 school year, Michigan replaced the Michigan Educational Assessment Program (MEAP) assessment with the Michigan Student Test of Educational Progress (M-STEP) assessment for grades 3-8. Starting in the 2018-2019 school year, Michigan uses the PSAT-8 to assess student learning in ELA and math for 8th grade students. We standardize these test scores by grade, year, and subject. In high school, students are tested only in 11^{th} grade. Through the 2014-2015 school year, all 11th graders in Michigan took the ACT, but starting with the 2015-2016 school year Michigan administered the SAT for 11th graders. Because students take the SAT only once in grade 11 for accountability purposes, we standardize SAT scores by subject and year and collapse them to the school-by-year level.

We also estimate the effects of Partnership on school-level high school four-year graduation and dropout rates. High school graduation and dropout are binary indicators based on the status of a student at the end of their expected graduation year, when a student is considered to be an on-time graduate, "other completer" (e.g., GED), dropout, or continuing on in high school. Because we only observe these outcomes once for each student, we collapse these binary variables to school-by-year means representing the proportion of a given cohort of students who dropped out and graduated on time, respectively.