



# Working Paper: School Improvement in Ohio and Missouri: An Evaluation of the School Turnaround Specialist Program

*Daniel Player<sup>1</sup>, Veronica Katz<sup>1</sup>*

---

School turnaround has received much attention and funding, but little empirical evidence exists as to how and whether it works. This study follows 34 elementary and middle schools in Missouri and Ohio throughout the course of their participation in the University of Virginia School Turnaround Specialist Program (STSP). Our evaluation of the STSP employs a triple differences model, comparing participating schools to non-participating schools within the same district as well as comparable schools in different districts. On average, participating schools experienced statistically significant improvements in student achievement after completing the two-year STSP. Improved student achievement is not concentrated within specific performance categories, suggesting that participation in the STSP is associated with increases in overall student performance rather than focusing on students at the margin of proficiency.

---

<sup>1</sup>University of Virginia  
Curry School of Education  
405 Emmet St. South  
Charlottesville, VA 22904

Updated May 2013.

Center on Education Policy and Workforce Competitiveness  
University of Virginia  
PO Box 400879  
Charlottesville, VA 22904

CEPWC working papers are available for comment and discussion only. They have not been peer-reviewed.  
Do not cite or quote without author permission.

We are grateful to Jim Wyckoff, Tom Dee, and numerous seminar participants at UVA and at the AEFPP conference for useful comments. We are also grateful to LeAnn Buntrock and William Robinson of the Partnership for Leaders in Education for helpful information about the implementation of the school turnaround program.

**SCHOOL IMPROVEMENT IN OHIO AND MISSOURI: AN EVALUATION OF THE SCHOOL  
TURNAROUND SPECIALIST PROGRAM**

*By Daniel Player & Veronica Katz, University of Virginia*

**Introduction**

Student achievement lags in many schools throughout the United States. In the decade since the enactment of the No Child Left Behind Act, the increased emphasis on student achievement has highlighted the shortcomings of many persistently low-performing schools. Federal investments of approximately \$5 billion to turn around 5,000 low-performing schools further underscores the magnitude of the need, at least from the federal government’s perspective (“No child left,” 2012).

In response to this need for turnaround assistance, some states and districts have sought external partners to assist with turnaround efforts in schools. One of these partners, the University of Virginia School Turnaround Specialist Program (STSP) has worked with more than 250 schools in more than 65 urban and rural districts over the past nine years. To date, there has not been a large-scale longitudinal evaluation of the schools that have participated in the STSP.

This paper presents an evaluation of a single cohort of 34 STSP schools from Ohio and Missouri. The findings from the evaluation suggest that the schools experienced significant gains in student achievement over the two years in which they participated in the STSP. The patterns of the gains suggest that they were not limited to students right on the threshold of proficiency, yielding some evidence that the changes were not simply a “bubble-student” phenomenon.

**Background**

*School Turnaround*

Although there is not a universally agreed upon definition of school turnaround, it is distinguished from school improvement and comprehensive school reform (CSR) as a process of

change focused on quickly improving schools that persistently fail (Hassel, Hassel, Arkin, Kowal, & Steiner, 2010; Hassel, Hassel, & Rhim, 2007). Whereas CSR and school improvement models are characterized by slow, incremental changes over time, school turnaround aims to dramatically alter the trajectory of a school's student achievement in less than three years (Herman et al., 2008).

Early school turnaround studies consist largely of case studies focusing on schools that successfully changed their achievement trajectory. One case study examined successful, high-achieving schools in urban, poverty-stricken areas and found that the schools where the district office played a substantial role experienced the most rapid gains (Johnson et al, 1999). Another case study examined 15 historically low-performing schools that had sustained growth for at least 2 years and found that leadership changes were a common feature in many of the schools (Duke, n.d.). Other case studies focus heavily on leadership, quick gains, district support, and data-driven restructuring models (Public Impact, 2007).

Although turnaround programs can offer a variety of policy and curriculum changes, one emergent theme in turnaround research is district and school leadership. Prior research suggests that turnaround leaders take action-based approaches to leadership (Kowal & Hassel, 2005). Some literature has looked for evidence of commonalities in turnaround leaders across public and private industries, including concentrating on a few changes with big, fast payoffs and implementing practices proven to work with previous low-performing students without seeking permission for deviations from district policies (Kowal & Hassel, 2005). The qualitative research on turnaround schools and businesses supports the idea that districts and school leadership both play a key role in turnaround success.

The What Works Clearinghouse makes four recommendations for school turnaround based on analysis of ten case studies but cautions repeatedly that the level of evidence is weak (Herman et

al., 2008, page 6). According to WWC’s comprehensive review of the extant literature, successful school turnarounds are enabled by strong leadership, data-driven instructional focus, quick gains, and a committed staff (Herman et al., 2008). These four recommendations are thought to be conjunctive, such that effective school turnaround requires the implementation of all four recommended practices.

Following the 2008 WWC review, few large-scale empirical studies have documented the effect of whole-school reforms in turning around low-performing schools. A notable exception is Dee’s (2012) recent evaluation of School Improvement Grants (SIGs) in California. Dee employs a fuzzy regression discontinuity model, leveraging the eligibility requirements used to identify schools given prioritized access to SIG interventions. Specifically, California’s schools were deemed SIG-eligible if they qualified for Title I funding, were among the “lowest achieving”<sup>1</sup> schools, and demonstrated “lack of progress”<sup>2</sup> (Dee, 2012, page 6). Dee evaluates post-intervention performance at each treatment threshold and finds significant test-based improvement among schools identified as “lowest-achieving” but does not find similar results at the “lack of progress” threshold (Dee, 2012, page 22). The author also uses a difference-in-differences estimation strategy to evaluate the effect of each SIG reform model<sup>3</sup> on school performance and finds achievement growth concentrated in

---

<sup>1</sup> “Lowest achieving” schools are defined for each school level (i.e. elementary, middle, and high school) using the following AYP proficiency-rate thresholds: “29.97 percent or below for elementary schools, 22.44 percent or below for middle schools, and 37.31 percent or below for high schools” (Dee, 2012, page 5).

<sup>2</sup> “The California Department of Education based its ‘lack of progress’ definition on its school-level, test-based Academic Performance Index (API) [...]. Specifically, for each of the 3,652 PLA-eligible schools, the state summed the annual API growth from five baseline years (i.e., AY 2004-05 through AY 2008-09). Schools for whom this summed growth measure was below 50 (or missing) were labeled as ‘lack of progress’ schools” (Dee, 2012, page 5).

<sup>3</sup> SIG schools must select one of four reform models: closure, restart, transformation, or turnaround. With the restart model, schools reopen under the supervision of a charter or education management organization. The transformation model requires replacing the school principal in addition to multiple instructional and personnel reforms. Schools adopting the turnaround model face the most

schools adopting the “turnaround” model, an aggressive restructuring model that requires replacing the principal and at least half of the school’s prior personnel (Dee, 2012, page 27). Taken together, these results suggest whole-school reform can alter student performance when the reform is robust and targeted at the lowest-performing schools.

As described in the following section, the STSP is a whole-school reform program. However, unlike the federally prescribed turnaround model, the STSP does not require a fixed proportion of staff turnover. Rather, the STSP focuses on developing leadership capacity at the school and district level. To this end, this study contributes to a growing body of research evaluating the impact of effective leadership on student outcomes. To date, the extant research on school leadership has focused primarily on school principals, finding a positive albeit indirect relationship between principal effectiveness and student outcomes (Brewer, 1993; Hallinger, Bickman, & Davis, 1996; Hallinger & Heck, 1998). Observational studies have found improved student outcomes under the leadership of principals who dedicate a significant portion of their time to teacher recruitment and management (Beteille, Kalogrides, & Loeb, 2009; Grissom & Loeb, 2011; Horng, Klasik, & Loeb, 2010). Similarly, survey data suggests that teachers are more likely to remain in their school if they hold a favorable impression of their school administrators (Boyd et al., 2011; Grissom, 2011; Ladd, 2011) and improved teacher retention ostensibly contributes to improved student outcomes. In sum, effective school leadership has an influential impact on student outcomes. This study contributes to the existing literature by assessing the effect of improved school and district leadership on student outcomes.

### *STSP Program*

---

aggressive reforms. In addition to the changes mandated under transformation, the turnaround model also requires that schools replace at least 50% of the staff.

The first cohort of schools to participate in the STSP began in 2003. Each year since, the STSP has introduced a new cohort of schools. As of summer 2012, the STSP had worked with more than 250 schools in more than 65 districts across the country. Beyond school-level progress summaries in annual reports produced by the Partnership for Leaders in Education (PLE),<sup>4</sup> we are unaware of any published research that has attempted to document a causal impact of STSP.

One unique feature of the STSP among turnaround programs is the emphasis on building district-level support for the turnaround schools. The primary assumption behind the STSP is that district involvement and ownership of the turnaround effort is the key to its sustained success. This assumption aligns with what little research there is about the district's role in school support (Baroody 2011; Honig, 2003; Honig, Copland, Rainey, Lorton, & Newton, 2010; Waters & Marzano, 2006). The program institutes a number of initiatives to improve alignment between the district's and the school's goals, such as requiring districts to identify an executive-level District Shepherd to personally visit each turnaround school weekly and to act as a liaison between STSP schools and the district superintendent. District Shepherds are also tasked with ensuring STSP schools have access to organized and responsive data systems that will enable STSP schools to use student data (including attendance, formative assessments, and summative assessments) to drive student interventions. The program also expects districts to formulate specific turnaround plans and to follow up with schools on the plans they make.

In addition to developing district-level leadership skills, the STSP model also acknowledges that school leadership makes a difference in student achievement (Hallinger & Heck, 1996; Waters,

---

<sup>4</sup> Faculty and staff from the University of Virginia's (UVA) Curry School of Education and the Darden School of Business have collaborated since 2003 in the Partnership for Leaders in Education (PLE), a joint venture that seeks to develop the capacity of leaders in education. The STSP is one product of this collaborative effort.

Marzano, & McNulty, 2003; Witziers, Bosker, & Krüger, 2003). The STSP emphasizes the role of principals in setting expectations for instruction and providing teachers with feedback on their instructional practices. Principals are taught to share specific expectations for instruction and to regularly observe classroom instruction and provide detailed feedback. The program also emphasizes the importance of the principal setting and communicating clear goals for the school, recognizing key issues and developing solutions based on objective data, and pursuing bold actions.

### *Participation*

The STSP has evolved over the course of the nearly nine years that it has operated. The cohort included in this evaluation participated in a two-year program that included in-residence (at UVA) and in-field support for district personnel and school leaders. Sequencing is described in greater detail below.

**Summer 1.** In the summer prior to the first school year of the program, district leaders, principals, and teachers from turnaround schools attended a six-day executive education session at the Darden School of Business at the University of Virginia. During this time, participants focused on developing 90-day action plans for their schools and districts, specifically detailing how they plan to bring about change in their schools during the first half of the school year. They also received instruction on how to effectively engage and motivate a high performing team, and how to use student data to monitor student progress and diagnose the root causes of problems.

**Fall 1.** During the fall, each school was visited by a Darden/Curry staff member. During the visit, the staff member assessed the school's progress toward meeting their turnaround plan and provided feedback on the school's availability and use of student data for instructional decisions.

Throughout the two years of the program, STSP schools expected to be visited weekly by a district leader who assessed school progress and provide mentoring and support to principals.

**Winter 1.** School principals and school leadership teams (three to four people) met in the winter of the first year for two-and-a-half days of executive education from UVA faculty and district leaders. The instruction was targeted to needs that were identified through site visits and communication between the districts and the STSP staff. It was also focused on developing school leadership teams to build momentum. It culminated in the development of a second 90-day plan to build on each school’s progress during the second half of the academic year.

**Year 2.** The second year of the program was in many ways a repeat of the first year. School and district leaders received three days of executive education at the Darden School of Business in the summer of the second year during which they received instruction on building school culture and reinforcement on using student data to drive instructional practices. At the conclusion of the week they developed a 90-day plan for the first half of the second school year. Schools were each visited a second time by UVA staff in the fall of the second year, and school leadership teams gathered in the winter for a two-and-a-half day executive education session to develop their 90-day plan for the second half of the school year. Schools and districts also received “real-time support” throughout both years of the program, but particularly in the second year. This support consisted of multiple site visits by STSP staff and Darden and Curry faculty members to provide customized support to address specific needs.

## **Method and Data**

Schools were not randomly assigned to receive STSP. With this limitation, the estimation strategy must take account for other contemporaneous changes that could have been occurring

during the same period. Most notably, there could have been district-level policies or interventions that could have positively affected school performance. The evaluation must account for these. Likewise, low-performing schools could have been the target of additional state support. Therefore, we implement an estimation strategy that accounts for district-level influences as well as the potential state-level influences. This method is described below.

The impacts of STSP are calculated using a triple differences methodology in which changes in outcomes over time are contrasted between STSP schools (first difference) and non-participating schools in the same district (second difference). Those differences in changes are then contrasted with similar measures from comparable schools and districts in the state that did not participate in STSP (third difference).

The following model was used to estimate the impacts of STSP:

$$Y_{jkt} = \alpha + X_{jkt} \beta + d_1 P_t + d_2 T_j + d_3 S_{kt} + d_4 (T_j S_{kt}) + D_t \gamma + F_j + e_{jkt}$$

where  $j$  indexes schools,  $k$  indexes districts,  $t$  indexes time,  $X$  is a vector of school characteristics (percent Black, percent Hispanic, percent free/reduced lunch eligible, and student enrollment),  $D$  is a vector of year dummies to account for year-specific effects,  $P$  is an indicator for the post period years,  $T$  is an indicator equal to one for treatment-eligible schools in the post-intervention period, and  $S$  is an indicator for STSP districts in the post period.<sup>5</sup> The inclusion of school fixed effects ( $F$ ) introduces the first differences of the model and accounts for any time-invariant omitted variables. This fixed effect also implicitly accounts for any differences between the two states. Changes associated with participation in the STSP are captured with an interaction of the treatment-

---

<sup>5</sup> For simplicity, I have defined many of the indicators as being post-period specific. This is equivalent to defining these as interactions with a post indicator. For example, the indicator  $T$  in the model above is equivalent to interacting a time-invariant treatment indicator with  $P$ .

eligible indicator and an indicator for being in an STSP district. By construction, this indicator is equal to one only in the post period for schools that participated in STSP. Thus, the coefficient  $d_4$  represents the difference corresponding with STSP participation net of any other trends occurring in the STSP districts and the state during the same period. The regressions used stacked school panel data (2006-07 through 2010-11) and post-intervention years were defined as 2009-10 and 2010-11. Because the model includes multiple observations per school, all standard errors are corrected to account for the clustering across schools over time.

A key to this method is to identify a set of comparison districts and “treatment-eligible” schools on which to identify the third difference. These schools serve as an important comparison because they would be facing the same accountability pressure and motivation to improve school performance as the STSP schools faced prior to their participation. We used propensity score matching models to first identify other schools in the state that are comparable to the STSP-participating schools.<sup>6</sup> We estimated the following propensity model separately for each state:

$$\Pr(STSP)_j = a + P_j b_1 + X_j b_2 + dG_j + e_j$$

where  $P$  is a quadratic of baseline performance levels,  $G$  is a measure of student achievement growth during the baseline period, and  $X$  is defined as above. We used nearest neighbor matching (with replacement) to identify the five closest matches for each treatment school and designated them as “treatment-eligible.” Those schools and their corresponding districts were then included in the main impact estimation described above.

---

<sup>6</sup> We also explored identifying “treatment-eligible” by first identifying comparison districts and selecting the lowest 40% of schools in those districts as treatment-eligible. The results presented in this paper are comparable and slightly more conservative.

The triple difference methodology allows us to rule out some other explanations for changes in school performance that were unrelated to STSP such as regression to the mean or other general responses to accountability pressures. All of the comparison schools from non-treated districts would have been facing the same possibility of mean regression since they were beginning from a low achievement level. The comparison schools would have also had the same accountability pressures as the schools that received the STSP, ruling out the possibility that the school was simply responding to pressure. Likewise, the longitudinal nature of the analysis allows us to rule out the possibility that the improvements were happening prior to the introduction of STSP.

*Sample*

The schools that participated in STSP and are part of this evaluation were from Ohio and Missouri. The participating Ohio schools were all elementary or K-8 schools from two urban districts. The districts each approached the STSP and identified the schools that were in need of turnaround. One district identified the 15 lowest-performing schools in the district to participate in the program. A second urban district identified four persistently low-performing schools that were, in their estimation, in greatest need of turnaround. These were among the schools with the lowest performance ratings, but not necessarily the four lowest-performing, in the district.

Nineteen elementary and middle schools from Missouri also completed the STSP program during the same period. STSP schools in Missouri spanned grades K-8. Unlike Ohio, the participating schools in Missouri were nominated at the state level rather than the district level. The state identified underperforming schools to be roughly regionally representative within the state. This resulted in a sample of schools that was urban and rural and represented a much larger group of districts than the schools in Ohio. Missouri's participating schools were grouped in 10 districts, 2 of which were large urban districts and the remaining 8 were small rural districts. For the purposes

of the evaluation, we restrict the treatment sample to elementary and middle schools serving grades K-8 that have at least one non-STSP school of the same level in the district (to permit within-district comparisons). This restriction results in a final sample of 15 STSP schools in 8 districts in Missouri.<sup>7</sup>

8

### *Data*

All data are at the school level and are collected from the state departments of education and merged with demographic data from the Common Core of Data. All outcomes and demographic data were collected for the school years 2005-06 through 2008-09 (pre-period) and 2009-10 through 2010-11 (post-period) from Ohio. The comparable Missouri data were available for one year fewer in the pre-period. For the main analysis, we include the years available for both states (2006-07 through 2010-2011). However, in some state specific analyses we include the full available data.

This study relies on three primary outcome measures that reflect student performance on state exams and the average daily student attendance. The primary outcome of interest is the school's performance rating. Each state has unique proficiency standards. Ohio calculates a state performance index for each school, which is a weighted average of the percentage of students scoring in each proficiency category (Untested, Limited, Basic, Proficient, Accelerated, and Advanced). Students in the Advanced category receive the highest weight and each descending category receives slightly less weight (Ohio Department of Education, 2011, page 3). A school can receive a maximum performance index score of 120. Missouri also calculates a performance index called the Missouri Assessment Program (MAP) index. This is similar to the weighted average

---

<sup>7</sup> The omitted schools were from two small rural districts with only one school per level.

<sup>8</sup> One Missouri school was closed at the end of the first year of the intervention. We retain this school in the sample although it contributes only one year of impacts. The estimates are largely unchanged if we omit this school.

calculated by Ohio in which proficiency categories (Below Basic, Basic, Proficient, and Advanced) are given increasing weight (Missouri Department of Elementary and Secondary Education, 2012, pp. 8-9). Schools can receive a maximum MAP index score of 500.<sup>9</sup> To create comparable data across the two states, we standardized each state's performance index by year to have mean zero and standard deviation one. In addition to the performance score, we also use raw proficiency rates as outcomes. For comparability across states, these are grouped into four categories: Below Basic, Basic, just Proficient, and above Proficient. In grade-level analysis, we are only able to classify students as proficient or not.

There are several potential limitations associated with the use of school-level achievement data, especially when comparing across states. A primary concern is the comparability of achievement outcomes across states, as each state likely differs with respect to exam content and rigor, test administration practices, and the accuracy with which data is collected and reported, among other issues. Additionally, achievement tests are an admittedly crude measure of student performance. Nonetheless, Ohio and Missouri both utilize their state exam as an important indicator of overall school performance. In keeping with the fact that two states included in our analysis identified STSP schools on the basis of their state achievement test performance, we employ the same metric to estimate the change in outcomes associated with participation in the STSP. In other words, our outcome measures are consistent with the measures valued by each state. We address comparability across states by standardizing achievement outcomes by grade and year within each state, such that standardized achievement outcomes for each state are relative to the within-state average achievement outcome. Within-state standardization also accounts for any noise associated with state achievement exams, as all schools within a given state are subjected to the same exam and

---

<sup>9</sup> The maximum score would occur only if all students attained advanced proficiency

thus equally affected by the noisiness of the measure. If anything, test measurement error would tend to result in underestimation of the change in outcomes corresponding with participation in the STSP.

### *Descriptive Statistics*

Overall, the 19 participating schools from Ohio were all among the lowest-performing schools in the state. All fell within the bottom 10 percent according to the state performance index (Figure 1) and 14 of the 19 schools were within the bottom 5 percent. The Missouri schools were also low-performing, but they were not as low relative to state averages as were the schools in Ohio. Nine of the 15 participating schools fell within the bottom 10 percent (Figure 1). All but one school fell within the bottom 30 percent of elementary and middle schools statewide, and the remaining school was in the bottom half.

Prior to the intervention (2008-09 school year), schools in the treatment sample were well below the state average in performance for both states (Table 1).<sup>10</sup> In Ohio, the schools were well below the other schools in their districts on a number of metrics including overall performance, the percentage of students scoring Proficient or above, and percentage of students Below Basic. The within-district contrast was not as pronounced in Missouri. In fact, while the STSP schools in Missouri were below the state average on the performance index, they were not statistically below the other schools in their districts. The only metric in which STSP schools differed from the non-STSP schools in their district was in the percentage of students eligible for free/reduced lunch.

---

<sup>10</sup> For Ohio, the state comparisons are made against all non-STSP elementary schools in the state that had performance data for 2009 (N=1716). In Missouri, comparisons are made against all non-STSP elementary and middle schools with performance data in 2009 (N=1717).

In order to calculate triple difference estimates, we had to first identify comparable schools in the state that could proxy for STSP schools in the absence of the treatment. The resulting matches revealed important differences between Missouri and Ohio. In Ohio, the STSP schools were so low-performing and were so similar to each other that they only matched to 29 unique schools.<sup>11</sup> These comparison schools were closer to the STSP schools in performance and demographics than were the non-STSP schools in the STSP districts, but they were still statistically different on a number of performance metrics (Table 1, column 3). Indeed, the Ohio STSP schools were nearly without peers in terms of their performance and demographics.<sup>12</sup> In contrast, the Missouri schools matched with a total of 66 unique schools. These comparison schools were quite similar, on average, to the STSP schools. With the exception of the percentage of students eligible for free/reduced lunch, there were no statistically significant differences in terms of performance or demographics between the STSP schools and their non-district comparisons.

## Results

The trends over time for STSP schools and their comparisons suggest an impact of the program post 2009 (Figure 2). Prior to the program, the STSP schools were on a similar trajectory as the other schools in their districts and similar schools in the state. After 2009, the STSP schools begin to make visible improvements in their performance. The comparison schools do not make a similar improvement over this same period. All three groups of comparison schools (non-STSP in the same districts, treatment-eligible schools outside the districts, and non-treatment-eligible schools outside the districts) show parallel trends post 2009 that are relatively flat. By the conclusion of the

---

<sup>11</sup> A key constraint was that the comparisons had to come from another district. Many of the lowest-performing non-STSP schools in Ohio were in the two urban districts that participated in STSP.

<sup>12</sup> As an alternative in Ohio, we did the same analysis using the bottom 40% of schools in all urban districts as the comparisons. The results were very similar to what is presented below.

two-year intervention, the STSP schools have completely closed the gap with the other treatment-eligible schools in the state and are outperforming them on the state performance index.

State specific trends further decompose these suggestive findings. The STSP schools in Ohio began to increase at a time when their comparison schools began a slight decline (Figure 3). By the second year of the intervention, the Ohio schools had surpassed the non-district treatment-eligible schools by a significant margin. In Missouri, the STSP schools also closed the gap although the difference was not as dramatic.

### *Triple Difference Estimates*

The essence of the triple difference estimation is to formalize and quantify the closing of the gap that is illustrated in Figures 2 and 3 and to determine whether the differences were in fact practically and statistically significant. Participants in STSP saw a statistically significant increase of .36 standard deviations in the overall state performance score (Table 2). These schools saw an increase of 2.26 percentage points in students scoring above Proficient on their state tests and a reduction of 4 percentage points of students scoring Below Basic, both statistically significant differences. This suggests that the observed improvement was not concentrated among the “bubble students” scoring just below Proficient at the Basic level. In fact, Basic is the only category that did not see a significant change. Likely the students moving out of Basic and into Proficient were replaced by a similar number of students moving up out of the Below Basic category.<sup>13</sup> There was no evidence that schools saw a statistically significant increase in student attendance rates.

Changes in subject-specific proficiency rates tell a qualitatively similar story as our main estimates (Table 3). With the exception of third grade math, each grade and subject saw positive and

---

<sup>13</sup> Schools could manipulate scores by reducing the number of tested students. However, there was also no statistically significant impact on the number of untested students (not shown).

statistically significant gains in the percentage of students scoring Proficient or higher. There is no strong evidence that students improved relatively more in math or reading.

As described above, STSP schools in Ohio were quite different from STSP schools in Missouri in terms of their initial achievement levels and in the ways in which they were selected to the program. Thus, there is reason to believe that the observed improvements could vary by state. The state-level results were in a consistent direction, but differed in terms of size and statistical significance (Table 4). Ohio schools saw statistically significant gains of around .45 standard deviations on the state performance index while Missouri schools saw a much smaller gain of .19 standard deviations that was not statistically significant. The pattern across proficiency rates tells a similar story; the point estimates in Missouri are consistently smaller than those in Ohio and not statistically distinguishable from zero. Ohio also saw a statistically significant gain in average daily attendance rates of nearly one percentage point.

As expected, the school-level point estimates varied considerably across the sample (Figure 4). The estimated differences were positive for 28 of the 34 total schools and negative for 6 schools. It is interesting to note that the two schools with the smallest (most negative) point estimates were from Ohio and the two schools with the largest point estimates were in Missouri. Given the overall state means, these results are somewhat surprising. However, the pattern of school-level point estimates suggests that the differences in Missouri tended to be small, and the two large positive outcomes were not typical of the Missouri STSP schools. In contrast, the Ohio schools tended to witness larger positive changes and the two negative point estimates were not typical of the Ohio schools. Overall, there was no statistically significant correlation between the school's baseline achievement level and the size of the observed improvement. In other words, the lowest-performing STSP schools did not appear to experience greater gains relative to the higher-performing STSP

schools. However, it is worth noting that all but one of the schools with negative point estimates were in the top half of the sample in terms of their baseline achievement (Figure 4, right panel).

### *Robustness Checks*

We performed a number of robustness checks to explore the possibilities that the observed differences arose either by chance or by some other factor unrelated to STSP.

### **Examination of SIG schools**

During the period of this intervention, the federal government provided an unprecedented level of investment in persistently low-performing schools. Twelve elementary schools in Ohio received Tier I or Tier II school improvement grant (SIG) funding, including six STSP schools. Likewise, 23 elementary and middle schools in Missouri received awards, although none were program participants. The STSP schools that received SIG funding had considerable extra resources to devote to activities beyond their participation in STSP. If the resources alone had an impact above and beyond the STSP then the effects of STSP could be confounded with the other benefits of the SIG funding. To explore the possible effects of confounding resources with the STSP program, we restricted the analysis to Ohio schools that had not received SIG funding. The point estimates were reduced slightly from the main Ohio results (.39 versus .45), but were still highly statistically significant. We also ran an exploratory difference-in-differences regression with just the Ohio schools that received SIG funding (including schools at all levels) and the estimated STSP effect size among those schools was .53 and was statistically significant at the  $p=.01$  level. Therefore, the STSP schools that received SIG funding experienced larger growth than non-STSP schools in the post period. Taken together, these results suggest that STSP had an effect independent of additional funding that came with SIG grants.

### **Falsification Tests**

If the STSP had an impact, we would not expect to see any reported “impacts” of the program at times other than in the post period. Although the graphical representation suggests that the observed differences were seen only after 2009 (Figure 2), we formalized the test by running the triple difference regressions and defining the post period first as 2009 and then as 2008, one and two years before the program was implemented. The estimated impacts were 0.00 and -0.02 respectively and in neither case were they statistically significant.

As an alternative falsification test we also estimated difference-in-difference “impacts” on a sample that included only the non-STSP “treatment-eligible” schools. The only statistically significant finding was an increase of 1.5 percentage points in the percent of students scoring Proficient. This suggests there may have been a contemporaneous change in the percentage of students scoring Proficient among low-performing schools that corresponded with the implementation of STSP. The main triple difference estimations account for these changes however through the inclusion of year-specific fixed effects.

### **Other Factors**

One possible explanation for changes in school performance could be that the student demographics changed considerably as a result of the STSP. For instance, it could be that motivated parents began to enroll their children in the STSP schools when they learned the school was going through a turnaround effort. If that were the case, changes in proficiency could have little to do with student learning and could instead reflect changes in student demographics. Unfortunately, it is very difficult to completely rule out this possibility. As a check, we estimated the “impacts” of STSP on available demographics. There were no statistically significant changes in the percentage of Hispanic

students, students eligible for free/reduced lunch, or total enrollment that corresponded with the implementation of STSP (Table 5). There was a reduction in the percentage of Black students that was marginally significant ( $p=0.07$ ). However, the change was relatively small in magnitude and corresponded to a less than 2 percent change.

## Discussion

The measures of success we include in this study might not capture the full extent of changes that have occurred in the schools, but they are the measures that the state and federal government use to gauge school success. Ideally, the analysis would include measures of student discipline, school climate, and teaching practices. In qualitative interviews with the principals in five of the schools in Ohio, principals and teachers pointed to the increased use of student achievement data to plan instruction as the biggest factor that contributed to the success of the school. Principals we interviewed also discussed a shift in teacher attitudes toward greater self-efficacy as they began to see the small improvements that students were making. Most reported a significant improvement in student discipline and teacher collaboration. To the extent that these things translate to unobservable effects on students (e.g., promotion of citizenship), the measures here may understate the true changes that have taken place in the schools.

### *Have these schools turned around?*

Unfortunately, there is no single definition for a school that has turned around. The schools in this sample saw statistically significant improvements relative to similar schools, but can they be classified as having been turned around? By one definition, a turnaround could be thought of in absolute terms: a school has turned around only if it has reached some predetermined level of proficiency. For illustrative purposes, we could take the state average as the level a school must meet

in order to be classified as a successful turnaround. As of 2011, the STSP schools are still approximately 1.5 standard deviations below the state average in standardized performance scores. Proficiency rates in math and reading were statistically below the state average in every grade, with a near 40 percentage-point difference in fifth grade math proficiency. Despite the progress they have made, the schools have not radically turned around if we use the state mean as the benchmark.

If we think of turnaround in terms of a significant change in direction, the evidence above suggests that the schools in our sample have improved their condition. To illustrate, the schools in STSP were 18 percent more likely than non-STSP schools to be in the top 10 percent of growth for the state, and STSP schools in Ohio were nearly 50 percent more likely than non-STSP schools in Ohio to be in the top 10 percent.<sup>14</sup> STSP schools have also closed the performance gap considerably. For instance, the Ohio schools closed nearly 2/3 of the gap with non-STSP schools between 2009 and 2011. This suggests that the schools are making meaningful progress, but they still have much ground to make up.

In terms of cost, STSP costs a fraction of the full amount of the average SIG grant. Across the two states in our study, the average SIG grant award was more than \$650,000 per recipient in 2010-11 compared to less than \$40,000 for the same year of STSP. While the point estimates observed in this study are similar in size to the overall SIG impacts measured by Dee (2012), they come at small fraction of the cost.<sup>15</sup>

*What can we learn about turnaround implementation?*

---

<sup>14</sup> Estimates based on a logistic regression that includes a quadratic of baseline performance, student demographics, and total enrollment. Full results available on request.

<sup>15</sup> As noted above, some of the STSP schools received SIG grants. However, the non-SIG STSP schools saw nearly the same impacts on average as the SIG STSP schools.

Unfortunately, the study lacks the richness of implementation data to be able to comment very much on the best way to implement school turnaround. However, the results consistently pointed to larger effects in Ohio than Missouri. Anecdotally, STSP staff that worked with the schools from both states reported that the district support for Ohio schools was critical to the success of those schools. In Missouri, the schools had been selected by the state and the districts did not have as much ownership in the turnaround effort. Consequently, the STSP schools in Missouri were less likely to receive support and follow-up from the central office.

*Limitations and Extensions*

The treatment and comparison schools in this study were not chosen randomly. Therefore, there is the possibility that the results were influenced by selection bias. In other words, the districts and schools that participated in this program may have been different from the comparison schools in unobservable ways that influenced their ability to respond to the STSP. Because the districts (Ohio) and state (Missouri) identified the schools, the direction of this potential bias is unclear. For example, if district/state leaders were identifying the schools that were the least likely to be able to turn around without outside assistance, the estimates presented here could understate the true differences associated with participation in the STSP. On the other hand, if district/state leaders were identifying the schools that were the most likely to be successful in the program, the estimates may overstate the changes wrought by the STSP.

While we cannot completely rule out the possibility of selection bias, it is helpful to provide background on how the schools were selected. In both states, federal and state accountability pressure was mounting for a number of the low-performing schools. In Ohio, two districts approached UVA to partner in a turnaround effort. One of the districts approached UVA because a key person in the district had moved from the Charlottesville area and was familiar with the program.

The second district found out about UVA through web searches. Throughout the rest of the state, other districts sought external turnaround partners. In other words, these were not the only two districts in the state that were motivated to identify external turnaround partners. Fortunately, the triple difference methodology accounts for district-level characteristics that could favorably influence the participating schools while inter-district comparisons include schools that were in a similar academic situation but did not have a chance to be selected.

In Missouri the school and district selection was less clear. Key people in the state learned of the program through interactions with the program in other contexts. The state then identified schools based on geographic representativeness and the school's willingness to participate in the program. As with Ohio, other turnaround partners were utilized for other underperforming schools throughout the state, ruling out the possibility that these were the only districts/schools motivated to improve.

One other limitation of this analysis is that it does not formally seek to explain what was happening in the schools that participated in STSP. We know relatively little about how these schools were bringing about change. For instance, we do not have teacher-level data that would allow us to observe whether changes in teacher composition were a key factor in successful schools. Likewise, without student-level data we cannot formally rule out the possibility that some of the improvement was driven by changes in student composition that corresponded with the implementation of STSP.

With these data, we cannot yet observe schools beyond their second year of participation in the program. Therefore, we cannot yet observe whether these effects will persist, improve, or decline over time. Future research will follow these schools over a longer horizon to see whether the results appear to have been sustainable.

A final limitation of this study is that it includes schools from only two states that participated in STSP. Therefore, the external validity of these findings is limited. In future years we will expand the sample to include new districts as they participate in STSP.

## **Conclusion**

The schools examined as part of this study demonstrated statistically and practically significant growth in student achievement within two years of participating in STSP. On average, schools moved up more than 1/3 of a standard deviation in the state performance rating, which was borne out in greater percentages of students passing math and reading tests in all grades. Schools that participated in STSP also saw gains in the percentage of students scoring above proficient on the state exams and reductions in the percentage of students who scored below proficient. Schools in Ohio also saw a small but statistically significant increase in the average daily student attendance.

Changes associated with participation in the STSP were heterogeneous across schools. The observed differences were unrelated to baseline performance measures, but were clearly larger in Ohio than they were in Missouri. This suggests important differences in the implementation of STSP across the two states.

The results of this study yield evidence that STSP can facilitate meaningful change in persistently low-performing schools. Taken more generally, it suggests that relatively low-cost interventions can yield promising results among persistently low-performing schools. This is good news to policy makers in light of the federal and state investments for these schools. However, the study does not indicate that the typical STSP school is able to move among the top-performing schools within two years. Further research and follow-up will help illuminate whether STSP brings about enough momentum to sustain sufficient improvement to close the full achievement gap.



## References

- Baroody, K. (2011). Turning around the nation's lowest-performing schools: Five steps districts can take to improve their chances of success. *Center for American Progress*.
- Beteille, T., Kalogrides, D., & Loeb, S. (2009). *Effective schools: Managing the recruitment, development, and retention of high-quality teachers*. (CALDER Working Paper No. 37). Washington, DC: The Urban Institute.
- Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, 48(2), 303-333.
- Brewer, D. J. (1993). Principals and student outcomes: Evidence from US high schools. *Economics of Education Review*, 12(4), 281-292.
- Dee, T. (2012). *School turnarounds: Evidence from the 2009 stimulus*. (Working Paper No. 17990). Cambridge, MA: National Bureau of Economic Research.
- Duke, D. (n.d.). *Keys to sustaining successful school turnaround*. Unpublished manuscript. Charlottesville, VA: Darden/Curry Partnership for Leaders in Education. Retrieved August 2007, from [http://www.darden.edu/html/standard.aspx?menu\\_id=39&styleid=3&id=3215](http://www.darden.edu/html/standard.aspx?menu_id=39&styleid=3&id=3215)

- Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, 113(11), 2552-2585.
- Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, 48(5), 1091-1123.
- Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *The Elementary School Journal*, 96(5), 527-549.
- Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, 32(1), 5-44.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157-191.
- Hassel, B. C., Hassel, E. A., & Rhim, L. M. (2007). Overview of restructuring. In H. J. Walberg (Ed.), *Handbook on restructuring and substantial school improvement* (pp. 9-22). Lincoln, IL: Center on Innovation and Improvement.
- Hassel, E. A., Hassel, B. C., Arkin, M. D., Kowal, J. K., & Steiner, L. M. (2010). *School Restructuring: What Works When? A guide for education leaders*. Naperville, IL: Learning Point Associates.

- Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., & Redding, S. (2008). *Turning Around Chronically Low-Performing Schools: A practice guide*. (NCEE No. 2008-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.
- Honig, M. (2003). Building policy from practice: District central office administrators' roles and capacity for implementing collaborative education policy. *Educational Administration Quarterly*, 39(3), 292–338.
- Honig, M., Copland, M. A., Rainey, L., Lorton, J. A., & Newton, M. (2010). *Central office transformation for district-wide teaching and learning improvement*. Manuscript submitted for publication, Center for Study of Teaching and Policy, University of Washington, Seattle, WA.
- Hornig, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, 116(4), 491-523.
- Johnson, J.F., and Asera, R. (Eds.). (1999). *Hope for urban education: a study of nine high-performing, high-poverty, urban elementary schools*. Washington, DC: Policy Studies Associates and The University of Texas at Austin, The Charles A. Dana Center.
- Kowal, J.M., and Hassel, E.A. (2005). *Turnarounds with new leaders and staff*. Washington, DC: The Center for Comprehensive School Reform and Improvement.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33(2), 235-261.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. Prepared for The Wallace Foundation. Retrieved April 2007 from <http://www.wallacefoundation.org/WF/KnowledgeCenter/KnowledgeTopics/EducationLeadership/HowLeadershipInfluencesStudentLearning.htm>

Missouri Department of Elementary and Secondary Education. (2012). *MSIP 5: Guidance document for the fifth version of the Missouri School Improvement Program (MSIP 5)*. (Draft version).

*No child left behind funding*. (2012, April 24). Retrieved from <http://febp.newamerica.net/background-analysis/no-child-left-behind-funding>

Ohio Department of Education. (2011). *Guide to Understanding Ohio's Accountability System: 2010-11*.

Public Impact. (2007). *School Turnarounds: A Review of the Cross-Sector Evidence on Dramatic Organizational Improvement*. Lincoln, IL: Center on Innovation and Improvement. Accessed March 29, 2013, <http://www.centerii.org/survey/downloads/Turnarounds-Color.pdf>.

Walberg, H. J. (Ed.). (2007). *Handbook on restructuring and substantial school improvement*. Lincoln, IL: Center on Innovation and Improvement.

Waters, T., & Marzano, R. J. (2006). *School district leadership that works: The effect of superintendent leadership on student achievement*. (A Working Paper). Denver, CO: Mid-continent Research for Education and Learning.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. (A Working Paper). Aurora, CO: Mid-continent Research for Education and Learning.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398–425.

Figure 1. STSP Schools relative to State

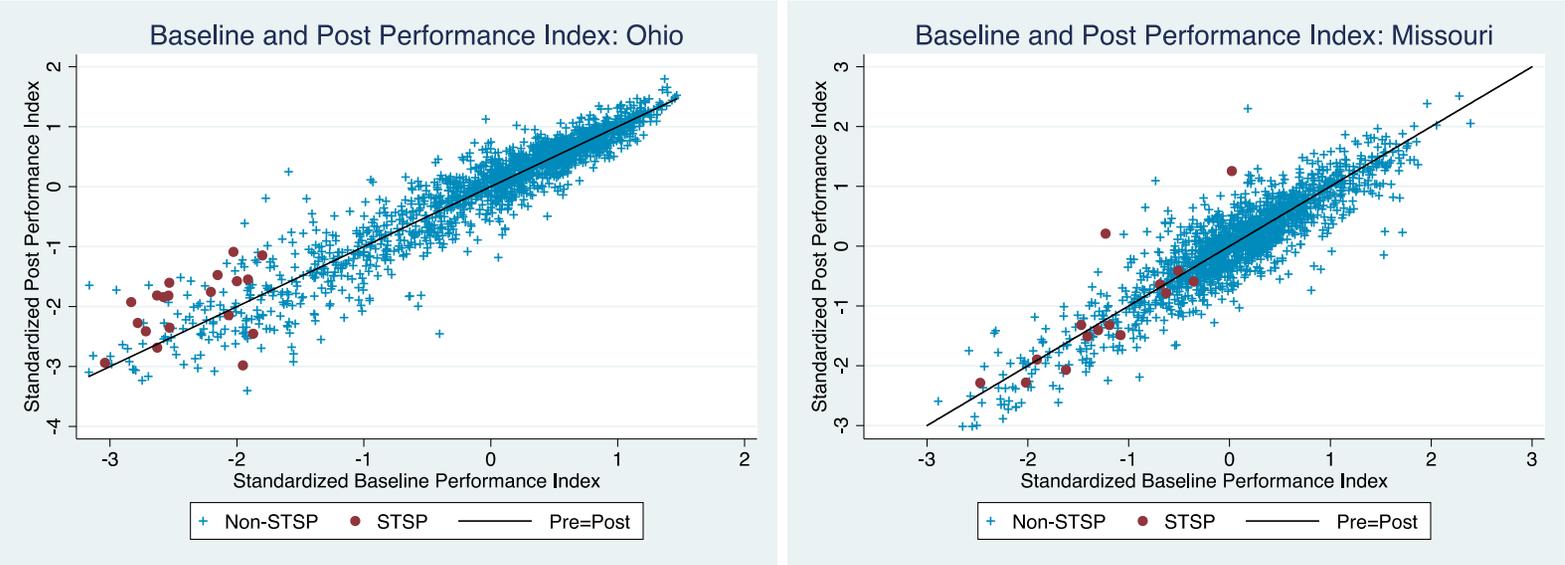
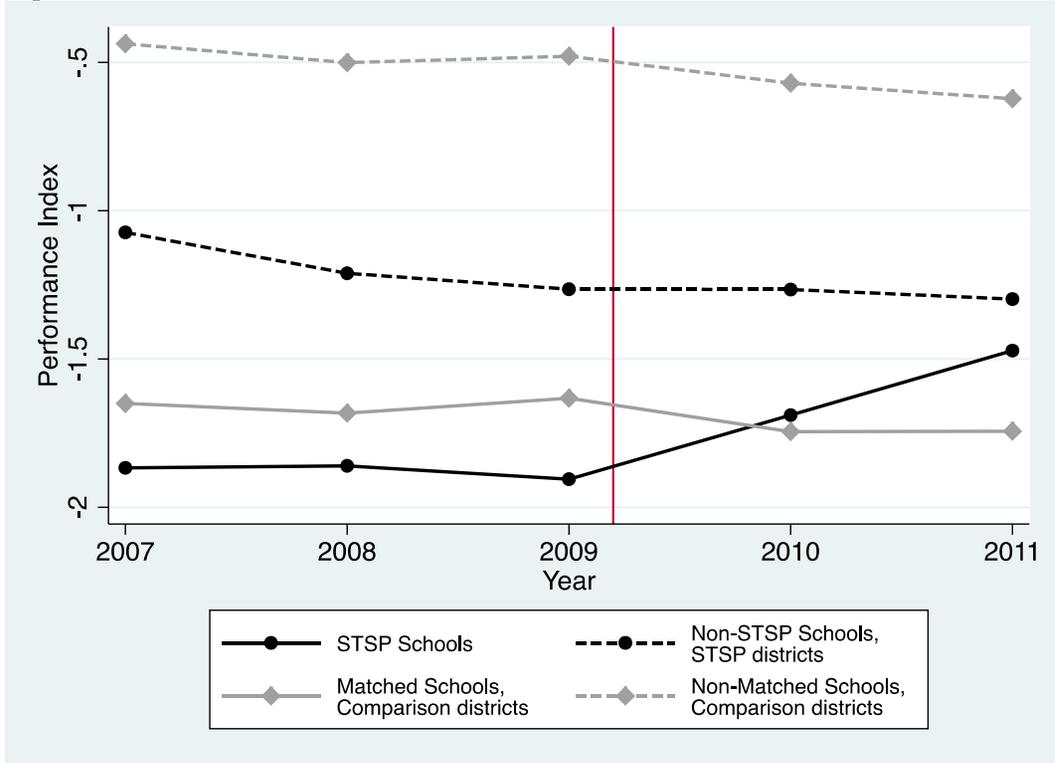


Figure 2. Performance Index over time for STSP and non-STSP schools



Note: Comparison means were weighted to reflect the composition of the treatment sample.

Figure 3. Performance Index over time for STSP and non-STSP schools by State

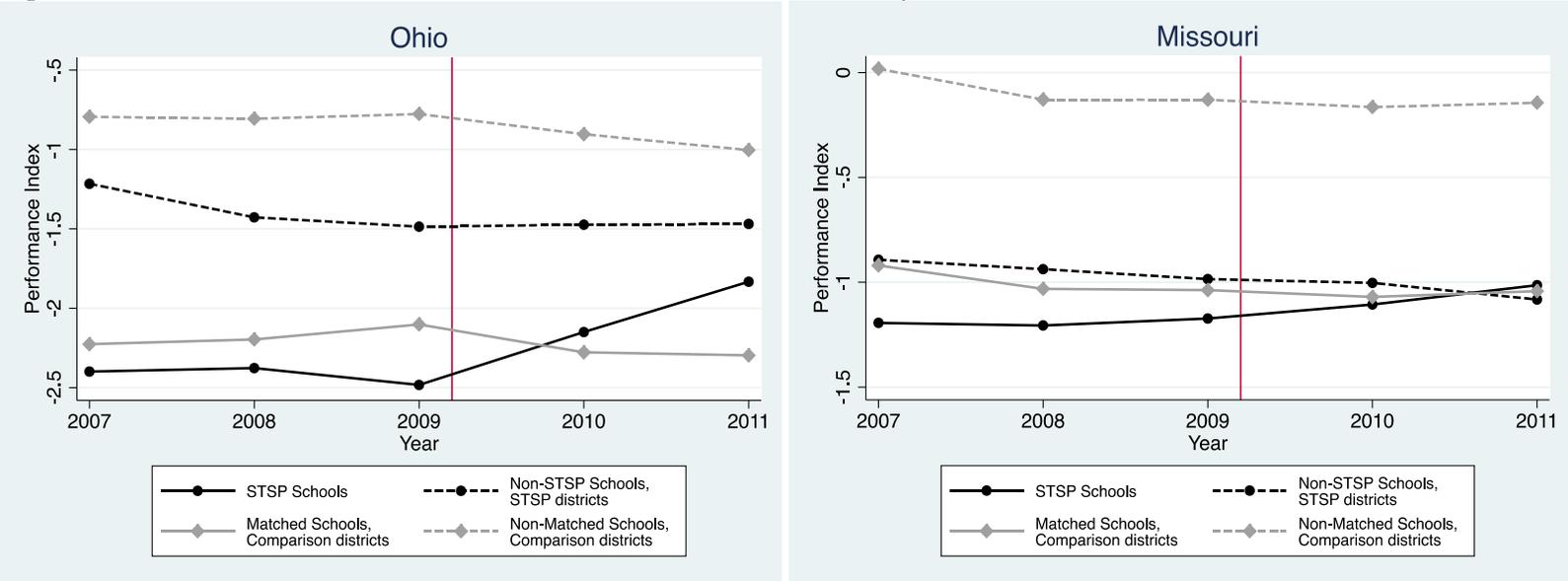
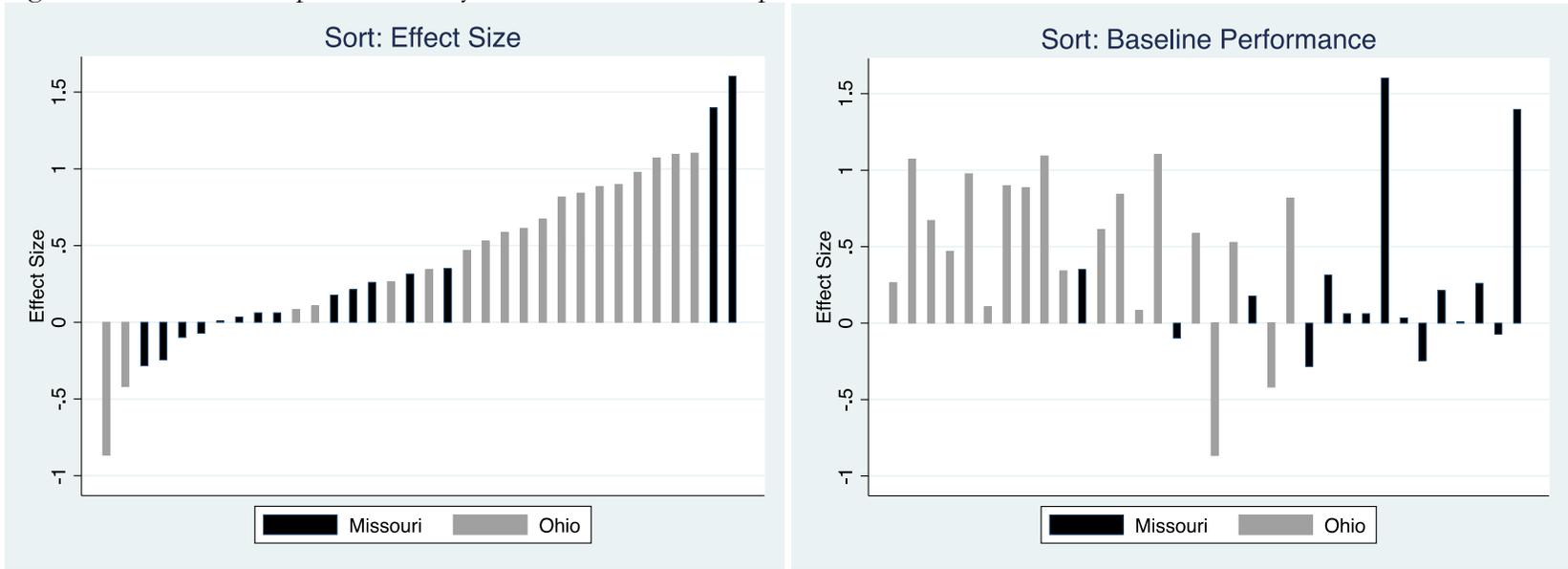


Figure 4. School level impacts sorted by effect size and baseline performance



**Table 1. Baseline (2008-09) Characteristics of STSP Schools and Comparisons**

	STSP Schools	Untreated in STSP District	Comparisons Outside STSP		State Average
			District		
<b><i>Ohio</i></b>					
Std. Performance Index	-2.48	-1.49 ***	-2.10 ***		0.00 ***
% Above Proficient	7.96	17.43 ***	11.60 ***		42.46 ***
% Proficient	21.82	29.03 ***	25.17 **		32.17 ***
% Basic	33.82	31.57	30.85 **		16.19 ***
% Below Basic	34.95	21.61 ***	32.23		9.05 ***
% Free/Reduced lunch	81.45	63.56 ***	78.18		44.28 ***
% Black	81.61	65.74 **	76.34		18.00 ***
% Hispanic	3.76	8.65	4.65		3.29
School Enrollment	511.00	469.67	460.14		430.91
N	19	89	29		1716
<b><i>Missouri</i></b>					
Std. Performance Index	-1.17	-1.12	-1.10		0.05 ***
% Above Proficient	4.48	5.45	4.66		12.22 ***
% Proficient	18.80	19.73	20.44		33.51 ***
% Basic	56.26	53.14	54.84		45.22 ***
% Below Basic	20.46	21.68	20.06		9.05 ***
% Free/Reduced lunch	63.08	40.81 ***	77.99 ***		45.50 ***
% Black	68.47	73.57	69.23		16.33 ***
% Hispanic	8.09	8.21	6.89		3.72 ***
School Enrollment	355.07	328.81	406.26		359.11
N	15	156	66		1717

\* p<.10, \*\* p<.05, \*\*\* p<.01 for difference between comparison and STSP schools

**Table 2. Main Triple Difference Impacts**

VARIABLES	Performance Index	% above proficient	% Proficient	% Basic	% below basic	Daily attendance %
STSP eligible*Post*STSP Dist	0.36*** (0.11)	2.26** (1.05)	2.07** (1.04)	-0.00 (1.15)	-4.04*** (1.36)	0.73 (0.67)
Post*STSP Dist.	0.01 (0.03)	-0.24 (0.40)	0.97*** (0.36)	-0.12 (0.38)	-0.61 (0.44)	0.74*** (0.20)
Post*STSP eligible	0.02 (0.05)	-0.11 (0.45)	1.45*** (0.51)	-0.41 (0.50)	-0.92 (0.66)	-0.13 (0.63)
Post	-0.21*** (0.03)	0.76** (0.37)	0.22 (0.30)	-0.67** (0.30)	-0.33 (0.36)	-0.01 (0.17)
Observations	3,049	3,049	3,049	3,049	3,049	3,049
R-squared	0.066	0.073	0.061	0.097	0.053	0.023

Robust standard errors in parentheses

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

Regressions also include % Black, % Hispanic, % Free/reduced Lunch Eligible, student enrollment, and school and year fixed effects.

**Table 3. Grade-level impacts**

VARIABLES	ELA Proficiency Rates			Math Proficiency Rates		
	3 <sup>rd</sup> grade	4 <sup>th</sup> grade	5 <sup>th</sup> grade	3 <sup>rd</sup> grade	4 <sup>th</sup> grade	5 <sup>th</sup> grade
STSP eligible*Post*STSP Dist	5.71** (2.50)	10.61*** (2.40)	4.18* (2.29)	0.76 (2.89)	5.32* (2.82)	6.81*** (2.62)
Post*STSP Dist.	-0.81 (1.03)	-1.77* (0.91)	0.24 (1.01)	3.29*** (1.16)	0.11 (1.10)	0.18 (1.05)
Post*STSP eligible	0.38 (1.32)	0.38 (1.22)	1.72 (1.23)	2.91* (1.68)	1.46 (1.49)	-0.83 (1.44)
Post	-1.24 (0.81)	3.30*** (0.71)	-6.33*** (0.80)	-3.71*** (0.95)	0.80 (0.87)	1.04 (0.85)
Observations	2,637	2,640	2,594	2,637	2,640	2,594
R-squared	0.034	0.055	0.105	0.071	0.044	0.061

Robust standard errors in parentheses

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

Regressions also include % Black, % Hispanic, % Free/reduced Lunch Eligible, student enrollment, and school and year fixed effects.

**Table 4. State-Specific Impacts**

VARIABLES	Performance Index	% above proficient	% Proficient	% Basic	% below basic	Daily attendance %
<i>Ohio</i>						
STSP eligible*Post*STSP Dist	0.45*** (0.16)	2.75** (1.38)	4.05*** (1.44)	0.13 (1.06)	-6.49*** (2.26)	0.95*** (0.35)
<i>Missouri</i>						
STSP eligible*Post*STSP Dist	0.19 (0.15)	2.11 (1.72)	0.22 (1.50)	-1.10 (2.15)	-1.23 (1.30)	0.13 (1.02)

Robust standard errors in parentheses

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

Each coefficient is from a separate regression. Regressions also include the necessary interactions for the triple difference as well as % Black, % Hispanic, % Free/reduced Lunch Eligible, student enrollment, and school and year fixed effects.

**Table 5. Measuring Impacts on Demographics**

VARIABLES	% Hispanic	% Black	% FRPL	Enrollment
STSP eligible*Post*STSP Dist	0.55 (0.47)	-1.44* (0.79)	0.42 (2.64)	13.13 (11.21)
Post*STSP Dist.	-0.10 (0.19)	0.25 (0.32)	2.70** (1.06)	-9.42** (4.52)
Post*STSP eligible	-0.05 (0.26)	-0.18 (0.44)	-3.98*** (1.47)	-6.39 (6.24)
Post	1.62*** (0.20)	-0.36 (0.33)	4.80*** (1.12)	-6.22 (4.73)
Observations	3,451	3,451	3,451	3,451
R-squared	0.053	0.004	0.101	0.024

Standard errors in parentheses

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

Regressions also include school and year fixed effects.