The Impact of Inclusive STEM High Schools
on Student Achievement

Jennifer Gnagey
Weber State University
Stéphane Lavertu
The Ohio State University

This study is one of the first to estimate the impact of "inclusive" science, technology, engineering, and mathematics (STEM) high schools using student-level data. We use multiple statistical strategies to estimate the effect on student achievement from 2 years of attendance at six such high schools in Ohio. The results indicate that two schools had positive effects on science achievement that appear to come at the expense of achievement in social studies. The other schools had negligible or, often, negative effects across both STEM and, particularly, non-STEM subjects. These results are consistent with studies indicating that inclusive STEM schools typically focus on problem-based, personalized learning rather than science and mathematics content. The analysis also reveals the importance of accounting for students' prior test scores in science, in addition to math and reading, when estimating models that use only 1 year of prior test score data—something that existing studies fail to do.

Keywords: *inclusive STEM high schools, student achievement*

IMPROVING science, technology, engineering, and mathematics (STEM) education has become a national priority. The National Academies' 2007 publication *Rising Above the Gathering Storm* asserts the importance of addressing the eroding achievement of U.S. students in mathematics and science relative to students in other countries, as well as the importance of addressing racial and gender achievement gaps in these fields. Subsequent reports have echoed these concerns—including the *Report of the Academic Competitiveness Council* (Academic Competitiveness Council, 2007) and *The Opportunity Equation* (Carnegie Corporation of New York, 2009)—and policy makers and other political actors have suggested STEM-focused schools as one possible solution. For example, the America COMPETES Act of 2007 explicitly called for providing "assistance to the states for the costs of establishing or expanding public, statewide specialty schools for mathematics and science," and in 2010, a report by the President's Council of Advisors in Science and Technology called for the creation of 1,000 more STEM-focused high schools within the next decade.

An emerging model is the "inclusive" STEM high school. This type of high school combines nonselective admission policies with a STEM-focused curriculum to seed interest in the STEM fields and expand the STEM workforce pipeline (Means, House, Young, & Wang, 2013). It has proliferated

over the last decade, enjoying support from a variety of political actors and education reformers, such as the Obama administration and the Bill and Melinda Gates Foundation. Yet, due to its recent emergence, there is little research that examines its academic impact. To our knowledge, Wiswall, Stiefel, Schwartz, and Boccardo's (2014) study of 30 selective and nonselective STEM high schools in New York City is the only published study to use student-level data.[1] They find positive impacts of STEM school attendance on some math and science exams. However, although their use of student-level data enables them to better account for selection bias than analyses that employ school-level data, they do not account for students' prior achievement in science. As our analysis reveals, that is a potentially important omission.

This study estimates the impact on academic achievement of six inclusive STEM high schools established in Ohio since 2006. It employs 2006–2013 student-level administrative data from the Ohio Department of Education to estimate the impact of 2 years of attendance at these high schools on 10th-grade standardized test scores in STEM subjects (math and science) and non-STEM subjects (reading and social studies). First, we estimated ordinary least squares (OLS) regression models comparing the achievement of STEM school students with the achievement of corresponding students in traditional public schools in STEM

students' districts of residence. Second, we estimated OLS models for which we used different propensity score–matching techniques to create comparison groups. Finally, for one school, we were able to use the results of admission lotteries as an exogenous source of variation in STEM school attendance to identify the causal impact of STEM school attendance on student achievement. Consistent with prior studies of student achievement, the models all account for prior student test scores in math and reading, in addition to the typical battery of demographic covariates. Additionally, contrary to existing research on STEM high schools, we compare estimates from these models to estimates from models that account for prior student test scores in science.

There is some debate in the literature about the validity of the nonexperimental techniques that we employ, but recent evidence from Forston, Gleason, Kopa, and Verbitsky-Savitz (2014) suggests that propensity score–matching estimates will be similar to experimental estimates if the pretreatment variables employed are highly predictive of posttreatment outcomes. Lagged test scores, particularly in the same subject as the outcome score, are strongly predictive of future test scores. Indeed, recent research has shown that bias is minimal in such "value added" estimates of school quality (e.g., see Abdulkadiroğlu, Angrist, Dynarski, Kane, & Pathak, 2011; Angrist, Pathak, & Walters, 2013; Deming, 2014).[2] In our case, the 10th-grade graduation tests contain much of the same content as the eighth-grade tests that we use to account for students' educational histories. This strengthens the validity of our analysis for the two STEM subjects (math and science) as well as reading and represents a marked improvement over existing research that has failed to account for prior science scores that are highly predictive of performance on the achievement outcomes of interest.

The results indicate that, on average, attendance at these inclusive STEM high schools had a negative impact on student achievement across all four subjects—particularly the non-STEM subjects, reading and social studies. Although there is significant heterogeneity in effects across the six schools, only two schools are associated with statistically significant achievement benefits—around a 0.15–standard deviation advantage in science—but this impact is accompanied by a 0.1– to 0.25–standard deviation disadvantage in social studies. However, two schools are associated with large negative effects across all four of the tested subjects, with the negative effects being most pronounced in the non-STEM subjects. There is also evidence that African American students bear the brunt of these negative effects, whereas there were no significant differences in impact between male and female students.

The results are quite similar across the various models, whether estimates are based on all available STEM school and feeder district students, matched samples, or a restricted sample based on the availability of test scores in all subjects. However, the estimated effects of STEM school attendance depend significantly on whether models account for students' eighth-grade test scores in science, in addition to test scores in math and reading. The results that we report in the preceding paragraph are based on models that account for students' prior science scores because these models more plausibly account for prior student interest and aptitude in STEM fields. Indeed, our analysis suggests that nonexperimental studies of STEM schools that fail to account for prior science test scores should be viewed with skepticism. For example, the two independent STEM schools in our sample seem to have a positive impact on math achievement in models that account only for prior reading and math achievement. However, the magnitude of these impacts is cut in half and no longer approaches conventional levels of statistical significance when models account for students' eighth-grade science scores. It is also noteworthy that accounting for prior science scores has a similar impact on estimates of achievement in non-STEM subjects.

Finally, it is worth emphasizing that all schools in the analysis were established in 2006 or later and that our estimates for each school are based on one to six student cohorts. The literature indicates that new charter schools tend to underperform for their first 3 to 5 years of operation, as compared with observationally similar traditional public schools, but then catch up to and sometimes surpass them afterward (Bifulco and Ladd 2006; Hanushek, Kain, Rivkin, & Branch, 2007; Sass 2006; Zimmer et al., 2009). There is reason to believe that these initial years were particularly trying for some of the STEM schools in our study, as some of them underwent construction or conducted classes in alternative buildings during a portion of the observation period. Accordingly, we estimated linear time trends to examine achievement progression in the five schools with multiple cohorts. The results show mostly null trends. Among the few discernable trends, all but one were negative.

As we discuss below, these results are not all that surprising. LaForce et al. (2014) show inclusive STEM schools do not generally emphasize coursework in science and mathematics. Instead, they typically focus on problem-based and personalized learning and the development of life skills. However, political actors often speak of STEM education as a critical way to improve student learning in subjects such as science and math. Indeed, our analysis indicates that students who score highly on science and math exams are far more likely to select into some of these schools, which is consistent with the notion that a substantive emphasis in these subjects is of import to parents. The results suggest that parents and policy makers may not always be getting from these schools what they think they are getting.

The remainder of this article is organized as follows. The next section provides background on STEM high schools in the United States, and the third describes the schools that are the subject of this study. The fourth section describes the

data, and the fifth discusses the empirical methods. The sixth section reviews the results, and the last offers concluding thoughts.

STEM High Schools in the United States

Specialized STEM high schools in the United States date back to the establishment of Stuyvesant High School in 1904 and Bronx High School of Sciences in 1938. The late 1960s and early 1970s saw a surge in the establishment of specialized STEM schools, with the birth of magnet schools in the wake of the civil rights movement. In 1977, North Carolina became the first state to establish a statewide residential STEM high school for talented students, and Louisiana, Illinois, and Mississippi followed suit in the mid-1980s (Eilber, 1987; National Consortium for Specialized Secondary Schools of Mathematics, Science, and Technology, 2014). The rapid expansion of STEM magnet and state residential high schools in the 1980s led to the establishment of the National Consortium for Specialized Secondary Schools of Mathematics, Science, and Technology in 1988 (Atkinson, Hugo, Lundgren, Shapiro, & Thomas, 2007; Thomas & Williams, 2010). These schools continued to proliferate through the 1990s and into the 21st century. As of 2010, 28 states had public residential STEM high schools, and there are many more local STEM magnet and charter schools (Thomas & Williams, 2010).

The nonselective, "inclusive" STEM high school is an emerging model that marks a significant departure from previous approaches (Carnegie Corporation of New York, 2009; Means, Confrey, House, & Bhanot, 2008). Prior to the early 21st century, a majority of STEM-focused high schools, whether publicly or privately funded, were selective, with admission decisions based on competitive exams or previous academic achievement (Means et al., 2008). These schools have mainly sought to cultivate and strengthen existing STEM talent and interest. Inclusive STEM high schools, however, distinguish themselves by their nonselective admission policies. In cases of oversubscription, admission decisions are typically made by random lottery. Consequently, they also are thought to help address gender and racial STEM interest and achievement gaps. Studies have shown that inclusive STEM high schools tend to serve more racially diverse student bodies than both their selective counterparts and traditional public schools (Means et al., 2008; Means et al., 2013).

Inclusive STEM high schools are difficult to define because they do not operate under a single umbrella philosophy or organizational structure (Lynch, Behrend, Burton, & Means, 2013). Most recent studies, including those cited above, identify these schools from the use of nonselective admission policies, a school's self-proclaimed emphasis on the STEM fields, and perhaps a school's affiliation with an organized STEM education initiative. These schools vary

significantly in their educational practices, but a systematic analysis of 25 inclusive STEM schools across the country conducted by the University of Chicago's Center for Elementary Math and Science Education (see LaForce et al., 2014) revealed that inclusive STEM schools have some common characteristics.

First and foremost, LaForce et al. (2014) found that these schools feature problem-based learning, interdisciplinary instruction, student autonomy, and "rigorous learning," which often entails mastery learning and a staff-created curriculum that features real-world applications. These schools also emphasize establishing a positive school culture, developing skills that students can use in their everyday lives and future careers (e.g., technological proficiency, communication, and collaboration), personalized learning (e.g., differentiation of instruction based on ability and relevance to students' lives), and a connection between the school and local community (e.g., partnerships with external educational and business organizations). Thus, LaForce et al. (2014) discovered that what makes these schools "STEM" is not necessarily a greater emphasis on STEM subjects. Although that is the case in some inclusive STEM schools, what makes them "STEM" schools is primarily the use of problem-based, interdisciplinary, and personalized learning approaches—none of which is unique to this recent STEM movement.

Inclusive STEM High Schools in Ohio

The Bill and Melinda Gates Foundation has played a lead role in propagating the inclusive STEM high school model by funding large-scale initiatives in Texas, Ohio, North Carolina, Tennessee, and Washington throughout the early 21st century, often supporting state-initiated programs. Our study focuses on six such high schools in Ohio, established through state and local public-private partnerships and featuring the Bill and Melinda Gates Foundation as an important benefactor.

A major university and two nonprofit organizations in Ohio established a STEM-focused early-college high school in 2006, drawing students from multiple districts in a large metropolitan area. This high school was inclusive and employed a lottery-based admission system. One year later, these nonprofits committed funds and received significant financial support from the Bill and Melinda Gates Foundation and the State of Ohio to establish the Ohio STEM Learning Network (OSLN), which is anchored by a series of STEM "platform" schools across the state. The platform schools, including the original STEM high school established in 2006, were designed to serve as laboratories for STEM education and to disseminate best practices. They are located throughout the state so that they can tailor education to local needs.

Platform schools must commit to inclusive admission policies and five broad "design principles" established by

TABLE 1
STEM High Schools in the Analysis

School	Type	First Year of Operation	Cohorts Tested in Grade 10[a]	Grade Spans[b]	Enrollment[b]	Teacher FTEs[b]	Student:Teacher Ratio
A	Independent	2006–2007	6	9–12	394	20	19.7
B	Independent	2009–2010	3	6–12	416	28.5	14.6
C	District	2009–2010	3	7–12	995	62.8	15.8
D	District	2008–2009	4	9–12	289	10	28.9
E	District	2010–2011	2	9–12	494	21.2	23.3
F	District	2011–2012	1	9–12	422	18.7	22.6

Note. FTE = full-time equivalent; STEM = science, technology, engineering, and mathematics.
[a]2012–2013 or prior.
[b]As of 2012–2013.

the OSLN: "form new skills and sharp minds for a new century," "engage partnerships to accelerate capacity and broaden opportunity," "start and stay small," "make STEM literacy attainable and desirable for all," and "drive scalable and sustainable innovations" (OSLN 2014). There are few other requirements. This study focuses on the six OSLN platform schools that were still part of the OSLN in 2012–2013 and were at least 2 years old by the end of the 2012–2013 school year, given that this was the last year of available data.[3] All of these high schools also are (or eventually would be) state-designated STEM schools. To achieve a state "STEM" designation, schools must, among other things, provide the following:

(3) Evidence that each school will offer a rigorous, diverse, integrated, and project-based curriculum to students in any of grades six through twelve, with the goal to prepare those students for college, the workforce, and citizenship, and that does all of the following:

(a) Emphasizes the role of science, technology, engineering, and mathematics in promoting innovation and economic progress;
(b) Incorporates scientific inquiry and technological design;
(c) Includes the arts and humanities;
(d) Emphasizes personalized learning and teamwork skills. ("STEM School Grants," 2007)

Thus, Ohio's STEM schools are required by law to feature problem-based and personalized learning and the development of life and career skills, which is consistent with the practices of inclusive STEM schools across the country.

Additionally, the Ohio Department of Education states that proposals to establish a state-sanctioned STEM school must do the following:

1. Create a public school (in any of the Grades 6–12) to help generate a talent base that will establish Ohio as a magnet and global leader in attracting, educating, and producing the next generation of scientists, engineers, and other professionals needed to create tomorrow's innovations for the betterment of all citizens;

2. Foster increases in the number of Ohio citizens studying and working in STEM fields; and

3. Foster increases in all students developing stronger skills in problem solving, innovation, and teamwork. (Ohio Department of Education, 2015)

Presumably, these goals entail increasing proficiency in the STEM fields—including science and mathematics. Indeed, just below the delineation of these goals on the Ohio Department of Education's website, the agency also emphasizes its role in helping schools with their educational programs in math and science.

Table 1 provides some basic information about each school that we examine in this study. As the table indicates, two of the schools in this study are state-chartered independent high schools (called "state STEM schools") that receive a portion of district funds tied to their students and that rely heavily on private contributions, whereas the rest were established and are run by traditional public school districts. Additionally, some of these schools are relatively new, whereas others are traditional public schools that converted to STEM schools. Finally, the table reveals that, as of the 2012–2013 school year, Schools B and C included middle school grades. It is worth noting that all but one of the student cohorts analyzed in this study entered these schools in Grade 9 and that these schools added lower grade levels only later. The sole exception is that one cohort from School B entered in the seventh grade. Although this may affect this group's eighth-grade scores, because we use a value-added approach, this will not interfere with out estimation of treatment effects in ninth to 10th grade.

Once again, these schools received the state's "STEM school" designation and should thus, to some extent, emphasize problem-based and personalized learning, develop life and career skills, and generate interest and proficiency in the STEM fields. In practice, however, there is great variation in how these schools operate. For example, Schools A, B, and E were part of LaForce and colleagues' (2014) study, which revealed that all three schools emphasize project-based

learning, mastery learning, student autonomy, and flexible schedules. But these schools also differ from one another. For example, unlike School E (the district school), Schools A and B (the two independent state-chartered schools) also noted in surveys that students must complete cognitively demanding work. Additionally, these three schools (A, B, and E) differ significantly from the others in our analytic sample. For example, as the analysis below reveals, they are far more likely than the others to attract students with high proficiency levels in science and math.

These STEM schools have received significant resources from government and private actors. For example, in 2007, Ohio House Bill 119 allocated $13 million to STEM schools and programs, and the Bill and Melinda Gates Foundation added $12 million to create the network of STEM hubs and platform schools (OSLN, 2012). Indeed, anecdotes abound that schools have sought the STEM designation simply to secure the millions of dollars in grants that the state distributes to these schools. In addition, Race to the Top funds, as well as a multitude of financial and other gifts[4] from private and public STEM school partners, have subsidized STEM schools and their students (e.g., through the provision of supplies)—often above and beyond the funding that these schools get from the typical local, state, and federal sources.

Unfortunately, systematic, building-level financial data are unavailable for all six schools, but an analysis of enrollment and student:teacher ratios revealed that these STEM high schools vary significantly in size and instructional resources. During the 2012–2013 school year, the median public high school in the state enrolled 465 students and employed 28.6 teachers—a student-teacher ratio of 16.26.[5] As Table 1 indicates, four of the six STEM schools are "small" in that they have enrollments lower than the median of the traditional district high school and only two have student:teacher ratios lower than the state median. Indeed, the district schools are either large (School C)—which conflicts with the "small school" model that OSLN has sought to implement for platform schools—or small with high student:teacher ratios (Schools D and F). Indeed, these figures suggest that the financial benefits of being an OSLN-platform STEM high school—the focus of this analysis—come in the form of equipment or access to community resources (including expertise from nearby universities, for example), as opposed to teacher labor.

In summary, although political actors have touted inclusive STEM high schools as potential solutions to low student achievement in math and science—especially, achievement gaps in these subjects—case studies of these schools suggest that their focus is not generally on increasing student achievement in these subjects. Additionally, although there have been millions of dollars in public and private money spent on these schools, they do not generally invest more resources in the most important school-based educational input: teachers. These facts suggest that Ohio's STEM platform schools are unlikely to generate large overall gains in student achievement and that any gains in STEM subjects (science and math) could come at the expense of achievement in non-STEM subjects. Finally, the variability in these schools' designs, educational practices, and resources likely correlates with significant variability in their effectiveness when it comes to STEM instruction.

Data

The analysis employs student-level administrative data collected by the Ohio Department of Education from the 2005–2006 (fiscal year [FY] 2006) to 2012–2013 (FY2013) school years. These data include standard student-level information on the building and district of attendance, demographics, and scale scores on state standardized tests. The tests are administered at the end of the school year in grades 3–8 and grade 10. A passing grade on the 10th-grade tests is required for a high school diploma. Because all students are tested in both eighth and 10th grade, we focus on the estimation of 2-year STEM high school treatment effects.[6] The first STEM platform school admitted its first freshman class in FY2007, so the FY2006–FY2013 data allow us to track up to six cohorts of students at each platform school from eighth through 10th grade.

Table 2 displays summary statistics for the sample of STEM school students as well as the sample of feeder district students—the control group in this study. For district STEM schools, the operating district is the feeder district,[7] whereas the independent STEM school feeder districts include all districts from which at least one student in a cohort is included in the sample of STEM students. Test scores presented in the table, as well as those used in the analysis below, have been normalized to have a mean of zero and a standard deviation of one within each grade, subject, and year. Additionally, the student sample on which the table is based is restricted to students for whom we had 10th-grade test scores across all subjects, as this restricted sample is the focus of the analysis below. However, the descriptive statistics (and the results of the analysis) are similar for the full sample of students.

Following Tuttle, Teh, Nicholas-Barrer, Gill, and Gleason (2010), we use a quasi intent-to-treat approach for defining our samples of STEM and feeder district students. Specifically, any student who was enrolled in a STEM school in ninth grade is included in the STEM school sample regardless of whether she or he stayed through the end of 10th grade. Similarly, the sample of feeder district students includes those who attended the feeder district in ninth grade regardless of whether they remained in the same school, transferred within the district, or transferred outside the district in 10th grade. This sample definition mitigates bias from differences in school attrition rates between STEM school and feeder district students.

TABLE 2

Descriptive Statistics for Student Sample Used in the Analysis

	School A		School B		School C	
	STEM	Feeder	STEM	Feeder	STEM	Feeder
Mean eighth-grade math score (z score)	0.62	0.05	0.41	0.15	−0.61	−0.13
Mean eighth-grade reading score (z score)	0.60	0.03	0.41	0.14	−0.52	−0.06
Mean eighth-grade science score (z score)	0.58	0.00	0.59	0.09	−0.78	−0.28
Percentage female	0.46	0.50	0.34	0.51	0.61	0.56
Percentage African American	0.28	0.29	0.20	0.17	0.92	0.64
Percentage Hispanic	0.04	0.04	0.00	0.02	0.01	0.01
Percentage other non-White	0.11	0.07	0.12	0.07	0.02	0.06
Percentage economic disadvantage	0.42	0.40	0.27	0.32	0.86	0.57
Percentage limited English proficiency	0.04	0.05	0.00	0.01	0.04	0.02
Percentage disability	0.05	0.11	0.08	0.11	0.23	0.16
Percentage gifted	0.64	0.33	0.18	0.25	0.07	0.24
Percentage attended charter in eighth grade	0.05	0.01	0.30	0.02	0.04	0.04
Cumulative school attrition rate[a]	0.21	0.11	0.30	0.06	0.18	0.11
Student count, *n*	306	34,722	142	11,449	256	3,322

	School D		School E		School F	
Mean eighth-grade math score (z score)	−0.35	−0.65	0.67	0.05	0.06	0.01
Mean eighth-grade reading score (z score)	−0.26	−0.56	0.57	0.09	0.13	0.07
Mean eighth-grade science score (z score)	−0.46	−0.81	0.86	0.20	0.30	0.20
Percentage female	0.45	0.53	0.36	0.51	0.83	0.54
Percentage African American	0.73	0.71	0.26	0.39	0.46	0.37
Percentage Hispanic	0.12	0.12	0.02	0.02	0.02	0.02
Percentage other non-White	0.03	0.03	0.11	0.11	0.10	0.13
Percentage economic disadvantage	0.95	0.95	0.27	0.40	0.49	0.40
Percentage limited English proficiency	0.04	0.05	0.03	0.04	0.06	0.04
Percentage disability	0.11	0.18	0.05	0.13	0.05	0.13
Percentage gifted	0.28	0.14	0.0	0.18	0.27	0.22
Percentage attended charter in eighth grade	0.11	0.04	0.00	0.01	0.02	0.01
Cumulative school attrition rate[a]	0.21	0.29	0.11	0.12	0.19	0.09
Student count, *n*	130	7,096	188	318	63	182

Note. The table presents descriptive statistics for the restricted sample of students used to estimate the ordinary least squares models based on unmatched samples. Specifically, the sample is restricted to students for whom we had 10th-grade test scores in all four subjects and who attended the science, technology, engineering, and mathematics (STEM) school or a school's feeder district. The table presents the mean test scores (standardized by year, grade, and subject) in standard deviation units for all students who attended the STEM school or the traditional public school district in which STEM school students resided. It also reports the percentage of the STEM school or feeder district students who fit into the demographic categories and who previously attended charter schools.

[a]For feeder district schools terminating in ninth grade, only out-of-district transfers are counted toward the school attrition rate. Schools E and F share a feeder district in which the primary district high school closed after fiscal year 2011. Therefore, only out-of-district transfers between fiscal years 2011 and 2012 for students in this feeder are counted toward the school attrition rate.

Cumulative school attrition rates—defined as the percentage of students in the sample who did not attend the same school for two full academic years[8] in ninth and 10th grade—are also displayed in Table 2. As the table indicates, Schools A, B, and C have significantly higher attrition rates than their feeder district schools, whereas School D's attrition rate is significantly lower than that in its feeder district schools. These differences support the use of the quasi intent-to-treat approach.

Overall, Table 2 demonstrates that there are significant differences between the STEM school students and their counterparts in traditional public schools and that these differences also vary significantly among the six schools. For example, the two independent STEM schools—Schools A and B—draw from numerous districts, which is why there are many feeder district students for comparison; in addition, their students—as well as students who attended district Schools D and E—have significantly higher prior test

scores than their counterparts attending traditional public schools in their districts of residence. However, the students at School C—a district STEM school that, unlike the others, is more of a neighborhood school than a choice school—tend to have lower eighth-grade test scores, are more economically disadvantaged, are more likely to be African American, are more likely to be female, and are less likely to be labeled gifted when compared with their feeder district counterparts.[9] Finally, students attending independent STEM schools, as well as district School D, were more likely than their traditional district counterparts to have attended charter schools.

It is important to note that the analysis is based on only a subset of students attending the STEM schools and feeder districts. In particular, the sample of STEM school students included in the regressions is only about three-fifths of the sample of attendees due to the combination of unmatched eighth-grade records and missing test scores and demographic information. The data are particularly limited for Schools C and D, as only 50% of their students are included in the analysis. This raises legitimate concerns about the accuracy of the information in the data set and the external validity of our findings. To the extent that data are inaccurate or not missing at random, our results may not reflect the impact of the STEM schools across the entire population of attendees. Given recent concerns about data manipulation in Ohio,[10] omission of poorly performing students is a particular concern. However, Schools C and D, for which the data are most limited, also exhibit the poorest performance, which mitigates the concern that poorly performing students were systematically excluded from the data.

Empirical Methods

Our empirical strategy consists of estimating student growth models comparing the achievement of students who attended one of the six STEM schools with students who attended traditional public schools in the feeder districts. Specifically, the analysis entails the estimation of models that include all feeder district students (with some restrictions) as well as models estimated with a smaller, matched sample. We also used an instrumental variables approach to estimate the impact of STEM school attendance for one of the six schools. This section reviews the procedures that we used to conduct these analyses.

Student Achievement Model

We primarily employ growth models to address potential selection bias and to identify STEM school treatment effects. Specifically, the treatment effect estimates are based on the following model of student achievement:

$$y_{ig} = \beta_0 + \lambda y_{ig-2} + \beta x_{ig} + \gamma D_{ig} + \varepsilon_{ig},$$

where y_{ig} is student i's test score in 10th grade and y_{ig-2} is a vector of eighth-grade test scores.[11] The vector x_{ig} represents observable student-level characteristics and includes indicators for gender, race, economic disadvantage, limited English proficiency, disability status, gifted status, and charter school attendance in eighth grade, as well as cohort by district of residence fixed effects. D_{ig} indicates whether or not a student attended a STEM school, and ε_{ig} is a random error term. The parameter of interest is the STEM school treatment effect, γ. Essentially, students in these models are exactly matched according to their cohort and district of residence. The lagged test scores control for the pretreatment effects of observable inputs and the time-constant effect of the unobservable initial endowment that affects student achievement levels. Treatment effects are identified under the assumption of common trends conditional on observables.

To evaluate heterogeneity in school effectiveness, school-specific treatment effects are estimated with a slightly modified version of the model:

$$y_{ig} = \beta_0 + \lambda y_{ig-2} + \beta x_{ig} + \gamma^s D_{ig}{}^s + \varepsilon_{ig},$$

where D_{ig}^s is a vector of dummy variables indicating attendance at each of the STEM high schools (or at one of two STEM school types, depending on the analysis) and where γ^s is a vector of treatment effects by school (or by school type, depending on the analysis). In subsequent analyses, we also estimated all models using student samples restricted by race and gender to examine how STEM school attendance affected the achievement of student groups that are thought to benefit from inclusive STEM schools. Finally, we estimated STEM school achievement trends over time using models that estimate a linear trend in STEM school performance.

The STEM school students included in the first set of models are those who attended one of the six STEM schools and for whom we had 10th-grade test scores in all subjects. The feeder district students included in these models are those attending a traditional public school in ninth grade in a district where at least one resident in their cohort attended a STEM school and for whom we had 10th-grade test scores in all subjects. More explicitly, following Angrist et al. (2013), feeder district students are restricted to those enrolled in traditional public schools—not charter schools—in their respective districts of residence. The results are similar if the samples are not restricted to students with all four test scores.

Matching Analysis

The sample of feeder district students described above may not be an appropriate control group for evaluating STEM school performance, because STEM school students are not representative of their respective cohorts and districts

of residence. We conducted additional analyses based on matched samples to address this potential problem. Because matching improves balance in the distribution of student characteristics across treatment and control groups, we minimize the implicit extrapolation of treatment effects to dissimilar groups of students.

In general, the identifying assumptions underlying matching techniques are very similar to those underlying simple OLS regression (Angrist & Pishke, 2009). As such, matching is an effective strategy for addressing bias due to selection on observable characteristics. We employ lagged test scores as matching variables and in subsequent regressions on matched samples, so the matching treatment effect estimates presented below also are primarily identified under the assumption of common trends conditional on observables.

A broad array of matching techniques can be applied in program evaluation contexts. Two of these techniques, exact and propensity score matching, are particularly common in the school choice literature (Angrist et al., 2013; Dobbie & Fryer, 2013; Tuttle et al., 2010; Woodworth, David, Guha, Haiwen, & Lopez-Torkos, 2008). We report the results based on propensity score–matching techniques because the dimensionality problem associated with exact matching required us to drop many student observations. Propensity score matching entails matching observations on their probability of selecting into treatment (i.e., their propensity score) rather than on vectors of covariates and relying on the balancing property of the propensity score to achieve desirable sample balance.

After matching and evaluating the balance of matched samples, we estimated treatment effects based on these matched samples using the models described above. We also weighted these samples to account for imbalance in the number of students in the treatment and control groups.

Procedure for Matching Based on Propensity Scores

First, we estimated probit models of STEM school attendance that are some variant of the following form:

$$\pi_i = \Pr\left(D_{ig} = 1 \mid y_{ig-2}, x_{ig}\right)$$
$$= \Phi(\theta_0 + \theta_1 y_{ig-2} + \theta_2 x_{ig}),$$

where Φ represents the normal cumulative distribution function. Due to the heterogeneity of the student populations across the six platform STEM high schools and their respective feeder districts, we estimated separate probit propensity score models for each STEM high school. As in the achievement model, the demographic variables (x_{ig}) include indicators for gender, race, economic disadvantage, limited English proficiency, disability status, and gifted status, as well as an indicator of charter school attendance in eighth grade. Finally, models also include cohort by district of residence

fixed effects if schools drew from multiple districts or had multiple cohorts tested in 10th grade.

The analysis based on the larger, unmatched sample of students indicates that it is important to control for baseline test scores in math, reading, and science. For this reason we included all of these test scores in the propensity score models when possible.[12] Additionally, we obtained the greatest balance when we modified the above model by interacting all test score variables with one another, including the square of each test score variable and by interacting demographic characteristics with one another. We also estimated propensity scores using models with no interaction terms, using linear probability models, and using separate models for each school-cohort combination. We selected the propensity score model presented here because it provides superior covariate balance. (The results of the probit propensity score models are presented in Appendix A.)

We used two techniques to match students on the basis of their propensity scores—1:1 nearest-neighbor matching within caliper and radius matching, both with replacement. Specifically, we first matched students exactly on cohort and district of residence and then applied the two matching techniques within these exactly matched subgroups. To ensure common support over the range of propensity scores among treatment and control groups, we required matched controls to fall within a specified caliper around the estimated propensity score of the treatment observation to which it is matched. Specifically, given the results from Cochran and Rubin (1973) and following Rosenbaum and Rubin (1985), we specified a caliper equal to 0.25 standard deviations of the pooled standard deviation of the estimated propensity scores across treatment and control students.

Following Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd (1998), we calculated balance statistics separately for each propensity score model. We evaluated covariate balance according to the standardized bias in covariate values between pre- and postmatching treatment and control groups (see Rosenbaum & Rubin, 1985; Stuart, 2010). (The procedure and balancing results are presented in Appendix B.) Additionally, because radius matching is one to many and because both matching techniques involve replacement (as some STEM schools draw from the same districts), balance statistics and subsequent regressions are calculated with weights for the matched control group where each control student i is given a weight, w_i, equal to

$$w_i = \sum_{m=1}^{M_i} \frac{1}{N_m},$$

where M_i is the total number of treatment observations to which control student i is matched and where N_m is the total number of control students matched to the mth treatment student to which control student i is matched. This makes the total weight of the matched control sample equal to the number of students in the matched treatment sample.

Using Lottery Results as an Instrument

We were able to obtain data on admission lotteries for one of the schools (School A). Thus, for this school, we also estimated the average effect of attending School A for those students who attended as a result of winning the lottery. We used a two-stage least squares regression in which lottery outcomes served as an instrument for STEM school attendance (see Angrist & Imbens, 1994). Specifically, restricting the analysis to students who applied to attend this school—some of whom won a seat and some of whom did not—in the first stage, we regressed STEM school attendance on the lottery outcome and the full battery of covariates from the models described above. We then used the estimated probability of attendance for lottery winners and losers in the second stage, along with all covariates, to estimate the impact of STEM school attendance across the three cohorts for which we had admission data. Standard errors are corrected for the fact that the probability of attendance is estimated rather than measured directly.

The primary threat to internal validity is that admissions were not in fact random. To test for this possibility, we regressed each pretreatment covariate on the lottery outcome. None of the regressions yielded a coefficient estimate with a magnitude of 0.04, and none of the coefficients approached conventional levels of statistical significance. These tests support the internal validity of this design. Thus, this limited analysis's internal validity complements the analysis that we describe above, which has superior external validity.

Results

The results presented in this section are from models limited to students for whom we have 10th-grade test scores across all four subjects. We report only these results so that we can compare outcomes across subjects and because they are broadly similar to those that we obtained using all available student data across all specifications. First, we present results from the OLS models that employ the larger, unmatched sample of students, followed by a similar analysis based on samples created with the two propensity score–matching techniques described above. We then present the results of models estimated individually for males, females, African Americans, and Caucasians, as well as by student cohort within each school.

OLS Models Based on Unmatched Sample

Table 3 presents the estimated impact of STEM school attendance on achievement in math, science, reading, and social studies using the larger, unmatched sample of students. Each column presents the estimated effects from three separate OLS regressions: one that estimates a single STEM school effect ("all STEM schools"), one that estimates separate parameters for independent and district STEM schools ("disaggregated by STEM school type"), and one that estimates separate parameters for each STEM school. Odd-numbered columns include prior test scores in math and reading. Even-numbered columns include prior test scores in math, reading, and science. Heteroskedasticity-robust standard errors are reported in parentheses below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: $^*p < .05$ or $^\wedge p < .10$.

The first row of Table 3 reveals that, overall, 2 years of attendance at the STEM schools had a negative impact on student achievement in 10th grade. This negative impact is more pronounced when models include students' eighth-grade test scores in science[13]—and the negative impact is substantively significant in the non-STEM fields, for which STEM school attendance is associated with a disadvantage of around 0.11 standard deviations in reading and 0.15 standard deviations in social studies. As the propensity score model results in Appendix A reveal, students with higher test scores in science are more likely to attend five of the six STEM schools—all else held equal—so accounting for prior science scores addresses selection bias in models of STEM school achievement.

Table 3 also reveals significant heterogeneity across schools. For example, independent STEM schools have null effects in math and reading (which appear positive when models include lags only for math and reading scores), a positive impact in science, and a negative impact on social studies. District STEM schools, however, have substantively significant negative impacts across all subjects. Finally, the schools within these two categories also have impacts that vary significantly. For example, one independent STEM school has null effects across all subjects, whereas the other has a positive impact on science achievement (0.15 standard deviations) and a negative impact on achievement in social studies (−0.09 standard deviations). Additionally, the substantial negative impact of district STEM schools is driven in large part by two schools—Schools C and D. Indeed, district STEM School E has a positive impact on science achievement, but like the independent STEM school with such an impact, this impact is countered by a negative impact in social studies.

There are a couple of general takeaways. First, when one accounts for students' prior test scores in science, only two schools have a positive impact on achievement—both in science only—and this significant impact in science appears to come at the expense of achievement in social studies. Second, the results indicate that two of the four district STEM schools have significant negative impacts.

OLS Models Estimated With Matched Samples

Table 4 presents the results of student growth models estimated with the matched samples. For each subject, we present the results of models estimated via 1:1 "nearest neighbor" and 1:N radius propensity score–matching methods. Once

TABLE 3
Estimates of STEM School Effects: Unmatched Sample

	Math		Science		Reading		Social Studies	
	Lag: M, R	Lag: M, R, S	Lag: M, R	Lag: M, R, S	Lag: M, R	Lag: M, R, S	Lag: M, R	Lag: M, R, S
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All STEM schools	**−0.03**^	**−0.06**[*]	−0.01	**−0.06**[*]	**−0.07**[*]	**−0.11**[*]	**−0.11**[*]	**−0.15**[*]
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Independent STEM	**0.08**[*]	0.04	**0.11**[*]	**0.04**^	**0.07**[*]	0.00	0.00	**−0.06**[*]
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
District STEM	**−0.14**[*]	**−0.15**[*]	**−0.13**[*]	**−0.14**[*]	**−0.19**[*]	**−0.20**[*]	**−0.21**[*]	**−0.22**[*]
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
A. Independent STEM	**0.09**[*]	0.04	**0.05**^	−0.01	**0.06**[*]	0.00	0.00	−0.04
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
B. Independent STEM	**0.07**[*]	0.02	**0.29**[*]	**0.15**[*]	**0.07**^	0.01	0.00	**−0.09**[*]
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
C. District STEM	**−0.23**[*]	**−0.23**[*]	**−0.27**[*]	**−0.26**[*]	**−0.30**[*]	**−0.29**[*]	**−0.26**[*]	**−0.25**[*]
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
D. District STEM	**−0.11**[*]	**−0.12**[*]	**−0.11**[*]	**−0.14**[*]	**−0.17**[*]	**−0.19**[*]	**−0.16**[*]	**−0.19**[*]
	(0.04)	(0.04)	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)	(0.04)
E. District STEM	−0.04	−0.05	**0.13**[*]	**0.09**[*]	−0.05	−0.08	**−0.21**[*]	**−0.25**[*]
	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
F. District STEM	0.07	0.05	−0.05	−0.10	−0.01	−0.03	−0.07	−0.10
	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.08)	(0.07)
Students, *n*								
Treated	1,140	1,085	1,140	1,085	1,140	1,085	1,140	1,085
Control	64,316	56,907	64,316	56,907	64,316	56,907	64,316	56,907
Total	65,456	57,992	65,456	57,992	65,456	57,992	65,456	57,992

Note. Each column presents the estimated treatment effects from three separate ordinary least squares regressions: all science, technology, engineering, and mathematics (STEM) schools; disaggregated by STEM school type; and disaggregated by school. Odd-numbered columns include prior test scores in math (M) and reading (R); even-numbered columns include prior test scores in math, reading, and science (S). Robust standard errors are reported below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: [*]$p < .05$. ^$p < .10$.

again, each column presents the estimated effects from three separate regressions: "all STEM schools," "disaggregated by STEM school type," and "disaggregated by school." Unlike the previous table, however, all models account for prior test scores in math, reading, and science; the regressions are weighted to account for matching with replacement; and nonrobust standard errors are reported below the regression coefficients.

Table 4 reveals that models estimated with the matched samples yield results similar to those presented in Table 3. That the impacts are broadly similar lends us additional confidence that the results of the models based on the larger, unmatched sample are valid.

OLS Models by Race and Gender

The six STEM platform schools in this study are inclusive high schools that, to some extent, are meant to address STEM interest, achievement, and attainment gaps by race and gender. Although the descriptive statistics in Table 2 (and the propensity score model in Appendix A) clearly indicate that students who do well on science exams (and who are therefore less likely to be female, African American, or Hispanic) are more likely to enroll, the geographic locations and inclusive policies of these schools seem to expand access to a STEM-focused curriculum. Given the purpose of these STEM schools, it may also be that their impact varies by gender and race. In Tables 5 and 6, we present the results of the OLS models based on the unmatched samples (as in Table 3) but restricted to male, female, Caucasian, or African American student samples. (There are too few Hispanic students to estimate these models.) Once again, all models account for prior test scores in math, reading, and science, and robust standard errors are reported below the regression coefficients.

Table 5 reveals that for math and science, the negative STEM school effects had a disproportionate impact on African American students. The results for Caucasian

TABLE 4

Estimates of STEM School Effects: Matched Sample

	Math		Science		Reading		Social Studies	
	1:1 NN	Radius	1:1 NN	Radius	1:1 NN	Radius	1:1 NN	Radius
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All STEM Schools	**−0.08***	**−0.07***	**−0.06***	**−0.07***	**−0.07***	**−0.10***	**−0.15***	**−0.17***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Independent STEM	0.00	0.01	0.01	0.01	0.01	−0.02	−0.07	**−0.08***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
District STEM	**−0.13***	**−0.13***	**−0.11***	**−0.12***	**−0.13***	**−0.16***	**−0.21***	**−0.24***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
A. Independent STEM	0.00	0.01	−0.04	−0.05	0.01	−0.04	−0.04	−0.07
	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
B. Independent STEM	0.03	0.04	**0.16***	**0.17***	−0.02	0.03	**−0.14**^	−0.12
	(0.06)	(0.08)	(0.08)	(0.07)	(0.07)	0.07	(0.07)	(0.09)
C. District STEM	**−0.22***	**−0.21***	**−0.26***	**−0.25***	**−0.27***	**−0.29***	**−0.27***	**−0.25***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
D. District STEM	**−0.15***	**−0.17***	**−0.22***	**−0.17***	**−0.22***	**−0.21***	**−0.22***	**−0.22***
	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
E. District STEM	−0.03	−0.07	**0.15***	**0.09**^	0.06	0.01	**−0.19***	**−0.27***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
F. District STEM	0.02	0.03	−0.06	−0.12	0.03	−0.01	−0.01	−0.08
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)
Students (weighted), *n*								
Treated	1,051	1,051	1,051	1,051	1,051	1,051	1,051	1,051
Control	1,051	1,051	1,051	1,051	1,051	1,051	1,051	1,051
Total	2,102	2,102	2,102	2,102	2,102	2,102	2,102	2,102

Note. The table presents the estimated science, technology, engineering, and mathematics (STEM) school effects from ordinary least squares regressions with samples based on the 1:1 "nearest neighbor" (NN) matching methods (odd-numbered columns) and radius matching methods (even-numbered columns). Each column presents the estimated effects from three separate regressions: all STEM schools, disaggregated by STEM school type, and disaggregated by school. All models account for prior test scores in math, reading, and science. Standard errors appear in parentheses below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: *$p < .05$. ^$p < .10$.

students are null. Males appear to be more negatively affected in math than women, but both genders experienced negative impacts in science. As Table 6 illustrates, the results for reading are similar to those for science, but it appears that STEM school attendance also harms the achievement of Caucasian students in social studies—though to a lesser extent than it harms African American students. One reason for this differential impact by race appears to be that the very worst, district-run STEM schools disproportionately serve African American students.

OLS Models With Time Trends

The estimates presented above to a significant extent are based on student performance data from these schools' early years of operation. Indeed, if it takes 3 to 5 years for schools to get established (e.g., see Bifulco & Ladd 2006; Hanushek et al. 2007; Sass 2006; Zimmer et al., 2009), then only two schools were "established" when the latest cohorts that we analyze took the 10th-grade tests at the end of the 2012–2013 school year. To examine whether the negative results that we unearthed are attributable to these schools' growing pains, we estimated linear time trends in STEM school performance for each of the four subjects. For each STEM school, the time trend variable takes the value of zero in the year in which the school's first cohort entered 10th grade and increases by one in each subsequent year. School F is omitted because only a single cohort was observed.

Table 7 presents the results for our time trend analysis. Among the STEM subjects, there are no statistically significant trends. Outside the STEM subjects, there is a negative trend in reading at School A and negative trends in social studies at Schools A and E. The only positive trend is observed at School C in social studies, but its

TABLE 5

Estimates of STEM School Effects by Gender and Race: Unmatched Sample

	Math				Science			
	Male	Female	White	Black	Male	Female	White	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All STEM schools	**−0.09**[*]	−0.03	−0.01	**−0.11**[*]	**−0.04**[*]	**−0.08**[*]	0.02	**−0.13**[*]
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Independent STEM	0.02	0.06	0.02	0.08	0.05	0.02	0.03	**0.08**[^]
	(0.04)	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.05)
District STEM	**−0.20**[*]	**−0.10**[*]	**−0.09**[^]	**−0.17**[*]	**−0.13**[*]	**−0.15**[*]	0.01	**−0.20**[*]
	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)
A. Independent STEM	0.02	0.07	0.02	**0.11**[*]	−0.02	0.00	−0.03	0.05
	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)	(0.05)
B. Independent STEM	0.01	0.05	0.03	0.01	**0.18**[*]	0.09	**0.12**[*]	**0.17**[^]
	(0.05)	(0.07)	(0.05)	(0.09)	(0.06)	(0.07)	(0.05)	(0.10)
C. District STEM	**−0.31**[*]	**−0.17**[*]	**−0.52**[*]	**−0.21**[*]	**−0.27**[*]	**−0.25**[*]	**−0.31**[^]	**−0.25**[*]
	(0.06)	(0.05)	(0.19)	(0.04)	(0.06)	(0.04)	(0.18)	(0.04)
D. District STEM	**−0.15**	**−0.10**[^]	0.00	**−0.14**[*]	**−0.16**[*]	**−0.12**[*]	0.04	**−0.18**[*]
	(0.05)	(0.05)	(0.11)	(0.04)	(0.07)	(0.06)	(0.11)	(0.06)
E. District STEM	**−0.10**[*]	0.05	−0.06	−0.10	0.09	**0.12**[^]	**0.10**[^]	0.03
	(0.05)	(0.07)	(0.05)	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)
F. District STEM	0.04	0.04	0.06	−0.04	0.05	**−0.15**[*]	−0.10	−0.09
	(0.12)	(0.08)	(0.12)	(0.07)	(0.14)	(0.07)	(0.10)	(0.09)
Students, *n*								
Treated	564	521	441	524	564	521	441	524
Control	26,104	21,748	29,305	16,626	26,104	21,748	29,305	16,626
Total	26,668	22,269	29,746	17,150	26,668	22,269	29,746	17,150

Note. Each column presents the estimated effects from three separate ordinary least squares regressions: all science, technology, engineering, and mathematics (STEM) schools; disaggregated by STEM school type; and disaggregated by school. The regressions for each subject are restricted to samples of male, female, White, or Black students. All models account for prior test scores in math, reading, and science. Robust standard errors are report below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: [*]$p < .05$. [^]$p < .10$.

TABLE 6

Estimates of STEM School Effects by Gender and Race: Unmatched Sample

	Reading				Social Studies			
	Male	Female	White	Black	Male	Female	White	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All STEM Schools	**−0.11**[*]	**−0.10**[*]	−0.02	**−0.19**[*]	**−0.13**[*]	**−0.16**[*]	**−0.12**[*]	**−0.20**[*]
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Independent STEM	0.00	0.00	0.02	−0.04	−0.06	−0.06	−0.07	**−0.12**[*]
	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
District STEM	**−0.23**[*]	**−0.17**[*]	**−0.12**[*]	**−0.24**[*]	**−0.21**[*]	**−0.23**[*]	**−0.24**[*]	**−0.23**[*]
	(0.04)	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)
A. Independent STEM	−0.01	0.00	0.02	−0.06	−0.06	−0.01	−0.05	**−0.09**[^]
	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
B. Independent STEM	0.02	0.01	0.02	0.01	−0.05	**−0.19**[*]	**−0.11**[*]	**−0.20**[*]
	(0.05)	(0.07)	(0.05)	(0.07)	(0.06)	(0.08)	(0.06)	(0.09)

(continued)

TABLE 6 (CONTINUED)

	Reading				Social Studies			
	Male	Female	White	Black	Male	Female	White	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
C. District STEM	**−0.38**[*]	**−0.22**[*]	**−0.45**[*]	**−0.27**[*]	**−0.25**[*]	**−0.24**[*]	**−0.34**[^]	**−0.24**[*]
	(0.06)	(0.05)	(0.19)	(0.04)	(0.06)	(0.06)	(0.19)	(0.04)
D. District STEM	**−0.16**[*]	**−0.22**[*]	0.03	**−0.26**[*]	**−0.19**[*]	**−0.18**[*]	−0.07	**−0.22**[*]
	(0.07)	(0.10)	(0.15)	(0.07)	(0.06)	(0.06)	(0.09)	(0.05)
E. District STEM	**−0.11**[*]	0.02	**−0.11**[^]	−0.07	**−0.23**[*]	**−0.28**[*]	**−0.29**[*]	**−0.26**[*]
	(0.06)	(0.10)	(0.06)	(0.08)	(0.06)	(0.09)	(0.06)	(0.07)
F. District STEM	0.05	−0.07	−0.03	−0.04	0.17	**−0.16**[*]	−0.16	−0.02
	(0.12)	(0.08)	(0.11)	(0.10)	(0.15)	(0.08)	(0.11)	(0.10)
Students, *n*								
Treated	564	521	441	524	564	521	441	425
Control	26,104	21,748	29,305	16,626	26,104	21,748	29,305	14,078
Total	26,668	22,269	29,746	17,150	26,668	22,269	29,746	14,503

Note. Each column presents the estimated effects from three separate ordinary least squares regressions: all science, technology, engineering, and mathematics (STEM) schools; disaggregated by STEM school type; and disaggregated by school. The regressions for each subject are restricted to samples of male, female, White, or Black students. All models account for prior test scores in math, reading, and science. Robust standard errors are report below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: [*]*p* < .05. [^]*p* < .10.

TABLE 7
Estimates of STEM School Trends Over Time

	Math	Science	Reading	Social Studies
A. Independent STEM	0.02	−0.04	**0.15**[*]	**0.15**[^]
	(0.07)	(0.06)	(0.07)	(0.08)
A × Time Trend	0.01	0.01	**−0.05**[*]	**−0.07**[*]
	(0.02)	(0.02)	(0.02)	(0.02)
B. Independent STEM	−0.03	0.10	0.04	−0.11
	(0.07)	(0.07)	(0.06)	(0.07)
B × Time Trend	0.06	0.05	−0.03	0.03
	(0.05)	(0.05)	(0.04)	(0.05)
C. District STEM	**−0.30**[*]	**−0.18**[*]	**−0.28**[*]	**−0.39**[*]
	(0.06)	(0.06)	(0.08)	(0.08)
C × Time Trend	0.07	−0.07	−0.01	**0.13**[*]
	(0.05)	(0.05)	(0.05)	(0.06)
D. District STEM	0.01	−0.07	0.02	**−0.17**[*]
	(0.09)	(0.09)	(0.13)	(0.08)
D × Time Trend	−0.06	−0.04	**−0.11**[^]	−0.01
	(0.04)	(0.04)	(0.06)	(0.04)
E. District STEM	0.00	0.06	−0.15	**−0.18**[*]
	(0.06)	(0.05)	(0.05)	(0.05)
E × Time Trend	−0.10	0.05	0.14	−0.12
	(0.08)	(0.08)	(0.09)	(0.09)
Students, *n*				
Treated	1,022	1,022	1,022	1,022
Control	56,907	56,907	56,907	56,907
Total	57,929	57,929	57,929	57,929

Note. This table presents estimated science, technology, engineering, and mathematics (STEM) school and time trend effects from ordinary least squares regressions. Each column presents the effects from a single regression disaggregated by school. Robust standard errors are reported below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: [*]*p* < .05. [^]*p* < .10.

social studies scores remain very low relative to the comparison group.[14] Given that a linear trend is a strong assumption, we also estimated models with school by cohort dummy variables to examine trends over time, but no additional trends were discernable. These results are presented in Table C1. These results do not preclude these schools from improving going forward, but for the five schools with multiple cohorts—particularly the established independent STEM schools—it is clear that the results of this study are not due to typical achievement trajectories for newly established schools.

Two-Stage Least Squares Models for School A

School A is the only school from which we could obtain admission data. The school provided sufficiently informative lottery data for three student cohorts. Table 8 presents the second-stage results of the two-stage least squares models that include all covariates but prior science scores so that we can include the first cohort in the analysis. As the table indicates, the coefficient estimates for math and science are somewhat more positive, and the estimate for social studies is somewhat less negative than the estimates in Table 3 that account for prior science scores. However, the coefficient for reading achievement is quite a bit larger in magnitude. Unfortunately, the models lack statistical power, as standard errors are often 3 times as large as those in Table 3. Thus, we cannot be sure that the null results are due to there being no substantively significant effects. That the coefficients are roughly comparable to those in Table 3 is nonetheless reassuring.

TABLE 8

Two-Stage Least Squares Estimates of Attending STEM School A

	Math	Science	Reading	Social Studies
	(1)	(2)	(3)	(4)
STEM school attendance	0.05 (0.11)	0.02 (0.11)	−0.11 (0.12)	−0.03 (0.13)
n	212	211	211	211
Wald χ^2	462.62[*]	400.24[*]	285.60[*]	279.09[*]
R^2	0.69	0.65	0.57	0.57
Applicants, n	212	211	211	211
Lottery winners, n	153	152	153	152
Attendees (2 years), n	123	122	123	122

Note. The table presents two-stage least squares estimates of the effect of attending School A on academic achievement. The model is based on only three of the six cohorts. Standard errors are reported in parentheses. [*]$p < .05$ (two-tailed test). STEM = science, technology, engineering, and mathematics.

Conclusion

The results suggest that, overall, the first 2 years of attendance at Ohio's STEM platform high schools have had a negative impact on student achievement in STEM and, especially, non-STEM subjects. There is considerable heterogeneity in effects across the six schools, but only two schools can claim to boost achievement in any subject (science). Generally, student achievement suffered most in non-STEM subjects and among African Americans. One should not evaluate these schools solely on the basis of 10th-grade achievement tests, of course. These schools focus on problem-solving skills and group work that standardized exams—particularly Ohio's graduation tests—may not capture, for example. Indeed, if some of these schools simply help foster student interest and attainment in STEM fields (via early college opportunities, for example), they may very well achieve their goals of improving and broadening the STEM workforce pipeline in spite of comparable or lower student achievement.

However, promoters of these schools have long claimed that they improve achievement, citing student test scores as evidence that they help traditionally underserved populations beat the odds. Despite the limitations of test scores discussed above, standardized tests—particularly Ohio's graduation tests—do provide a useful barometer for gauging knowledge of basic concepts and principles. Regardless of other possible benefits of Ohio's STEM schools, it is clear that the STEM schools studied here do not, on average, improve achievement in these basic content areas. The consequences of this might be particularly significant because the purpose of these platform schools is to identify best practices and disseminate them.

These results may not be all that surprising. The Ohio STEM schools studied here, like their counterparts across the nation, may not emphasize math and science content as much as observers assume. They tend to focus on problem-based and individualized learning, which does not inherently favor STEM fields. Moreover, although these schools have received state and private funds as part of STEM initiatives, their student counts and student-teacher ratios are highly variable and do not suggest a clear advantage in terms of instructional resources. Indeed, some have very high student-teacher ratios as compared with non-STEM public high schools. It is perhaps unsurprising, therefore, that the two schools that have a positive impact on science achievement are also associated with a negative impact in social studies achievement.

The results presented here force us to reconsider both the definition of a high-quality STEM education and the ability of contemporary standardized tests to evaluate STEM education quality. Additionally, they illustrate how an analysis of achievement growth can lead to mistaken impressions if researchers account only for one prior year of reading and math achievement, as they often do. Accounting for prior science achievement appears to be important for controlling for self-selection into STEM schools, and studies that fail to do so may overestimate STEM school impacts.

Appendix A: Propensity Score Models

We estimated separate propensity score models for each school. Table A1 presents the results of these models. The sample in each model includes all students in the unmatched regression analysis, except for those with characteristics that perfectly predict non-STEM school attendance. As the results indicate, the characteristics that predict selection into treatment, as well as the nature of the relationships between particular characteristics and probability of treatment, vary by school.

TABLE A1

Results from Propensity Score Models Estimating Probability of STEM School Attendance

	School A[a]	School B	School C	School D	School E	School F
Eighth-grade math score	**0.19***	−0.07	−0.03	0.09	0.20	−0.36
	(0.05)	(0.07)	(0.06)	(0.06)	(0.13)	(0.21)
Eighth-grade reading score	**0.11***	0.06	−0.09	0.04	−0.07	−0.21
	(0.05)	(0.07)	(0.07)	(0.07)	(0.12)	(0.21)
Eighth-grade science score	**0.15***	**0.45***	−0.11	**0.12**$^\wedge$	**0.26***	**0.57***
	(0.05)	(0.10)	(0.09)	(0.07)	(0.15)	(0.26)
(Eighth-grade math score)^2	**−0.07**$^\wedge$	0.05	0.02	−0.07	0.13	−0.28
	(0.04)	(0.05)	(0.06)	(0.06)	(0.11)	(0.25)
(Eighth-grade reading score)^2	**0.05***	0.02	0.00	−0.03	−0.04	**−0.73***
	(0.03)	(0.05)	(0.06)	(0.07)	(0.11)	(0.27)
(Eighth-grade science score)^2	−0.05	0.04	−0.05	−0.08	0.03	−0.50
	(0.05)	(0.06)	(0.07)	(0.07)	(0.16)	(0.31)
Eighth-Grade Math × Eighth-Grade Read	−0.08	−0.05	**−0.17**$^\wedge$	0.01	−0.12	−0.28
	(0.06)	(0.10)	(0.09)	(0.09)	(0.18)	(0.39)
Eighth-Grade Math × Eighth-Grade Science	**0.19***	−0.06	0.06	0.13	−0.13	0.84
	(0.08)	(0.09)	(0.11)	(0.11)	(0.24)	(0.54)
Eighth-Grade Read × Eighth-Grade Science	−0.10	−0.10	−0.03	−0.04	0.21	0.98
	(0.06)	(0.09)	(0.10)	(0.11)	(0.23)	(0.48)
Female	−0.09	**−0.56***	−0.34	**−0.46**$^\wedge$	−0.36	**10.46***
	(0.13)	(0.24)	(0.27)	(0.26)	(0.38)	(0.80)
African American	−0.03	**0.64***	**0.41***	0.22	−0.10	0.40
	(0.12)	(0.18)	(0.17)	(0.27)	(0.21)	(0.45)
Hispanic	0.14	X	0.31	−0.02	−0.19	**−8.80***
	(0.18)		(0.41)	(0.45)	(0.54)	(1.24)
Other non-White	**0.33***	**0.55***	0.31	0.10	−0.20	**−9.57***
	(0.11)	(0.15)	(0.27)	(0.45)	(0.32)	(0.87)
Subsidized lunch	**−0.34***	**−0.61**$^\wedge$	0.25	−0.32	−0.03	0.71
	(0.16)	(0.27)	(0.28)	(0.37)	(0.36)	(0.82)
Female × Caucasian	−0.03	0.24	0.18	−0.11	−0.24	**−9.68***
	(0.14)	(0.25)	(0.31)	(0.23)	(0.39)	(0.96)

(continued)

15

TABLE A1 (CONTINUED)

	School A[a]	School B	School C	School D	School E	School F
Female × African American	0.16	0.25	0.27	0.15	0.04	**−9.22***
	(0.15)	(0.30)	(0.26)	(0.18)	(0.40)	(0.90)
Subsidized Lunch × Caucasian	0.26	0.39	0.06	0.08	−0.33	0.20
	(0.16)	(0.29)	(0.32)	(0.45)	(0.39)	(0.78)
Subsidized Lunch × African American	0.00	−0.26	0.08	−0.16	−0.40	−0.82
	(0.17)	(0.34)	(0.29)	(0.39)	(0.40)	(0.77)
Female × Subsidized Lunch	0.01	0.21	0.13	0.17	0.28	−0.12
	(0.10)	(0.18)	(0.14)	(0.21)	(0.27)	(0.44)
Limited English proficiency	0.21	X	0.26	−0.05	−0.11	**0.86**^
	(0.14)		(0.20)	(0.18)	(0.36)	(0.47)
Disability	0.05	0.12	0.03	−0.13	−0.24	−0.37
	(0.10)	(0.14)	(0.08)	(0.11)	(0.22)	(0.41)
Gifted	**0.18***	−0.26	**−0.17**^	0.06	0.06	−0.01
	(0.06)	(0.11)	(0.09)	(0.08)	(0.17)	(0.29)
Attend charter in eighth grade	**0.56***	**1.78***	0.14	**0.43***	X	**5.97***
	(0.13)	(0.15)	(0.14)	(0.12)		(0.46)
Pseudo R^2	0.17	0.25	0.08	0.06	0.12	0.20
p value for test of joint significance	0.00	0.00	0.00	0.00	0.00	0.00
Cohort-district fixed effects	52	49	3	4	2	X
Students used, n						
Treated	317	149	411	226	195	65
Control	36,211	12,125	3,488	7,572	334	190
Total	36,528	12,274	3,899	7,798	529	255

Note. X indicates a variable that is off the common support. Robust standard errors in parentheses. Bold indicates significance: *$p < .05$. ^$p < .10$. STEM = science, technology, engineering, and mathematics.
[a]Cohorts 2–6.

TABLE B1
Sample Balance

Note: In the control columns (C, C 1:1, C R) each cell shows the mean with the standardized percent bias in parentheses below it; bold indicates the bias value as printed.

	School A					School B					School C				
	Unmatched		Matched			Unmatched		Matched			Unmatched		Matched		
	T	C	T	C 1:1	C R	T	C	T	C 1:1	C R	T	C	T	C 1:1	C R
Eighth-grade math score	0.62	0.05 (**59.4**)	0.61	0.61 (−0.5)	0.48 (**12.7**)	0.41	0.15 (**26.8**)	0.38	0.39 (−0.6)	0.27 (**11.6**)	−0.61	−0.14 (**−59.4**)	−0.61	−0.50 (**−13.1**)	−0.52 (**−10.7**)
Eighth-grade reading score	0.60	0.03 (**60.0**)	0.58	0.60 (−1.6)	0.47 (**12.5**)	0.41	0.14 (**30.0**)	0.40	0.33 (7.5)	0.29 (**11.9**)	−0.52	−0.06 (**−52.9**)	−0.52	−0.42 (**−11.2**)	−0.45 (−7.0)
Eighth-grade science score	0.58	0.00 (**60.6**)	0.56	0.56 (0.2)	0.43 (**13.5**)	0.59	0.09 (**53.2**)	0.57	0.48 (**10.3**)	0.34 (**24.9**)	−0.78	−0.28 (**−61.7**)	−0.78	−0.66 (**−15.4**)	−0.72 (−7.4)
Female	0.46	0.50 (−9.5)	0.47	0.44 (5.3)	0.49 (−4.9)	0.34	0.51 (**−35.2**)	0.35	0.33 (3.3)	0.44 (**−19.8**)	0.61	0.55 (**11.5**)	0.60	0.56 (9.5)	0.57 (6.1)
African American	0.28	0.29 (−0.7)	0.29	0.30 (−1.5)	0.31 (−5.1)	0.20	0.17 (7.9)	0.23	0.21 (4.1)	0.22 (0.6)	0.92	0.65 (**71.0**)	0.92	0.87 (**13.1**)	0.91 (3.6)
Hispanic	0.04	0.04 (0.3)	0.04	0.04 (0.0)	0.04 (1.2)	0.00	0.02 (**−18.9**)	0.00	0.00 (0.0)	0.00 (0.0)	0.01	0.01 (0.5)	0.01	0.02 (−7.4)	0.01 (1.0)
Other non-White	0.11	0.07 (**16.6**)	0.11	0.13 (−8.1)	0.11 (−0.6)	0.12	0.07 (**18.0**)	0.10	0.12 (−5.5)	0.08 (9.0)	0.02	0.06 (**−22.6**)	0.02	0.04 (**−14.7**)	0.03 (−7.3)
Subsidized lunch	0.42	0.40 (4.1)	0.43	0.44 (−1.3)	0.45 (−3.4)	0.27	0.32 (**−10.5**)	0.25	0.25 (0.00)	0.31 (**−13.5**)	0.86	0.57 (**69.3**)	0.86	0.87 (−2.8)	0.86 (0.5)
Limited English proficiency	0.04	0.05 (−5.4)	0.04	0.03 (0.0)	0.05 (−6.2)	0.00	0.01 (**−14.7**)	0.00	0.00 (0.0)	0.00 (0.0)	0.04	0.02 (**11.3**)	0.03	0.05 (**−12.3**)	0.03 (1.2)
Disability	0.05	0.11 (**−22.7**)	0.05	0.04 (3.6)	0.007 (−5.6)	0.08	0.11 (**−11.1**)	0.09	0.09 (0.0)	0.09 (−0.2)	0.23	0.16 (**19.0**)	0.24	0.22 (4.9)	0.21 (6.9)
Gifted	0.64	0.33 (**63.6**)	0.64	0.67 (−6.3)	0.58 (**11.3**)	0.18	0.25 (**−17.6**)	0.14	0.14 (**15.8**)	0.19 (1.9)	0.07	0.24 (**−47.1**)	0.07	0.07 (0.0)	0.09 (−4.7)
Attend charter in eighth grade	0.05	0.01 (**20.4**)	0.05	0.03 (8.0)	0.04 (−1.3)	0.30	0.02 (**82.1**)	0.21	0.23 (−4.7)	0.19 (7.2)	0.04	0.04 (−0.2)	0.04	0.04 (2.0)	0.04 (0.3)
Mean standardized bias (abs)		**12.8**		0.7	1.3		**12.6**		0.9	1.7		**31.8**		7.6	4.0
n (weighted for matched)	306	34,722	301	301	301	142	11,449	124	124	124	256	3,322	255	255	255

TABLE B1 (CONTINUED)

	School D					School E					School F				
Eighth-grade math score	−0.35	−0.65	−0.35	−0.32	−0.34	0.60	0.01	0.58	0.62	0.57	0.01	−0.03	−0.01	−0.02	−0.08
		36.0		−3.8	−2.2		**66.5**		−4.9	1.8		4.6		0.8	7.8
Eighth-grade reading score	−0.26	−0.56	−0.27	−0.26	−0.22	0.50	0.06	0.47	0.42	0.45	0.13	0.03	0.15	0.09	0.04
		34.5		−1.0	−5.0		**48.5**		6.0	2.7		**10.9**		6.8	**12.6**
Eighth-grade science score	−0.46	−0.81	−0.45	−0.52	−0.45	0.78	0.16	0.76	0.76	0.76	0.24	0.16	0.26	0.16	0.18
		44.5		8.0	−0.3		**71.3**		−0.2	−0.5		8.6		**10.1**	7.9
Female	0.45	0.53	0.45	0.44	0.48	0.36	0.51	0.37	0.35	0.38	0.83	0.54	0.81	0.81	0.82
		−17.3		3.1	−5.3		**−30.1**		3.3	−1.6		**63.3**		0.0	−1.1
African American	0.73	0.71	0.73	0.72	0.71	0.26	0.39	0.26	0.23	0.28	0.46	0.37	0.45	0.45	0.43
		4.9		1.7	2.7		**−29.0**		5.8	−4.6		**17.6**		0.0	4.6
Hispanic	0.12	0.12	0.12	0.07	0.10	0.02	0.02	0.02	0.00	0.01	0.02	0.02	0.00	0.02	0.01
		0.0		**14.6**	6.8		0.2		**13.0**	2.1		−0.5		**−13.6**	−6.6
Other non-White	0.03	0.03	0.03	0.03	0.04	0.11	0.11	0.10	0.13	0.12	0.10	0.13	0.10	0.02	0.10
		−1.3		0.0	−4.1		−1.2		−8.7	−4.6		**−11.5**		**27.1**	2.5
Subsidized lunch	0.95	0.95	0.96	0.95	0.96	0.27	0.40	0.27	0.25	0.25	0.49	0.40	0.50	0.55	0.50
		−1.7		7.0	−0.1		**−28.5**		3.5	4.1		**13.9**		**−10.4**	−0.7
Limited English proficiency	0.04	0.05	0.04	0.00	0.03	0.03	0.04	0.03	0.01	0.03	0.06	0.04	0.03	0.05	0.07
		−5.9		0.0	4.3		−7.9		9.0	−4.1		8.6		−7.6	**−17.3**
Disability	0.11	0.18	0.11	0.09	0.10	0.05	0.13	0.05	0.05	0.06	0.05	0.13	0.05	0.05	0.05
		−20.9		**18.9**	3.3		**−27.4**		1.9	−1.7		**−28.1**		0.0	1.9
Gifted	0.28	0.14	0.29	0.29	0.24	0.40	0.18	0.39	0.35	0.38	0.27	0.22	0.26	0.29	0.23
		37.1		0.0	**12.8**		**49.0**		8.6	2.4		**11.6**		−8.0	5.5
Attend charter in eighth grade	0.11	0.04	0.09	0.09	0.08	0.00	0.01	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.0003
		28.1		4.4	3.6		**−11.2**		0.0	0.0		10.0		0.0	−0.3
Mean standardized bias (abs)		**19.1**		3.0	3.4		**28.7**		5.0	2.3		**16.2**		7.0	5.7
n (weighted for matched)	130	7,096	128	128	128	188	318	185	185	185	63	182	58	58	58

Note. Bolded values indicate a standardized bias above an absolute value of 10 (the threshold above which some researchers consider bias to be elevated). Abs = absolute; C = control; R = radius; T = treated.

TABLE C1

Estimates of STEM School Effects Over Time

	Math	Science	Reading	Social Studies
A: Cohort 2	0.10	0.06	0.05	0.02
	(0.07)	(0.06)	(0.07)	(0.07)
A: Cohort 3	−0.03	**−0.14***	**0.13**^	**0.17***
	(0.07)	(0.06)	(0.07)	(0.08)
A: Cohort 4	−0.02	0.00	0.01	0.03
	(0.09)	(0.07)	(0.08)	(0.07)
A: Cohort 5	0.03	−0.09	**−0.15***	**−0.29***
	(0.08)	(0.07)	(0.07)	(0.06)
A: Cohort 6	**0.11**^	0.09	−0.10	**−0.17***
	(0.07)	(0.06)	(0.06)	(0.07)
B: Cohort 1	0.01	**0.18***	0.09	−0.08
	(0.07)	(0.07)	(0.07)	(0.08)
B: Cohort 2	−0.10	−0.06	−0.11	**−0.19***
	(0.08)	(0.07)	(0.07)	(0.05)
B: Cohort 3	**0.12**^	**0.28***	0.02	−0.03
	(0.07)	(0.07)	(0.06)	(0.07)
C: Cohort 1	**−0.30***	**−0.20***	**−0.39***	**−0.47***
	(0.06)	(0.07)	(0.09)	(0.10)
C: Cohort 2	**−0.24***	**−0.22***	**−0.15***	**−0.16***
	(0.06)	(0.05)	(0.07)	(0.06)
C: Cohort 3	**−0.16***	**−0.34***	**−0.38***	**−0.19***
	(0.07)	(0.06)	(0.07)	(0.06)
D: Cohort 1	−0.02	−0.01	0.16	−0.12
	(0.17)	(0.07)	(0.26)	(0.12)
D: Cohort 2	−0.04	−0.02	−0.10	**−0.20***
	(0.07)	(0.09)	(0.09)	(0.07)
D: Cohort 3	**−0.15***	**−0.35***	**−0.25***	**−0.19***
	(0.07)	(0.10)	(0.09)	(0.09)
D: Cohort 4	**−0.18***	−0.10	**−0.27***	**−0.19***
	(0.06)	(0.07)	(0.10)	(0.07)
E: Cohort 1	0.00	0.06	**−0.15***	**−0.18***
	(0.06)	(0.05)	(0.05)	(0.05)
E: Cohort 2	**−0.10**^	**0.11**^	−0.02	**−0.30***
	(0.05)	(0.06)	(0.08)	(0.07)
F: Cohort 1	0.03	−0.09	−0.15	**−0.12**^
	(0.07)	(0.06)	(0.05)	(0.07)
Students, *n*				
Treated	1,085	1,085	1,085	1,085
Control	56,907	56,907	56,907	56,907
Total	57,992	57,992	57,992	57,992

Note. This table presents estimated cohort fixed effects for each school from ordinary least squares regressions. Each column presents the effects from a single regression disaggregated by school. Robust standard errors are reported below the regression coefficients. Bolded coefficients are significant at the following levels for a two-tailed test: *p < .05. ^p < .10.

Appendix B: Procedure for Evaluating Balance

Simple *t* tests on the differences in mean characteristics between treatment and control groups can be misleading because the dissipation of significant differences that result from matching may simply be due to loss of statistical power rather than to improvements in sample balance. Thus, we examined balance by calculating standardized bias using the following formula:

$$\frac{\left(\bar{x}_t - \bar{x}_c\right)}{\sqrt{\dfrac{s_{tu}^2 + s_{cu}^2}{2}}},$$

where s_{tu}^2 is the sample variance of variable x in the unmatched treatment group and where s_{cu}^2 is the sample variance of variable x in the unmatched control group. The differences in means across treatment and control groups for variable x is given by $\left(\bar{x}_t - \bar{x}_c\right)$. We calculated this difference in means for both the unmatched and the matched samples, and the matched and unmatched mean differences are respectively divided by the same denominator, $\sqrt{\dfrac{s_{tu}^2 + s_{cu}^2}{2}}$.

This provides a standardized measure of covariate balance in units of pooled unmatched sample standard deviations, with zero indicating perfect balance and larger deviations from zero, either positive or negative, indicating larger imbalances. This measure is not influenced by sample size and can be compared across the matched and unmatched samples (Stuart, 2010).

Table B1 presents standardized bias calculations. Bolded figures indicate a standardized bias above an absolute value of 10: a threshold above which some researchers consider bias to be elevated. The postmatching sample includes only those treatment and control observations that were successfully matched. A significant drawback of evaluating balance via standardized bias is that, unlike *t* tests, there is no well-established threshold for distinguishing between acceptable and unacceptable levels of standardized bias. Nevertheless, Table B1 shows that both matching methods significantly reduce average absolute standardized bias across the covariates for each school. The overall mean of the absolute bias across all characteristics for each school go from a range of 13.1–30.8 for the unmatched samples to 1.9–16.0 for 1:1 matched samples and 2.2–10.9 for the radius-matched samples. Of particular importance, there are large reductions in the standardized bias across eighth-grade test scores.

authors alone are responsible for the design and execution of the analysis and for the contents of this article. Some of the results presented here will be disseminated via an OERC policy brief.

Notes

1. Young, House, Wang, Singleton, and Klopfenstein (2011) evaluated the impact of the first 31 inclusive science, technology, engineering, and mathematics (STEM) schools established by the Texas High School Project's T-STEM initiative. The authors employed school-level data and used statistical techniques to match STEM high schools to observationally similar traditional public high schools. Their regression analysis indicates that T-STEM high schools were associated with greater achievement in math and science than that of the matched traditional public schools.

2. Deming (2014) showed that bias is nonexistent if multiple years of prior test data are used but that there is some bias if only one prior year of test scores is used. But this may be due to his not accounting for other covariates in that specification. As Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011) and Angrist, Pathak, and Walters (2013) revealed in their study of charter schools, one baseline score may be sufficient if one also accounts for baseline demographics, as we do.

3. We excluded one district school because its implementation of a STEM program was difficult to determine and because the Ohio Auditor of State determined that it had manipulated its data.

4. For example, one STEM school is on a university campus. In addition to the building provided by the university, the school benefits from having direct access to various campus resources, such as faculty and their courses.

5. Our calculations are based on data publicly available on the Ohio Department of Education website.

6. Some have raised concerns about our using the results of the Ohio Graduation Test to examine student achievement growth, citing potentially insufficient "stretch" in the assessment that might impede our ability to estimate effects—particularly for students on the high end of the achievement distribution.

7. We omitted from the analysis the few students who attended a district STEM school by taking advantage of district open enrollment.

8. A full academic year for high school students during all years in the study period refers to continuous enrollment from the end of the first week in October through the administration of the Ohio Graduation Test in mid-March as defined in the Ohio Department of Education's annually published data reporting instructions, the Education Management Information System manuals (fiscal year 2007–2013 editions).

9. School C seems to have adopted the STEM model in an effort to turnaround a low-performing high school, so it is qualitatively different from the other schools in our study.

10. The Ohio Department of Education requires that a test score report be submitted for each enrolled student regardless of whether or not the student took the test. Test score reports for untested students should include a code indicating the reason why the student did not take the test (fiscal year 2007–2013 editions). The missing 10th-grade test scores at Schools C and D are largely due to missing test score reports rather than existing test score reports indicating untested students. A statewide audit of K−12 data reporting for the 2010–2011 school year found evidence of improper record keeping at nine school districts across Ohio (Yost, 2013).

11. We do not include math and reading test scores from prior grades, to protect our sample size. Additionally, we exclude prior scores on social studies tests because they were not administered for much of our panel. Finally, because our preferred specifications include eighth-grade test scores in science—and because the eighth-grade science test was not administered to the first cohort in our panel—we present results primarily from models estimated without the first year of data from School A. As we mention later, models that are not restricted in these ways yield similar results.

12. Grade 8 test scores in science are not available for the first cohort of one of the platform schools. To include this cohort in the analysis, we estimated a separate propensity score model that includes lagged test scores only in math and reading. However, to make the results that we present comparable across the various modeling approaches, we report the results of models that exclude this cohort from the analysis.

13. In results not presented here, we estimated models that control only for lagged math and reading test scores but that also exclude Cohort 1 from the sample. We find evidence that change in the estimated STEM school effects between the models that lag only math and reading scores and the models that also lag science scores are, in fact, driven by the inclusion of the additional lagged test scores and not by the exclusion of Cohort 1.

14. Given that School C is essentially a neighborhood school that converted to STEM school as part of what is ostensibly a turnaround effort, the trend analysis is particularly helpful because it shows that the STEM focus has not led to improvements over time.

References

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics*, *126*(2), 699–748.

Academic Competitiveness Council. (2007). *Report of the Academic Competitiveness Council*. Retrieved from http://eric.ed.gov/?id=ED496649

America COMPETES Act. 42 U.S.C. § 7381h (2007).

Angrist, J. D., & Imbens, G. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Angrist, J. D., Pathak, P. A., & Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics*, *5*(4), 1–27.

Angrist, J. D., & Pishke, J.-S. (2009). *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.

Atkinson, R. D., Hugo, J., Lundgren, D., Shapiro, M. J., & Thomas, J. (2007). *Addressing the STEM challenge by expanding specialty math and science high schools*. Retrieved from http://www.itif.org/publications/addressing-stem-challenge-expanding-specialty-math-and-science-high-schools

Bifulco, R., & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy*, *1*(1), 50–90.

Carnegie Corporation of New York. (2009). *The opportunity equation: Transforming mathematics and science education for citizenship and the global economy*. New York, NY: Author.

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics, Series A*, *35*(4), 417–446.

Deming, D. J. (2014). Using school choice lotteries to test measures of school effectiveness. *American Economic Review: Papers & Proceedings*, *104*(5), 406–411

Dobbie, W., & Fryer, R. G. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, *5*(4), 28–60.

Eilber, C. R. (1987). The North Carolina School of Science and Mathematics. *The Phi Delta Kappan*, *68*(10), 773–777.

Forston, K., Gleason, P., Kopa, E., & Verbitsky-Savitz, N. (2014). Horseshoes, hand grenades, and treatment effects? Reassessing whether nonexperimental estimators are biased. *Economics of Education Review*, *44*(1), 100–113.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., & Branch, G. F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics*, *91*(5–6), 823–848.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experiemntal data. *Econometrica*, *66*(5), 1017–1098.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies*, *64*(4), 605–654.

LaForce, M., Noble, E., King, H., Holt, S., & Century, J. (2014). *The 8 elements of inclusive STM high schools*. Chicago, IL: The University of Chicago.

Lynch, S. J., Behrend, T., Burton, E. P., & Means, B. (2013). *Inclusive STEM-focused high schools: STEM education policy and opportunity structures*. Retrieved from http://ospri.research.gwu.edu/conference-presentations

Means, B., Confrey, J., House, A., & Bhanot, R. (2008). *STEM high schools: Specialized science, technology, engineering, and mathematics high schools in the U.S.* Retrieved from http://www.sri.com/work/publications/stem-high-schools

Means, B., House, A., Young, V., & Wang, H. (2013). *Expanding access to STEM-focused education: What are the effects?* Retrieved from http://www.sri.com/work/publications/expanding-access-stem-focused-education-what-are-effects

National Academies. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Retrieved from http://www.nap.edu/catalog.php?record_id=11463

National Consortium for Specialized Secondary Schools of Mathematics, Science, and Technology. (2014). *Our history and founders*. Retrieved from http://www.ncsssmst.org/about-ncsssmst/our-history-and-founders

Ohio Department of Education. (2015). *STEM—science, technology, engineering and mathematics*. Retrieved from http://education.ohio.gov/Topics/Career-Tech/STEM

Ohio STEM Learning Network. (2012). *Ohio STEM advocacy kit*. Columbus, OH: Author.

Ohio STEM Learning Network. (2014). *OSLN design principles*. Retrieved from http://www.osln.org/about/promise-principles-practice/

President's Council of Advisors on Science and Technology. (2010). *Prepare and inspire: K−12 education in science, technology,* *engineering, and math (STEM) for America's future*. Retrieved from http://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sample methods that incorporate the propensity scores. *The American Statistician*, *39*(1), 33–38.

Sass, T. R. (2006). Charter schools and student achievement in Florida. *Education Finance and Policy*, *1*(1), 91–122.

STEM school grants—contents of proposals. 33 Ohio Rev. Code. § 3326.03 (2007).

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21.

Thomas, J., & Williams, C. (2010). The history of specialized STEM schools and the formation and role of NCSSSMST. *Roeper Review*, *32*, 17–24.

Tuttle, C. C., Teh, B.-R., Nicholas-Barrer, I., Gill, B. P., & Gleason, P. (2010). *Student characteristics and achievement in 22 KIPP middle schools*. Retrieved from http://www.mathematica-mpr.com/publications/redirect_PubsDB.asp?strSite=PDFs/education/KIPP_fnlrpt.pdf

Wiswall, M., Stiefel, L., Schwartz, A. E., & Boccardo, J. (2014). Does attending a STEM high school improve student performance? Evidence from New York City. *Economics of Education Review*, *40*, 93–105.

Woodworth, K. R., David, J. L., Guha, R., Haiwen, W., & Lopez-Torkos, A. (2008). *San Francisco Bay Area KIPP schools: A study of early implementation and achievement*. Retrieved from http://www.sri.com/work/publications/kipp-early-implementation-achievement-study

Yost, D. (2013). *Statewide audit of student attendance data and accountability system*. Retrieved from https://ohioauditor.gov/publications/Attendance_FINAL_2-11-13.pdf

Young, V., House, A., Wang, H., Singleton, C., & Klopfenstein, K. (2011). *Inclusive STEM schools: Early promise in Texas and unanswered questions*. Retrieved from http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_071087

Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T. R., & Witte, J. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition*. Washington, DC: RAND.

Authors

JENNIFER GNAGEY is an assistant professor in the John B. Goddard School of Business and Economics, Department of Economics, Weber State University, 1337 Edvalson Street, Ogden, UT 84408; jennifergnagey@weber.edu. Her research focuses on the economics of education.

STÉPHANE LAVERTU is an associate professor at the John Glenn College of Public Affairs at The Ohio State University, 110 Page Hall, 1810 College Road, Columbus, OH 43210; lavertu.1@osu.edu. His research focuses on school district governance, accountability systems, and the evaluation of education policy.