NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

AIR

Characteristics of Schools Successful in STEM: Evidence from Two States' Longitudinal Data

MICHAEL HANSEN

WORKING PAPER 97 • APRIL 2013

Characteristics of Schools Successful in STEM: Evidence from Two States' Longitudinal Data

Michael Hansen
American Institutes for Research

Contents

Acknowledgements

This work was generously supported with a research grant (#1059991) from the National Science Foundation. Excellent research assistance was provided by Habib Moody, Tommy Gonzalez, and Joel Mittleman. I acknowledge the Florida Education Data Warehouse and the North Carolina Education Research Data Center in providing access to the longitudinal data utilized in this study. The statements and views expressed in this report are those of the author and do not necessarily reflect the position of the National Science Foundation or any affiliated institutions.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Characteristics of Schools Successful in STEM: Evidence from Two States' Longitudinal Data

Michael Hansen

CALDER Working Paper No. 97
April 2013

Abstract

Current federal education policies promote learning in Science, Technology, Engineering, and Math (STEM) and the participation of minority students in these fields. Using longitudinal data on students in Florida and North Carolina, value-added estimates in math and science are generated to categorize schools into performance levels and identify differences in school STEM measures by performance levels. Several STEM-relevant variables show a significant association with effectiveness in math and science, including STEM teacher turnover, calculus and early algebra participation, and math and science instructional indices created from survey items in the data. Surprisingly, a negative association between students' STEM course participation and success in STEM is consistently documented across both states, in addition to low participation of underrepresented minority students in successful schools in STEM.

Keywords: STEM Learning, School Value Added, Underrepresented Minorities

Introduction

Promoting science, technology, engineering, and math (STEM) in public education is commonly viewed as a key strategy in maintaining America's competitiveness in the rapidly changing and increasingly global 21st century economy.[1] Federal education policymakers now show a keen interest in learning what programs and policies enable schools to obtain successful outcomes in STEM fields (National Research Council, 2011). Once these successful programs are identified, policymakers hope to scale them up to improve the quality of education in STEM fields and expand students' access to such programs. In spite of the policy aspirations, however, little empirical research has been conducted to document what successful schools in STEM fields look like, much less point to specific elements that enable their success.

This report estimates school effectiveness in science and math to identify and describe both successful and unsuccessful schools in STEM fields. Using longitudinal data from Florida and North Carolina, subject-specific value-added estimates are calculated to rank schools according to effectiveness in these fields. Once identified, this report documents school-level variables and STEM capacity measures at various points in the effectiveness distribution. The primary purpose of this analysis is to identify key variables that are associated with varying degrees of effectiveness in STEM fields to generate hypotheses on which schooling inputs may influence students' learning in STEM subjects. Although the analysis presented here is descriptive only, it aims to identify promising STEM policies that may be evaluated more formally in future research.

[1] For example, President Obama stressed this issue in his 2011 State of the Union address (retrieved from http://www.whitehouse.gov).

Background

The achievement of students in the United States in science and mathematics remains low in international comparisons, and stark racial achievement gaps persist.[2] Yet, interest in STEM appears to be gaining momentum; a 2008 survey report reveals the rapid growth of STEM schools in recent years—of 203 schools surveyed, the median year of opening was 2003 (Means Confrey, House, & Bhanot, 2008). Nonetheless, the concept of STEM is ill-defined, with the implementation details and programs goals of STEM-focused instruction varying widely across institutions. For instance, this same survey report highlights the distinction between "inclusive" STEM schools, which aim to serve a wide range of students, and "selective" STEM schools, which only admit highly qualified students. Unsurprisingly, inclusive STEM schools are much more demographically diverse than selective schools, with many more African-American and Hispanic students and students from economically disadvantaged backgrounds.

Education policy discussions increasingly focus on the urgent need for greater stress on STEM instruction, where the interconnections and real-world relevance of these fields are emphasized. In 2007 and 2010, the National Academies' Committee on Prospering in the Global Economy of the 21st Century released two reports focusing national attention on STEM education (The National Academies, 2007; 2010). These documents highlighted persistent deficiencies in K-12 education in science, math, engineering, and technology and advocated federal intervention to remedy these deficiencies. These recommendations included scholarships for academic programs that lead to teacher certification and professional development programs for current and future STEM teachers, a rigorous but voluntary

[2] The 2009 Program for International Student Assessment (PISA) in mathematics reveals American 15-year-olds performing worse than those of 30 other nations, including significantly poorer nations, including Estonia and Slovakia (OECD, 2010). According to the U.S. Department of Education, only 23% of bachelor's degrees in the United States are awarded in STEM fields, a number that has barely increased in decades, and 60% of students who enroll in college with the intent to pursue STEM fail to obtain a STEM degree (retrieved from http://dashboard.ed.gov). A 2006 brief from the National Science Foundation ("Science and Engineering Doctorates Hit All-Time High in 2005," retrieved from http://www.nsf.gov) documents that only 5.1% of doctorates in science and engineering in 2005 were awarded to blacks, Hispanics, and Native Americans, despite these ethnic groups constituting about one-third of the population, and this figure has changed little over time.

national curriculum in science and math, statewide specialty STEM schools, and additional resources for students and teachers to promote performance on Advanced Placement and International Baccalaureate examinations in high school.

Although these recommendations have some face validity, the supporting evidence for them is inconsistent, suggesting the importance of carefully structuring their implementation. For instance, recent research suggests that after controlling for the rigor of the remainder of a student's high school course load, additional Advanced Placement courses predict neither improved first-year college academic performance nor higher rates of retention to the second year of college (Klopfenstein & Thomas, 2009). The committee's recommendations for promoting STEM teachers' qualifications are too broad given the research literature's nuance of the relationship between teacher preparation and effectiveness, which suggests professional development in content, though not pedagogy, is modestly effective for secondary school mathematics instruction, but not in elementary school (Harris & Sass, 2010). Teachers certified in mathematics promote higher student performance in mathematics, but more general certification has much weaker effects; master's degrees emphasizing the study of content in mathematics and science promote student performance, but those in other fields, such as those emphasizing the study of education or pedagogy, do not (Goldhaber & Brewer, 1997). Very little is known, meanwhile, about the advantages of specialized STEM schools, though surveys have provided suggestive evidence that their graduates are more likely to complete a college major in mathematics, science, or engineering (Thomas, 2000). The preliminary results of an ongoing study of 14 STEM academies in Texas suggest that their students perform slightly better in science and mathematics than students in matched comparison schools (Young, 2011).

The dearth of high-quality empirical research on this policy initiative suggests that further analysis is urgently needed to properly tailor policy interventions in STEM, particularly analyses using longitudinal data that can trace the growth of students' trajectories over time. Existing research has not identified the systematic differences between strong and weak schools in STEM, and has yet to link student participation in advanced coursework and performance on standardized assessments. State longitudinal databases, such as those in Florida and North Carolina, are thus a powerful tool for viewing the landscape of STEM performance across entire populations of students and assessing the prospects for policy intervention. [3]

In the results that follow, several variables consistently show an association with success in STEM subjects across both states including the percentage of minority students, the percentage of students eligible for lunch assistance, turnover among the student body, and turnover among STEM teachers. Yet, a surprising result documented in both states is a significant negative association between STEM course participation and success in STEM, in addition to significantly lower levels of STEM course participation for underrepresented minorities in the highest performing schools. The results presented here invite further research into the causal relationships that determine learning in STEM subjects.

Data and Methodology

This report utilizes administrative educational data from longitudinal databases in Florida and North Carolina, which span the 2006-07 through 2008-09 school years. The longitudinal data tracks students over time across all schools in the public education system in both states, allowing for the estimation of school value-added estimates in STEM subjects.[4] The ability to estimate value-added is an important consideration for this study, and offers a clear advantage over other cross-sectional data sources (e.g., survey responses, results from the National Assessment of Educational Progress) by

[3] Note that standardized tests are only offered in math (and science in North Carolina); hence, school performance is based on either of these subjects only and not technology or engineering even though throughout the narrative of the text these schools are described as schools successful in STEM subjects. Several of the STEM metrics described below (e.g. course participation measures, certified STEM faculty), however, do include technology and engineering courses and teachers in creating these metrics.

[4] In the science and end-of-course subject tests, the method used is similar in spirit to, but not technically, a value-added model since the included prior test scores are those from different tests that are not vertically aligned; instead the estimates from these models are referred to as “quasi-value-added estimates.”

comparing current student achievement on the basis of past achievement. This is particularly compelling in the case of STEM-focused schools since they are commonly schools of choice that require parents or students to actively choose to attend the school; this choice may signal different levels of parent involvement or student interest in STEM subjects that could bias comparisons of otherwise similar groups of students. If this is indeed the case, students in STEM schools would show higher outcomes on key measures relative to students in non-STEM schools, but these differences would be due to selection bias and not necessarily reflect any difference in the school's actual performance. The value-added approach, by controlling for a student's past observed achievement along with other observable student characteristics, attempts to isolate changes in students' performance over time that is attributable to the school. While the resulting estimates still cannot be interpreted as causal, these growth-based value-added measures are generally viewed as better estimates of school inputs compared to status-based measures (e.g. achievement levels).[5]

Though the use of this longitudinal data offers some clear advantages, a major limitation is the omission of some key variables of interest, which is particularly true in the case of STEM school and program variables. Neither state database documents which schools have a formal STEM program, nor is there any information on school admission policies (i.e., selective or inclusive) or various STEM learning models. To remedy this deficiency, indicator variables on STEM schools in both states were generated by hand, using the following method. First, lists of STEM schools available from external sources were sought out; second, websites for all magnet schools in the states were visited to determine whether the magnet focus of the school was on STEM; and third, a search for STEM-related keywords in school names (e.g. "science," "career," "academy") was conducted and the websites of schools with these

[5] See, for instance, Tekwe et al. (2004). A parallel literature on estimating value-added teacher effects has grappled with whether these estimates are biased. Rothstein (2010) argues that the dynamic sorting of students over time on unobservable characteristics irretrievably biases the resulting estimates, though more recent evidence from Koedel & Betts (2011) suggests that transitory differences from sorting may bias single-year estimates but averaging performance across multiple years removes that bias.

keywords were visited to determine whether it had a STEM focus. A school identified as STEM by any of these three methods was considered a STEM school in the data.[6] No attempt was made to further distinguish STEM schools along additional dimensions, ignoring important qualitative variation in admissions, academic focus, and instructional programming. Hence, caution is warranted in interpreting the results as they pertain to STEM schools, as variation within that category is known to be substantive, as documented in Means, et al. (2008). Further details on the data in each state are described below.

Florida

The longitudinal data from Florida was provided from the Florida Education Data Warehouse, which maintains all administrative education data from the state. The standardized test outcomes in Florida are from the Florida Comprehensive Achievement Test for the Sunshine State Standards (FCAT-SSS), which is a series of vertically aligned, criterion-referenced tests used for accountability purposes in Florida. Students in grades 3 through 10 are tested at the end of each school year in both reading and math; because of the focus on STEM performance, only test scores in math are used in this report. These standardized test scores are the dependent variable in the value-added analysis.

The Florida data contains information on all teachers and tested students in the public education system. Student-level covariates used in estimating the value-added model include indicators on gender, race or ethnicity, eligibility for the free or reduced-price lunch program, and limited English proficiency. An additional indicator on student mobility (when not due to typical grade promotion) was created and used as a covariate in the model. School-level metrics on the percentage of teachers teaching any STEM courses (described below) who are certified in a STEM field and the percentage of

[6] The process of coding these STEM indicator variables was an admittedly subjective process, though efforts were made to apply the same classification uniformly across schools. I am confident that the schools identified as STEM through this method reflect actual STEM-focused schools, though if there are identification errors they likely arise from failing to identify a true STEM school (resulting in a false negative). This is certainly the case of STEM programs that do not encompass the entire school (e.g. a magnet program situated within traditional schools, or career academies); because the broader school was not focused on STEM, the school was not identified as a STEM school.

teachers new to the school (a year-specific measure of turnover among STEM faculty) are also calculated in the data.

Additionally, course membership files document the actual courses students are observed taking over time. In elementary schools, these course membership files are used to create an indicator variable on whether the school has a departmentalized or self-contained instructional model.[7] In middle and high schools, these membership files are used to create measures of course offerings and course participation. Course offerings measures are based on the variety of courses offered at the school in a given year; course participation measures are based on the number of students participating in given courses. Further details on these measures will be presented below in the discussion of methodology and in the appendix.

North Carolina

The North Carolina administrative data was compiled and distributed from the North Carolina Education Research Data Center, housed at the Child and Family Policy Center at Duke University. The standardized test scores in North Carolina are those on the End-of-Grade (EOG) tests administered as part of the state's ABC's accountability program. Math and reading tests are given annually to all students in grades 3 through 8. In addition to the math and reading tests, North Carolina has also administered an EOG science test to students in grades 5 and 8 only since the 2007-08 school year, which means two years of these scores are available for use in the analysis. Finally, North Carolina tests middle and high school students using standardized End-of-Course (EOC) tests for core courses required for graduation; those used in this study are from Algebra 1, Geometry, Algebra 2, Biology, Physical Science, Chemistry, and Physics.

[7] In coding the departmentalized indicator variable, a grade within a school and year is departmentalized if 50 percent or more of the students receive reading instruction and math instruction from different teachers.

Similar to the data from Florida, the longitudinal data in North Carolina also documents student demographics, eligibility for free or reduced-price lunch, and limited English proficiency. Course membership data is likewise used to create measures of course offerings and student participation. Finally, teacher-level certification data is used to ascertain which STEM teachers are certified to teach STEM courses,[8] and school-level measures on the percentage of STEM teachers who are new to the school are also computed.

Sample Description

The sample used in the study is limited to schools that serve at minimum one tested grade (grade 3 excluded) for each of the three years spanning the 2006-07 through 2008-09 school years.[9] In addition, schools had to also have valid student observations on course membership to be included in the sample. Only students with a valid prior test score in math were used; those missing prior math scores were dropped from the sample.

In the end, five different samples were employed in the analysis, each corresponding to a different test score as an outcome:

1. Florida schools with valid FCAT-SSS scores in math,
2. North Carolina schools with valid EOG scores in math,
3. North Carolina schools with valid EOG scores in science,

[8] The information on teacher certification varies across the states. In Florida, each course that the teacher is assigned to notates whether the teacher holds a certificate for that particular course. North Carolina, on the other hand, records teacher licensure and certification information separately but does not indicate which certificates cover which courses a teacher is assigned to; if a teacher held a certificate that had a STEM field specialty, that teacher was considered certified for all STEM courses taught. Because of this difference in variable construction, this variable is not comparable across the two states.

[9] The value-added methods in this report require students to have both a pretest and posttest score. Because grade 3 is the first tested grade in both states, it is used as a pretest score but not used as a posttest score. Test data from the 2005-06 school year is used to create the pretest scores for the students observed in the 2006-07 school year. For the North Carolina EOG Science sample, included schools only had to have valid test data from the most recent two years, since the EOG Science test was only administered in the last two years of the three-year time window considered here.

4. North Carolina schools with valid EOC scores in math subjects (Algebra 1, Geometry, and Algebra 2), and

5. North Carolina schools with valid EOC scores in science subjects (Biology, Physical Science, Chemistry, and Physics).[10]

Sample means for some of the key analysis variables are presented in Table 1; each column represents a different sample. The means reported for the *Student Characteristics* in the first eight rows of the table are those based on all student-year observations in each sample (because these samples span multiple years, many students are observed multiple times). The means reported in the *School Characteristics* section in the bottom nine rows of the table are simple means of all schools included in the sample.

The Florida FCAT-SSS sample is from a much larger state, and spans more grades than the other samples (grades 4-10); therefore, it includes far more students and schools than the other samples. Likewise, the North Carolina sample based on the EOG math test is considerably larger than the remaining North Carolina samples as the EOG math test covers more grades. Because the North Carolina sample based on the EOG science test only covers two tested grades (grades 5 and 8) during the last two school years, it is the smallest sample included in the study in terms of included student-year observations. The North Carolina EOC samples in math and science are limited to students in grades 6 or higher (though most students do not begin taking these EOC exams until high school); hence, standard K-5 elementary schools and many grade 6-8 middle schools will be excluded from these samples. Also, for the purpose of comparing schools that serve similar grades in estimation, the FCAT-SSS sample and the two North Carolina EOG samples (math and science) will each be broken into two mutually exclusive subsamples consisting of elementary grades (4 and 5) and middle and high-school grades (grades 6 to 10 in Florida, grades 6 to 8 in North Carolina). Consequently, schools that span both subsamples (e.g., a K-8

[10] The only EOC science exam administered in the 2006-07 school year was in biology. The EOC exams in physical science, chemistry, and physics were all administered in the following two years of the study.

school) will have two separate performance estimates in the same subject: one for elementary grades and one for middle and high-school grades.

White students (the omitted racial category) comprise the majority of students in North Carolina, and are nearly so in Florida. Asian students comprise a small share of students in both states. All remaining minority groups (i.e., black, Hispanic, or other racial groups) comprise the group of underrepresented minority students in STEM fields, which will be reported on later in the analysis. Additionally, STEM schools are uncommon in both states—constituting less than 2 percent of schools in the Florida FCAT and North Carolina EOG samples. Note that STEM schools are observed in higher proportions in the North Carolina EOC samples (in columns 4 and 5), which draw from schools serving grades 6 and higher, suggesting the specialty STEM school model is more common at the secondary level.

Methodology

The analysis proceeds by first estimating a school-level value-added estimate in each of the samples described above. In the case of the Florida FCAT and North Carolina EOG math samples, the following value-added regression was estimated:

$$(1) \quad y_{i,g,s,t} = \mathbf{y}_{\mathbf{i,s,t\text{-}1}}\mathbf{I}_{\mathbf{g}}\boldsymbol{\beta}_{\mathbf{1}} + \boldsymbol{X}_{\mathbf{i,g,s,t}}\boldsymbol{\beta}_{\mathbf{2}} + \mathbf{I}_{\mathbf{g}}\mathbf{I}_{\mathbf{t}}\boldsymbol{\beta}_{\mathbf{3}} + \mathbf{I}_{\mathbf{s}}\boldsymbol{\beta}_{\mathbf{4}} + \varepsilon_{\mathrm{i,g,s,t}}$$

This model predicts achievement for student *i* in grade *g* at school *s* at time *t* as a function of five components: 1) a vector of test scores in the immediately prior year, interacted with an indicator on grade ($\boldsymbol{y_{i,s,t-1}I_g}$); 2) a vector of student characteristics ($\boldsymbol{X_{i,g,s,t}}$);[11] 3) a grade-year fixed effect that removes systematic variation across different grade-year combinations ($\boldsymbol{I_g I_t \beta_3}$); 4) a school-level fixed effect that removes common residual differences that are attributable to specific schools ($\boldsymbol{I_s \beta_4}$); and 5) a random error ($\varepsilon_{i,g,s,t}$). The vector of prior test scores includes those observed in both math and reading, along with an indicator variable for those missing their prior reading score.[12] Included variables in the vector of student characteristics include indicator variables on gender, race or ethnicity (entered as a series of indicator variables), eligibility for the free or reduced-price lunch program, limited English proficiency, and whether the student newly moved to the school.[13] The parameters of interest in the model are those coefficients estimated for each school in the sample ($\widehat{\boldsymbol{\beta_4}}$), which are interpreted as the average residual difference in test scores between that school's students and the students in the mean school, all things equal. Note that one school fixed effect is estimated in each school, representing the student-weighted residual across all tested grades and years included in the sample.

The samples from North Carolina's EOG science test and EOC math and science tests differ from the model described above in that the tests are taken at specific points in time and are not part of a vertically aligned series of tests. As a result, these test scores cannot be used to compute a standard value-added estimate as the FCAT and EOG math test scores can be. For these samples a quasi-value-added approach is employed by using the student's most recent prior EOG test scores in both math and

[11] The inclusion of student demographic variables in a value-added model has aroused controversy in implementing these in practice, as they implicitly hold students to a different standard for growth depending on their demographic variables, which is at odds with policy objectives of holding all students to the same expectations (see Sanders, Wright, Rivers, & Leandro, 2009). Most value-added research includes these variables, as they marginally improve the model's fit overall, as I do here. Note, however, several studies have documented that value-added estimates that omit these variables are very highly correlated (correlations greater than 0.95) with those that include them (e.g., Aaronson, Barrow, & Sander, 2007; Ballou, Sanders, & Wright, 2004), and thus should not make any meaningful difference in the results I find here.

[12] Student missing their prior math score (e.g. they are new to the state public schools, were sick the day of the testing last year) were dropped from the analysis sample; see details under the *Sample Description* heading.

[13] Students' gains in learning are attributed to the school in which the student was tested, regardless of how long they attended the school prior to testing. Note that in the case of mobile students, this school will differ from the school and/or district in which they were tested in the year prior. Students who are mobile within the state public school system can be tracked with the longitudinal data, allowing prior test scores to be observed anywhere within the system.

reading as regressors, while using the EOG science score or EOC subject-test score as the dependent variable. Hence, the quasi-value-added model is identical to Equation (1) above, with the exception of using the alternate test score as the dependent variable. For the EOC exams, all math exams are evaluated within the same regression (interacting every covariate with the test subject), resulting in an aggregated school fixed effect for all EOC exams in math; the same process is followed in creating a school fixed effect for all EOC exams in science.

After estimating the value-added (or quasi-value-added) estimate for each of the analysis samples, [14] schools are binned into three categories based on these estimates: successful schools (those in the top 25 percent of the sample), average schools (middle 50 percent), and low-performing schools (bottom 25 percent).[15] Once schools are classified into performance categories, the schools in these groups are described and t-tests are conducted to see whether the group means for the successful and low-performing groups vary significantly from those in the average performance group.

Recall that many schools will appear in multiple samples, a different school fixed effect is estimated for each sample the school appears in. For instance, K-8 schools span both elementary and middle-school grades in the FCAT-SSS and EOG samples, and in North Carolina most schools have valid test scores in both math and science; for each sample in which the school appears, a separate value-added estimate is calculated and schools are categorized into performance groups according to these estimates.[16] Thus, schools that are included in multiple samples do not necessarily have the same

[14] These school-level fixed effects are estimates and thus are subject to error, which is predictably larger in schools with few students observed (and smaller in large schools with many students observed); to account for this error, I perform an empirical Bayes adjustment on the school fixed effects before determining where each school lies in the distribution. As a result of this adjustment, schools with imprecisely estimated fixed effect estimates are shrunk towards the distribution mean while those with a precisely estimated fixed effect are adjusted relatively little.

[15] Note that schools contained in the top or bottom quartiles of the distribution are not necessarily significantly different from the mean school's value-added performance, though this is true for the large majority of schools in either of the tails across all of the study samples. Likewise, many schools are included in the "average" group that have performance that is statistically different from the mean, but their performance is not different enough to put them in the top or bottom 25 percent of the error-adjusted distribution of value-added estimates.

performance classification across each measure. The implication of this decision to estimate school value-added separately for each of these samples is that schools may perform quite differently for each different outcome, which seems reasonable as schools that perform well in math may not necessarily perform well in science. As will be shown below, however, these value-added estimates are positively correlated across subjects for schools that appear in multiple samples.

In its essence, this investigation is an outlier study, where I look for significant differences across performance categories that may potentially explain some of the differences in outcomes, similar to a recent analysis of turnaround among low-performing schools (Hansen & Choi, 2011). Yet, caution is warranted in interpreting and generalizing these results, as value-added models cannot fully separate all school and student inputs that may be systematically correlated with achievement and hence may still contain some residual bias. Though these relationships cannot be interpreted as causal, understanding how schools successful in STEM subjects differ from low-performing schools may generate hypotheses on STEM learning that could be further evaluated with follow-up research.

STEM Metrics

When making these comparisons across successful, average, and low-performing schools in STEM, I want to particularly focus on metrics that may capture meaningful differences in either inputs or outcomes across schools. Several STEM metrics were created based on the data that were available, primarily using course membership data.

At the elementary level, available data on STEM inputs is scarce. The only meaningful measure of STEM capacity that could be captured was that on departmentalized instruction. In both states, the most common model of instruction in elementary schools is self-contained (where the same teacher

[16] All of the results tables presented below compare schools by samples only, and do not combine estimates across samples in any way. An alternative approach to this investigation could have been to estimate a latent variable that represents joint productivity in both math and science across all grades. My intention with this investigation, however, was to investigate differences across these various performance categories to see whether any consistent patterns emerged across subjects and/or grade-level differences, so I chose to keep the effects separate by the estimation sample.

teaches all academic subjects to the class). Departmentalized instruction uses a different teacher for each subject (or group of subjects), which should theoretically allow teachers to specialize within that area. As a result, one may hypothesize high-performing elementary schools in STEM have a disproportionately higher share of departmentalized instruction.

For the middle and high school grades (grades 6 and higher) in both Florida and North Carolina, the course membership information is much more varied and informative, allowing the creation of several STEM measures across schools. In creating these measures, caution was used to distinguish STEM courses into different categories: core STEM courses are academically oriented courses in core fields (e.g. algebra, biology, computer science); advanced core STEM courses are a subset of core STEM courses that are distinguished by being taught at the honors level, or are International Baccalaureate or Advanced Placement courses; vocational or technical STEM courses (abbreviated V/T throughout) are courses that apply STEM learning in specific ways (e.g. information technology, medical sciences, aerospace).[17] Under this categorization, virtually all science and math subjects are included in the core STEM category, while most technology and engineering classes are included in the V/T category. The course membership data was used to create 12 different variables, nested in four groups: course offering measures that represent the share of unique courses offered at a school that are STEM related; course participation measures that represent the level of student participation in STEM courses; URM participation measures that represent the level of participation in STEM courses for URM students; and course-specific participation measures that represent participation in specific advanced courses (i.e., calculus, AP or IB courses, or early algebra). The mechanical construction for each of these created STEM variables is described in the appendix.

A final measure of STEM inputs is available in North Carolina only: as part of the EOG, students fill out an accompanying survey that includes items on math and science classroom instruction. For this

[17] The Appendix details the specific course numbers used in each state to classify courses into the three STEM categories.

series of questions, students indicate whether certain classroom activities were common (whether lab experiments were conducted, whether computers or calculators were used, etc.) and the use of instructional time (whether most of the time was devoted to reading, discussion, or lecture, etc.). To reduce the dimensionality of these items (22 questions total), factor analysis was conducted and the first factor was retained for both math and science instruction. Virtually all of the survey items loaded positively into the instructional indices, though the weightings slightly favored participatory and applied learning over more passive activities.[18] These indices are included in the comparisons across successful, average, and low-performing schools in STEM.

Results

STEM vs. non-STEM Schools

Before binning schools into performance categories, Table 2 first makes some baseline comparisons between specialized STEM schools against those that do not have a STEM focus. Simple group means across schools from Florida are presented in columns 1-3 and North Carolina means are in columns 4-6. The Florida schools represented are those in the Florida FCAT sample; the North Carolina schools are from an amalgamation of all four North Carolina samples that removes duplicate school observations. Columns 1 and 4 present the means of all schools observed in the entire sample; non-STEM schools are in columns 2 and 5 and STEM schools are presented in columns 3 and 6.

Four observations from this table are particularly noteworthy. First, the key descriptors of the student bodies show some statistically significant variation between STEM and non-STEM schools, though the relationships are not necessarily consistent across states. For instance, STEM schools in both states have significantly smaller student bodies than the average school but they show some variation in the student body: Florida has significantly fewer students with limited English proficiency or who are

[18] Further information on the items included in creating these indices and the resulting factor weightings are documented in the Appendix.

eligible for the Free or Reduced-price Lunch program (FRL) in STEM schools compared to non-STEM schools; meanwhile North Carolina educates significantly more minority students in STEM schools than in non-STEM schools. Second, mean school performance on standardized tests does not significantly vary across groups (the reported statistic is a mean of mean performance across schools), though observed achievement varies significantly more across STEM schools in Florida. Third, the schools characteristics in both states show STEM schools are significantly more common at the high school (grades 9-12) level, and in Florida they are also more likely to be either charter or magnet schools. And fourth, the STEM metrics show students in STEM schools in North Carolina participate in all STEM courses (core, advanced, and V/T) at statistically significant higher rates than those in non-STEM schools; in Florida the increased participation in V/T STEM courses is statistically significant. The teacher workforce in North Carolina STEM schools also has a higher percentage of STEM faculty certified to teach in STEM subjects.

Performance Categories Based on Math Value-Added Estimates

Recall in order to estimate the school value-added estimates for the Florida FCAT and North Carolina EOG samples, the samples were subdivided into two groups: those serving elementary grades and those serving middle or high-school grades. Once segmented by school level and estimating the model, the schools are assigned to performance categories based on the school's value-added estimates in math. Key variables are compared across these performance categories and presented in Tables 3 and 4 for the elementary and middle or high school groups, respectively. Successful schools in math are presented in columns 1 and 2; average schools are presented in columns 3 and 4; low-performing schools are presented in columns 5 and 6. T-tests of significant differences in means across the three performance categories are conducted and notated in the table (asterisks indicate significant differences *greater* than the average schools category, daggers indicate significant differences *less* than the average schools group).

This comparison by performance group showed some evidence that was generally consistent across both states and both grade-level groupings. The first pattern to note is a general, though not universal, tendency for successful schools to show characteristics of higher socioeconomic status (significantly fewer minority students or fewer FRL-eligible students) or low-performing schools to serve more high-need students. Additionally, with the exception of North Carolina elementary schools, low-performing schools in both states have a significantly lower student re-enrollment rate (in other words, high student mobility across years), which is also commonly associated with high-need student populations and lower academic outcomes for mobile students (e.g., Rumberger & Larson, 2008). This consistent pattern between student demographics and value-added performance in math could either be due to a prevailing bias in the school-level value-added estimates that fail to capture these demographic differences at the schools, or it could be that these disadvantaged schools really provide a less rich educational experience; the methodology employed here cannot rule either alternative out. Note, however, that value-added measures and school mean achievement measures are not necessarily the same, though they are positively correlated; for instance, the pairwise correlation between these measures in math is 0.43 across all North Carolina elementary schools and increases to 0.70 across North Carolina middle schools.

Interestingly, low-performing elementary schools in math were significantly more common in rural settings compared to the average schools group; yet this relationship reversed in middle and high schools were low-performing schools were less common in rural settings (though not statistically significant for the North Carolina sample). The relationship between value-added math performance and charter schools also similarly changes across the school grade level: low-performing elementary schools in both states are significantly more likely to be charter schools (compared with the average schools group), but successful middle and high schools are more likely to be charter schools.

Finally, focusing on the STEM metrics shows some interesting, and some unexpected, patterns in the data. First, there is no evidence that STEM specialized schools perform better in math in either state. Second, the instructional indices in math and science generated from the student-level survey responses on classroom activities showed a statistically significant, positive relationship with math performance across all schools in both states. Third, turnover among STEM faculty appears to have a negative association with value-added math performance in both states (significantly lower turnover in successful Florida schools, significantly higher turnover in low-performing North Carolina schools). Yet, the measure on whether STEM faculty was certified to teach STEM courses did not show a consistent relationship across the two states: low-performing schools have significantly fewer certified faculty in Florida, but both successful and low-performing schools have significantly fewer certified faculty in North Carolina compared to the average schools group. A similar inconsistency across states is observed in the case of the variable on departmentalization in elementary schools: successful schools in Florida have a disproportionately higher frequency of departmentalization, yet low-performing schools are disproportionately more likely to be departmentalized in North Carolina.

The most surprising results, however, were the school mean course participation measures across these performance categories (presented in Table 4). In both states, both core STEM and V/T STEM course percent participation measures are lower in successful schools than they are in average and low-performing schools, and advanced course participation is also significantly lower in Florida's successful schools compared with average schools in the state (a similar pattern is observed in North Carolina's successful schools, though the difference from the average group mean is not statistically significant). This observed relationship runs counter to what one may expect of successful schools in STEM subjects—namely, that successful schools would have higher student participation rates in STEM courses—but that is countered in both states. A potential explanation for this counterintuitive result could be from endogenous course participation—schools with students struggling in STEM fields may encourage greater participation as a strategy to remedy the problem. Consistent with expectations, however, is the significantly lower frequency of Calculus, AP/IB courses, (in Florida only, not available with the EOG math data in North Carolina) and early Algebra 1 (in both states) participation among low-performing schools.

Finally, the course participation measures for underrepresented minority (URM) students were generally consistent across states; however, the relationship pointed in the opposite direction of current policy interest. Instead of observing higher URM participation in STEM courses among successful schools, as commonly recommended in discussions of STEM policy, URM participation rates were significantly lower among successful schools (relative to the average group) in core, advanced, and V/T STEM courses in both states. Note that these URM measures represent participation among URM students in the school only (i.e. these measures are conditional on URM status), which means that STEM participation in successful schools is significantly lower for the URM students that are there. This lower participation rate is in addition to the reduced number of URM students in successful schools compared with average and low-performing schools (see Table 2 above). The combined result, then, suggests the STEM courses most URM students participate in are disproportionately from average and low-performing schools in STEM. While this result is troubling, it cannot be separated from the contrary explanation that the schools identified as successful in this value-added framework are identified precisely because participation among URM students is unexpectedly lower in these schools, which would be the case if the coefficients on URM student achievement in the value-added methodology used here were systematically biased against URM students. Regardless of the explanation for this participation gap, further research into the source of this differential and URM students' access to successful STEM programming is certainly warranted based on the consistent patterns observed in these two states.

Performance Categories Based on Subject-Specific Quasi-Value-Added Estimates

The prior tables have only evaluated successful performance on the basis of a school's value-added estimate in math using the state accountability tests, but in North Carolina quasi-value-added measures based on alternative course tests can also be evaluated. Specifically, separate quasi-value-added estimates based on the EOG Science test, EOC exams in math subjects, and EOC exams in science subjects are all estimated in their corresponding analysis samples. The comparison of school means by performance category based are presented in Tables 5 and 6. Overall, the results in these tables were generally consistent with the results based on the math value-added estimates using the EOG math tests, though some results are worth highlighting here.

First, a school's value-added performance in math is significantly correlated with its estimated performance in science. In the EOG elementary sample, the pairwise correlation coefficient on the value-added estimates across schools is 0.33; in the EOG middle school sample, this measure increases to 0.49.[19] For the middle and high schools that appear in both the EOC math and science samples, this correlation is 0.62. Second, the math and science instructional indices showed a strong and consistent relationship with success in STEM when using the school estimates based on both math and science. And interestingly, the differential between the successful and low-performing groups are approximately equal for both indices; suggesting the indices may capture school qualities in STEM that are correlated across subjects, rather than simply a subject-specific measure of quality. These positive correlations across subjects suggest shared school factors (e.g. school leadership, instructional approach, laboratory facilities, peers) are important to school performance regardless of the outcome subject. Thus, caution

[19] Another way of thinking about the correlation of performance across subjects is to look at the incidence of schools considered successful in one subject, but low performing in the other. In the EOG elementary sample, 39 schools of the 1,303 (3.0%) included in both subject samples were considered low performing in math, but successful in science; 49 (3.8%) schools were considered low performing in science, but successful in math. In the EOG middle sample, the analogous numbers were 17 of 665 (2.6%) schools in both subject samples were low-performing in math, but successful in science; 20 (3.0%) schools were low-performing in science, but successful in math.

is warranted in generalizing the observed differences between schools, as they may be indicative of these school factors and are not necessarily indicative of subject-specific programming.[20]

Finally, a noteworthy point of departure in the quasi-value-added results is the level of STEM course participation in successful schools. As described above, core, advanced, and V/T STEM course participation were all lower in successful schools when performance was evaluated using math scores. Interestingly, this pattern reverses in advanced STEM course participation when schools are ranked based on science performance; using both the EOG and EOC tests, low-performing schools have significantly lower participation than average schools and successful schools have higher participation, though these differences are not always statistically significant. Yet, the pattern of lower core and V/T STEM course participation in successful schools continues (though the differences are not always statistically significant).

Conclusion

This analysis provides a descriptive snapshot of performance in STEM fields for schools in Florida and North Carolina. By using a value-added (or quasi-value-added) approach on the longitudinal data in these states, I attempt to assess differences in STEM outcomes that are attributable to schools. The method used here cannot be interpreted to provide answers on causal relationships due to non-random student sorting that may not be fully captured by covariates included in the model; however, the descriptive comparison of school means may hint at relationships among school inputs that potentially influence STEM outcomes that are of interest to policymakers, and point to areas of future research.

Based on this investigation, several variables show an association with success in STEM fields. Socioeconomic variables such as the percentage of minority students, and the percentage of students eligible for lunch assistance showed a generally consistent relationship with success in STEM in both

[20] Unfortunately, variables on differences in STEM-specific facilities or resources that might help distinguish between general school and specific program factors are not recorded in the state databases.

states. Variables capturing turnover among the student body and turnover among STEM teachers also showed a consistent relationship across states in which success was associated with stability in either variable. Indices on math and science instruction also showed a strong association with success in both math and science outcomes. Additionally, some of the STEM variables included in the analysis showed no clear relationship across states or tests; such was the case with the variable on specialized STEM schools, the percentage of STEM teachers who are certified, and departmentalization at the elementary level.

A surprising result that emerged was a significant negative association between STEM course participation (including advanced STEM courses) and success in STEM subjects, and this was a pattern observed in both Florida and North Carolina. Though the relationship in advanced STEM course participation was reversed when school performance was measured using science test outcomes in North Carolina, the systemic pattern across states is puzzling. The reasons for this negative association between participation and student learning cannot be addressed using this data alone, yet the result gives pause. A plausible explanation could be a quantity-quality tradeoff when schools offer STEM courses—the participation measures presented here are quantity measures and schools may compensate for low-quality courses with a higher quantity of them. In spite of these counterintuitive results with respect to course participation, enrollments in calculus, AP or IB courses, and early Algebra 1 all showed a strong association with success that was consistent with expectations—this may support a quality-quantity tradeoff hypothesis if these courses can be considered as proxies for course rigor. It's worth noting that current policy recommendations promote more rigorous coursework in STEM as a method to increase student outcomes, but are generally silent on whether students should increase the quantity of course taking in STEM subjects (e.g., National Research Council, 2011).

A troubling association revealed in this analysis is the low level of participation in STEM courses among URM students in successful schools, in addition to the lower levels of URM students in these schools to begin with. This negative association is at odds with policy interests that seek to both encourage STEM achievement overall and expand STEM participation for URM students. A positive association between these measures would imply a complementary relationship between success in STEM generally and participation among URM students specifically. The negative correlation actually observed in the data implies potential tradeoffs that may need to be addressed when promoting STEM policies. In other words, will the promotion of higher STEM achievement advantage students with a proclivity towards STEM fields at the expense of reducing learning for URM students? The descriptive evidence presented here suggests the answer may be yes.

Finally, this investigation has its limitations that must be addressed. First, precious few STEM-specific variables are captured in the longitudinal data; thus, many important relationships between STEM capacity and student outcomes cannot be explored here. Second, this analysis is only a descriptive snapshot of STEM performance across schools based on a value-added approach; though the approach explicitly attempts to reduce selection bias across from prior learning it may still include some residual bias. Third, the variation observed across schools for the variables investigated here is not determined exogenously, and therefore no causal relationship between success in STEM and the input variables is implied. And fourth, the value-added measures rely on standardized tests in science and math only and may not encompass all of the programmatic effectiveness in the STEM program. Similarly, inasmuch as school-wide resources and practices promote STEM learning, one must not assume that variations in STEM outcomes are entirely due to different STEM-specific inputs.

In summary, the analysis attempts to describe the lay of the land in terms of success in STEM fields and some basic descriptors of the schools that appear to be successful. While many of the relationships between success and school inputs were reasonably expected, several of the relationships revealed here warrant further research to understand how schools achieve success in STEM fields.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*, 95-135.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessments of teachers. *Journal of Educational and Behavioral Statistics*, *29*, 37-65.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources, 32*, 505-523.

Hansen, M., & Choi, K. (2011). *Chronically low-performing schools and turnaround: Evidence from three states*. CALDER Working Paper #60.

Harris, D. N., & Sass, T. R. (2010). Teacher training, teacher quality, and student achievement. *Journal of Public Economics*, *95*, 798-812.

Klopfenstein, K., & Thomas, M. K. (2009). The link between Advanced Placement experience and early college success. *Southern Economic Journal*, *75*, 873-891.

Koedel, C. & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy, 6*, 18-42.

Means, B., Confrey, J., House, A., & Bhanot, R. (2008). *STEM high schools: Specialized science, technology, engineering, and mathematics secondary schools in the U.S.* Arlington, VA: SRI International. Retrieved from http://ctl.sri.com/publications.

The National Academies. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Washington, DC: The National Academies Press.

The National Academies. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5.* Washington, DC: The National Academies Press.

National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Washington, DC: The National Academies Press.

OECD. (2010). PISA 2009 at a Glance, OECD Publishing. Retrieved from http://www.oecd.org.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics, 125*, 175-214.

Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education, 107*, 1-35.

Sanders, W. L., Wright, S. P., Rivers, J. C., & Leandro, J. G. "A response to criticisms of SAS EVAAS" SAS White Paper, 2009. Cary, NC: SAS Institute, Inc.

Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*, 11-36.

Thomas, J. (2000). First year findings: NCSSSMST longitudinal study. *NCSSSMST Journal*, 5(2), 4-5.

Young, V. M. (2011). *Inclusive STEM schools in Texas: Early promise and unanswered questions*. Proceedings from the Committee on Highly Successful STEM Schools or Programs for K-12 STEM Education, Board on Science Education and Board on Testing and Assessment, National Research Council. Retrieved from http://www7.nationalacademies.org.

Tables

Table 1. Descriptive Means of Data Samples Used in Analysis

State	Florida	North Carolina			
Sample	FCAT-SSS	EOG Math	EOG Science	EOC Math	EOC Science
Student Characteristics					
Percent Black	21.6%	27.6%	27.6%	30.9%	28.9%
Percent Hispanic	23.8%	9.9%	9.6%	7.1%	6.0%
Percent Asian or Pacific Islander	2.4%	2.3%	2.4%	2.3%	2.4%
Percent Other Ethnicity	3.8%	5.1%	4.9%	3.7%	3.4%
Percent FRL	46.8%	45.8%	45.6%	36.8%	34.3%
Percent LEP	3.9%	6.8%	6.7%	3.5%	2.8%
Percent New to School	10.9%	8.0%	10.5%	35.7%	9.7%
Student-year Observations	3,433,101	1,906,572	424,653	265,945	226,842
School Characteristics					
Enrollment	782.7	538	545	822	908
Met AYP	24.4%	51.2%	50.9%	35.1%	33.3%
Urban	27.5%	25.2%	25.2%	25.7%	21.1%
Town or Suburban	54.5%	27.8%	27.8%	28.0%	28.2%
Rural	17.7%	47.0%	47.0%	46.4%	50.7%
Percent STEM	1.6%	0.4%	0.4%	3.9%	4.4%
Percent Charter	8.2%	4.1%	4.4%	5.6%	4.8%
Percent Magnet	9.9%	5.7%	5.7%	6.0%	4.2%
School Observations	3,254	1,845	1,902	667	456

Note: Cell entries are sample means; variables listed under the student characteristics heading are weighted by students and those listed under the school characteristics heading are simple means. The FL FCAT-SSS sample includes students in grades 4-10; the NC EOG math sample includes students in grades 4-8; the NC EOG science tests include students in grades 5 and 8 only; the NC EOC math and science tests include all students in grade 6 or higher who completed the EOC test. Data in all samples span the 2006-07 to 2008-09 school years, with the exception of the NC EOG science test sample which spans the 2007-08 and 2008-09 school years.

Table 2. Descriptive School Means in Non-STEM and STEM Schools

	Florida			North Carolina		
	Total	Non-STEM Schools	STEM Schools	Total	Non-STEM Schools	STEM Schools
Student Body Characteristics						
Percent Minority	53.3%	53.4%	47.0%	41.9%	41.7%	52.6%*
Percent FRL	50.6%	50.8%	37.9%**	42.9%	43.0%	39.0%
Percent LEP	8.2%	8.3%	2.3%**	6.6%	6.6%	6.2%
Enrollment	783	787	493**	607	611	343**
Re-enrollment	72.1%	72.0%	75.3%	79.5%	79.6%	77.7%
Mean Reading Achievement	-0.16	-0.16	-0.16	-0.05	-0.05	-0.12
Standard Deviation Reading	0.54	0.54	0.72**	0.47	0.47	0.42
Mean Math Achievement	-0.13	-0.13	-0.13	-0.05	-0.05	-0.11
Standard Deviation Math	0.49	0.49	0.73**	0.52	0.52	0.49
School Characteristics						
K-5 School	46.3%	46.7%	17.0%**	43.3%	43.8%	9.7%**
6-8 School	14.9%	14.9%	17.0%	15.7%	15.9%	3.2%
9-12 School	10.0%	9.8%	28.3%**	15.4%	15.0%	48.4%**
Other grade configuration	28.7%	28.6%	37.7%	25.6%	25.4%	38.7%
Departmentalized (if elementary)	9.1%	9.2%	0.0%	15.6%	15.6%	16.7%
Met AYP Criteria	26.8%	26.7%	37.5%*	49.2%	49.4%	43.0%
Urban	27.5%	27.6%	22.6%	24.4%	24.4%	25.8%
Town or Suburban	54.8%	54.8%	56.6%	27.7%	27.7%	32.3%
Rural	17.7%	17.7%	20.8%	47.8%	47.9%	41.9%
Percent Charter Schools	8.2%	7.7%	37.7%**	4.0%	3.9%	6.5%
Percent Magnet Schools	9.9%	9.6%	26.4%**	5.2%	5.2%	6.5%
Select STEM Metrics						
Advanced STEM Course Participation	2.5%	2.5%	3.7%	1.6%	1.6%	6.2%**
Core STEM Course Participation	24.9%	24.9%	24.9%	23.5%	23.4%	32.4%**
V/T STEM Course Participation	4.3%	4.1%	11.3%**	2.4%	2.4%	5.8%**
Percent Certified STEM faculty	86.1%	86.2%	85.5%	70.4%	70.1%	79.8%*
Number of Schools Observed	**3,254**	**3,201**	**53**	**2,346**	**2,315**	**31**
Percentage of Students	**100.0%**	**99.0%**	**1.0%**	**100.0%**	**99.3%**	**0.7%**

*, **: group mean significantly different from the non-STEM group for p<0.05 and p<0.01, respectively.

Note: Cell entries are simple means across schools. Minorities do not include Asian or Pacific Islander racial groups. Included schools are all unique school observations in the Florida and North Carolina analysis samples. STEM schools are those externally identified as having a STEM-focused mission or curriculum.

Table 3. Descriptive Means by Performance in Math: Elementary Schools

	Successful Schools		Average Schools		Low-performing Schools	
State	Florida	North Carolina	Florida	North Carolina	Florida	North Carolina
Student Body Characteristics						
Percent Minority	57.8%**	38.6%‡	53.5%	44.0%	50.8%	42.1%
Percent FRL	52.7%†	43%‡	55.4%	46.2%	58.1%†	47.2%
Percent LEP	12.7%	7.8%	11.3%	7.4%	7.9%‡	6.0%‡
Enrollment	693.2	524	673.6	531	558.8‡	462‡
Re-enrollment	75.8%*	80.6%	74.7%	79.6%	70.3%‡	78.8%
School Achievement						
Met AYP	57.0%**	68.9%**	33.4%	55.0%	16.2%‡	47.0%‡
Average Reading	-0.1**	0.08**	-0.04	-0.06	-0.29‡	-0.17‡
Average Math	-0.13**	0.15**	-0.04	-0.07	-0.41‡	-0.25‡
Math Value-added Estimate	0.24**	0.14**	0.11	0.00	-0.06‡	-0.15‡
Average Science		0.15**		-0.06		-0.23‡
Science Value-added Estimate		0.07**		-0.01		-0.11‡
School Characteristics						
Urban	27.9%	31.5%	26.3%	26.9%	29.8%	18.2%‡
Town or Suburban	59.7%	26.8%	58.6%	29.5%	45.7%‡	20.8%‡
Rural	12.4%†	41.7%	15.1%	43.6%	24.5%**	61.0%**
Percent STEM Schools	0.8%	0.3%	0.4%	0.6%	1.0%	0.0%
Percent Charter Schools	4.8%	2.1%	5.0%	4.0%	16.6%**	10.1%**
Percent Magnet Schools	7.8%	8.2%	8.1%	6.1%	5.2%	2.1%‡
STEM Metrics						
Departmentalized	15.1**	12.6%	7.2%	14.2%	6.9%	21.8%**
Math Instructional Index		0.04**		-0.02		-0.08‡
Science Instructional Index		-0.04**		-0.08		-0.11‡
Number of Schools Observed	**477**	**336**	**955**	**672**	**477**	**336**
Percentage of Students	**26.7%**	**25.6%**	**51.9%**	**51.9%**	**21.5%**	**22.6%**

*, **: group mean significantly greater than the Average STEM group for p<0.05 and p<0.01, respectively.
†, ‡: group mean significantly less than the Average STEM group for p<0.05 and p<0.01, respectively.

Note: Cell entries are simple means across schools. Included schools are from the FL FCAT-SSS sample and NC EOG Math sample that serve any of grades 4 and 5 over the three-year period spanning the 2006-07 to 2008-09 school years. Successful (top 25 percent), average (middle 50 percent), and low-performing (bottom 25 percent) groups are based on math value-added estimates in the school . Minorities do not include Asian or Pacific Islander racial groups.

Table 4. Descriptive Means by Performance in Math: Middle and High Schools

	Successful Schools		Average Schools		Low-performing Schools	
State	Florida	North Carolina	Florida	North Carolina	Florida	North Carolina
Student Body Characteristics						
Percent Minority	47.4%	33.4%‡	49.0%	41.1%	58.8%**	49.7%**
Percent FRL	41.0%‡	38.2%‡	45.4%	44.4%	46.4%	49.0%*
Percent LEP	5.5%	5.1%	5.1%	5.6%	2.7%‡	4.2%‡
Enrollment	884.0‡	595	1244.9	622	229.9‡	378‡
Re-enrollment	79.5%**	84.8%	77.0%	84.7%	46.9%‡	62.3%‡
School Achievement						
Met AYP	34.3%**	55.9%**	5.0%	33.7%	0.7%	28.8%†
Average Reading	0.15**	0.13**	-0.10	-0.074	-0.98‡	-0.52‡
Average Math	0.18**	0.18**	-0.13	-0.105	-1.14‡	-0.62‡
Math VAM	0.08**	0.14**	-0.08	-0.009	-0.36‡	-0.20‡
Average Science		0.17**		-0.077		-0.52‡
Science VAM		0.06**		-0.016		-0.17‡
School Characteristics						
Urban	26.2%	18.8%	24.5%	19.0%	34.8%**	23.6%
Town or Suburban	55.1%	21.5%	52.7%	28.0%	49.0%	31.9%
Rural	18.6%	59.7%	22.8%	53.0%	16.2%‡	44.5%
Percent STEM Schools	2.5%	0.6%	2.7%	0.3%	2.9%	1.1%
Percent Charter Schools	13.5%**	11.6%*	9.6%	6.9%	9.8%	10.5%
Percent Magnet Schools	13.5%	2.2%	16.7%	5.6%	2.0%‡	2.8%
Core STEM Courses						
Percent Course Offerings	21.5%**	27.4%‡	26.0%	30.2%	27.8%‡	31.7%*
Percent Participation	20.3%**	32.2%‡	26.8%	36.8%	25.9%†	36.5%
Percent URM Participation	20.4%**	28.7%‡	26.9%	31.9%	25.9%†	32.6%
Advanced STEM Courses						
Percent Course Offerings	3.7%‡	0.7%‡	10.0%	1.4%	1.9%‡	1.1%
Percent Participation	2%‡	0.9%	3.8%	1.2%	0.5%‡	0.8%†
Percent URM Participation	1.6%‡	0.2%‡	2.8%	0.4%	0.4%‡	0.3%
V/T STEM Courses						
Percent Course Offerings	3.8%**	3.4%‡	11.0%	4.6%	5.7%‡	4.7%
Percent Participation	2.2%**	2.7%‡	5.6%	3.6%	3.9%‡	3.5%
Percent URM Participation	2.1%**	2.4%‡	5.1%	3.1%	3.7%‡	2.9%
Specific STEM Courses						
Calculus	7.4%		7.7%		0.7%‡	
AP/IB Science	27.1%		28.0%		1.0%‡	
AP/IB Math	5.8%		5.6%		0.4%‡	
AP/IB Math or Science	32.9%		33.6%		1.3%‡	
Early Algebra 1	41.7%**	27.0%**	34.3%	19.6%	10.8%‡	12.1%‡
STEM Teachers and Instruction						
Percent Turnover STEM Faculty	15.8%‡	15.8%†	21.1%	17.7%	22.4%	20.3%*
Percent Certified STEM faculty	89.1%	46.4%‡	88.9%	58.2%	77.3%‡	53.5%†
Math Factor		0.06**		-0.028		-0.30‡
Science Factor		0.08**		0.015		-0.18‡
Number of Schools Observed	**408**	**181**	**816**	**364**	**408**	**182**
Percentage of Students	**24.5%**	**26.7%**	**69.1%**	**56.2%**	**6.4%**	**17.1%**

*, **: group mean significantly greater than the Average STEM group for p<0.05 and p<0.01, respectively.
†, ‡: group mean significantly less than the Average STEM group for p<0.05 and p<0.01, respectively.

Note: Cell entries are simple means across schools. Included schools are from the FL FCAT-SSS sample and NC EOG Math sample that serve any of grades 6 through 10 in FL or grades 6 through 8 in NC over the three-year period spanning the 2006-07 to 2008-09 school years. Successful (top 25 percent), average (middle 50 percent), and low-performing (bottom 25 percent) groups are based on math value-added estimates in the school. Minorities do not include Asian or Pacific Islander racial groups.

Table 5. Descriptive Means by Performance in Science

	Successful Schools		Average Schools		Low-performing Schools	
Sample	5th Grade	8th Grade	5th Grade	8th Grade	5th Grade	8th Grade
Student Body Characteristics						
Percent Minority	33.4%‡	30.4%‡	39.9%	38.6%	52.7%**	56.1%**
Percent FRL	39.3%‡	31.7%‡	44.0%	43.4%	51.1%**	51.7%**
Percent LEP	6.8%	4.4%†	7.4%	5.4%	7.3%	4.7%
Enrollment	508†	672	538	625	452‡	379‡
Re-enrollment	81.6%**	84.7%*	79.8%	82.1%	75.8%‡	65.0%‡
School Achievement						
Met AYP	68.6%**	51.6%**	56.2%	34.7%	45.3%‡	28.0%†
Average Reading	0.12**	0.24**	-0.02	-0.05	-0.28‡	-0.56‡
Average Math	0.12**	0.23**	-0.03	-0.08	-0.29‡	-0.59‡
Math VAM	0.05**	0.03**	0.00	-0.02	-0.05‡	-0.12‡
Average Science	0.36**	0.44**	-0.03	-0.05	-0.50‡	-0.71‡
Science VAM	0.29**	0.24**	-0.01	-0.02	-0.30‡	-0.33‡
School Characteristics						
Urban	23.2%	26.6%*	26.2%	19.1%	28.4%	19.7%
Town or Suburban	24.7%	21.5%	26.9%	25.1%	28.7%	35.4%**
Rural	52.1%	52.0%	46.9%	55.8%	42.9%	45.5%†
Percent STEM Schools	1.2%*	0.6%	0.1%	0.3%	0.0%	0.6%
Percent Charter Schools	7.5%*	16.3%**	4.5%	7.3%	6.1%	6.7%
Percent Magnet Schools	3.8%	6.3%	6.0%	5.1%	6.2%	1.1%†
Core STEM Courses						
Percent Course Offerings		30.3%		30.1%		32.6%**
Percent Participation		36.4%		36.4%		38.7%**
Percent URM Participation		32.0%		31.9%		34.3%**
Advanced STEM Courses						
Percent Course Offerings		2.0%		1.6%		0.6%‡
Percent Participation		2.2%*		1.5%		0.4%‡
Percent URM Participation		0.8%**		0.4%		0.1%‡
V/T STEM Courses						
Percent Course Offerings		4.7%		4.7%		5.1%
Percent Participation		3.4%		3.7%		4.1%
Percent URM Participation		3.0%		3.2%		3.2%
Specific STEM Courses						
Early Algebra 1		26.8%**		18.3%		13.6%‡
STEM Teachers and Instruction						
Percent Turnover STEM Faculty	15.5%	20.0%*	15.6%	17.3%	17.7%*	20.5%**
Percent Certified STEM faculty		58.7%		58.8%		57.8%
Math Factor	0.03**	0.07**	-0.01	-0.03	-0.09‡	-0.28‡
Science Factor	0.01**	0.20**	-0.06	0.03	-0.14‡	-0.20‡
Departmentalized (if elementary)	16.7%		15.2%		15.9%	
Number of Schools Observed	**345**	**178**	**690**	**356**	**345**	**179**
Percentage of Students	**24.9%**	**29.2%**	**52.9%**	**54.3%**	**22.2%**	**16.5%**

*, **: group mean significantly greater than the Average STEM group for $p<0.05$ and $p<0.01$, respectively.

†, ‡: group mean significantly less than the Average STEM group for $p<0.05$ and $p<0.01$, respectively.

Note: Cell entries are simple means across schools. Included schools are from the NC EOG Science sample that serve grade 5 or 8 in over the two-year period spanning the 2007-08 to 2008-09 school years. Successful (top 25 percent), average (middle 50 percent), and low-performing (bottom 25 percent) groups are based on science quasi-value-added estimates in the school. Minorities do not include Asian or Pacific Islander racial groups.

Table 6. Descriptive Means by Performance on EOC exams in Math and Science

	Successful Schools		Average Schools		Low-performing Schools	
	Math EOC	Science EOC	Math EOC	Science EOC	Math EOC	Science EOC
Student Body Characteristics						
Percent Minority*	33.9%†	38.2%	39.7%	39.2%	55.0%**	47.5%**
Percent FRL	28.2%‡	28.9%‡	35.7%	36.2%	47.1%**	43.5%**
Percent LEP	4.3%	3.4%	4.0%	3.5%	4.2%	3.1%
Enrollment	943	1,079	907	992	522‡	563‡
Re-enrollment	84.8%**	81.9%	81.3%	80.6%	64.6%‡	64.1%‡
School Characteristics						
Met AYP	45.3%	45.9%*	39.7%	37.2%	31.9%†	30.4%
Urban	25.9%	22.8%	23.4%	21.6%	30.5%	19.3%
Town or Suburban	30.1%	30.7%	26.4%	24.7%	28.1%	32.5%
Rural	44.0%	46.5%	50.2%	53.7%	41.3%	48.2%
Percent STEM Schools	1.2%	2.6%	2.7%	3.1%	9.0%**	8.8%*
Percent Charter Schools	5.4%	7.1%	5.1%	2.6%	6.0%	6.1%
Percent Magnet Schools	7.9%	5.3%	5.2%	4.4%	4.3%	1.8%
Core STEM Courses						
Percent Course Offerings	29.0%	26.9%	28.7%	27.8%	31.8%**	30.3%**
Percent Participation	40.3%	40.4%‡	40.2%	42.0%	40.5%	40.7%†
Percent URM Participation	34.3%	34.8%	34.3%	35.5%	35.0%	35.8%
Advanced STEM Courses						
Percent Course Offerings	5.6%‡	8.6%	7.2%	8.0%	4.6%‡	5.1%‡
Percent Participation	6.2%	9.3%	7.3%	8.4%	4.6%‡	4.4%‡
Percent URM Participation	2.0%	2.9%	2.6%	2.6%	1.6%‡	1.3%‡
V/T STEM Courses						
Percent Course Offerings	6.1%‡	7.6%	7.3%	8.4%	6.7%	7.3%‡
Percent Participation	4.4%†	4.9%‡	5.1%	5.9%	4.8%	5.2%
Percent URM Participation	3.2%	3.3%‡	3.5%	3.8%	3.4%	3.5%
Specific STEM Courses						
Calculus	18.5%**	17.2%	12.8%	13.7%	7.6%‡	7.4%‡
AP/IB Science	1.9%	1.6%	1.0%	1.4%	0.7%	0.2%†
AP/IB Math	0.7%	1.1%	1.4%	1.4%	0.5%	0.3%
AP/IB Math or Science	2.6%	2.7%	2.4%	2.9%	1.2%	0.5%
Early Algebra 1	37.8%**	38.2%*	20.5%	12.4%	15.8%	7.6%
STEM Teachers and Instruction						
Percent Turnover STEM Faculty	18%†	19.9%	20.1%	19.8%	24.4%**	20.3%
Percent Certified STEM faculty	77.6%	81.7%	80.0%	84.6%	73.1%‡	76.2%‡
Number of Schools Observed	**166**	**114**	**334**	**228**	**167**	**114**
Percentage of Students	**28.6%**	**29.8%**	**55.4%**	**54.7%**	**16.0%**	**15.5%**

*, **: group mean significantly greater than the Average STEM group for p<0.05 and p<0.01, respectively.

†, ‡: group mean significantly less than the Average STEM group for p<0.05 and p<0.01, respectively.

Note: Cell entries are simple means across schools. Included schools are from the NC EOC Math and EOC Science samples that serve any of grades 6 or higher over the three-year period spanning the 2006-07 to 2008-09 school years. Successful (top 25 percent), average (middle 50 percent), and low-performing (bottom 25 percent) groups are based on quasi-value-added estimates in the school. Minorities do not include Asian or Pacific Islander racial groups.

Appendix

1. Categorizing coursework into core STEM, advanced core STEM, and V/T STEM

The course membership files from both states were used to categorize students' course enrollments into STEM categories. The data fields recording course enrollment varied by state, as did the actual course titles, though efforts were made to keep the classifications in each state as comparable as possible. The three STEM categories were defined based on the following definitions:

- Core STEM – academic course on subject knowledge in a STEM discipline
- Advanced core STEM – any core STEM class that was taught at an honors, AP, or IB level
- V/T STEM – any course that applied STEM-based learning or technologies in a vocational setting (though the vocational setting may not necessarily be in a STEM-related occupation)

The specific courses assigned to these categories are detailed for each state below:

North Carolina

Category	Course codes
Core STEM	2001-2599, 3001-3999
Advanced core STEM	Any core STEM course code with an academic level value of 5,7, or 8 OR if the course title included "honors", "AP", or "IB"
V/T STEM	6208, 6311-6313, 6340-6535, 6810, 6821-6828, 6851-6872, 7200-7399, 7511-7613, 7631-7633, 7651-7663, 7721-7743, 7901-7903, 7935-7936, 7972-7992, 8005-8999

Note: Course codes listed above can be linked to course titles using the *Outline of the Course Coding Structure for NC Public Schools*, prepared by the North Carolina Department of Public Instruction, available at http://www.ncpublicschools.org/docs/curriculum-instruction/home/2009-10coursecodes.pdf.

Florida

Category	Course codes
Core STEM	Course number of 0200-, 0201-, 121-, 120-, 20-, 7812010, or 7820010, or college course prefix of "AST", "BOT", "MCB", "BSC", "CHM", "CIS", "COP", "EVR", "GLY", "STA", "MAA", "MAC", "MAD", "MAP", "MTG", "MTB", "MAS", "MAT", "MGF", "MET", "PHY", "PSC", "ZOO", or course ID of 47308.
Advanced core STEM	Course was an "AP", "IB", or "honors" course, or was both a course taken at a university and a "core" course.
V/T STEM	Course number of 0103-, 1100-, 1802-, 18003-, 13014-, 05007-, 2301-, 129-, 0500920, 7912340, 7920010, 7980190, 870036-, 860-, 891-, 88272-, 8129-, 8106-, A010-, 8121-, A020-, 8112-, 8115-, 8111-, 8103-, 8118-, P150-, 8914-, 8913-, 8113-, P440-, 8916-, 8100-, 81003-, 81002- 8722-, 8725-, I480-, I469-, I470-, 8723-, I46-, I48-, 872-, 870-, 879-, I00-, 877-, I10-, 820-, I47-, 873-, B070-, 871-, 8209-, 8208-, 8207-, 8200-, 8212-, B078-, B079-, B070-, 8206-, 8208-, B077-, 8718-, I480-, 8417-, H170-, W170-, W170-, 8418-, H181-, H1702-, H1703-, V2004-, 8400-, 8499-, H000-, H179-, I150-, 8743-, I159-, I062-, 8721-, 8730-, I470-, I480-, 8754-, 8775-, 8751-, I490-, 8716-, 8730-, 8732-, I460-, 8736-, 3027-, 8743-, 8706-, I469-, 8775-, 8002-, 87003-, 85003-, I490-, 8766-, I470-, 8715-, 8709-, 8710-, I480-, 8766-, 8742-, 8751-, 8303-, 83004, 0501310-0501340, or 0500710-0500760, or college course prefix of "ACR", "AER", "ARR", "ATF", "AMT", "ASC", "ARC", "BCN", "BCV", "GRA", "CAP", "CDA", "CEN", "CET", "CGS", "OCA", "DEA", "EER", "EET", "EGN", "EGS", "ETD", "EMT", "FFP", "HIM", "HSC", "MUM", "PMT", "QMB", or "RTV", or course ID of 20499 or 20508.

Note: The titles of courses identified above can be found in Florida's *2008-2009 Course Directory*, Sections 3-5, prepared by that state's Department of Education, available at http://www.fldoe.org/articulation/CCD/files/sec3-high.pdf.

2. Using course membership data to create STEM metrics.

As described in the *Data and Methodology* section, course membership data in both states are used to create measures that capture variation in STEM offerings or participation across schools. The twelve created variables are constructed as follows:

1. Course offering measures—these represent the share of unique courses offered at a school that are STEM related in some way; three different course offering measures are defined in the following way:
   a. Core STEM course offerings = $\frac{Unique\ core\ STEM\ courses}{Total\ unique\ courses\ offered}$
   b. Advanced STEM course offerings = $\frac{Unique\ advanced\ core\ STEM\ courses}{Total\ unique\ courses\ offered}$
   c. V/T STEM course offerings = $\frac{Unique\ V/T\ STEM\ courses}{Total\ unique\ courses\ offered}$

2. Course participation measures—these represent the level of student participation in STEM subject courses (weighted by semester course length) as a share of student participation in all courses at a school; likewise three different participation measures are calculated:
   a. Core STEM course participation = $\frac{Total\ student-semesters\ in\ core\ STEM\ courses}{Total\ student-semesters\ in\ all\ courses}$
   b. Advanced STEM course participation = $\frac{Total\ student-semesters\ in\ adv.\ STEM\ courses}{Total\ student-semesters\ in\ all\ courses}$
   c. V/T STEM course participation = $\frac{Total\ student-semesters\ in\ V/T\ STEM\ courses}{Total\ student-semesters\ in\ all\ courses}$

3. Underrepresented minority (URM) participation measures—these represent the level of participation in STEM courses among the school's underrepresented minority students

(non-Asian minorities); these metrics mirror the course participation metrics described above:

a. URM core STEM participation = $\frac{Total\ URM\ student{-}semesters\ in\ core\ STEM}{Total\ URM\ student{-}semesters\ in\ all\ courses}$

b. URM advanced STEM participation = $\frac{Total\ URM\ student{-}semesters\ in\ adv.STEM}{Total\ URM\ student{-}semesters\ in\ all\ courses}$

c. URM V/T STEM participation = $\frac{Total\ URM\ student{-}semesters\ in\ V/T\ STEM}{Total\ URM\ student{-}semesters\ in\ all\ courses}$

4. Course-specific participation measures—these represent the level of enrollment in specific courses that are generally viewed as particularly rigorous:

a. Calculus participation (only in schools that serve grades 11 and 12)

$$= \frac{Total\ student{-}years\ in\ calculus}{\frac{1}{2}(Total\ grade\ 11\ \&\ grade\ 12\ students)}$$

b. AP/IB STEM participation (only in schools that serve grades 11 and 12)

$$= \frac{Total\ student{-}years\ in\ AP\ or\ IB\ STEM\ courses}{\frac{1}{2}(Total\ grade\ 11\ \&\ grade\ 12\ students)}$$

c. Early Algebra 1 participation (only in schools that serve grade 8)

$$= \frac{Total\ student{-}years\ in\ Algebra\ 1\ for\ students\ in\ grade\ 8\ or\ lower}{Total\ grade\ 8\ students}$$

3. Creating instructional indices from survey responses using factor analysis

In North Carolina, students taking the EOG tests also filled out an accompanying survey on the classroom instructional practices in math and science. Factor analysis was used to reduce the dimensionality of these items into a single index of instructional quality in each subject. The variables appearing in the data, the information contained in the variable, and the resulting factor weighting applied are recorded in the following tables. Factor loading values represent the importance of a specific item in the instructional index—larger values have greater weight, negative values decrease the index value.

Science Items

Variable Name	Question (1 = Yes, 0 = No)	Factor loading
SCICOMP	Used Computers, Calculators, or Other Machines to Learn Science in Science Class	0.6447
SCIEXPER	Completed a Science Experiment or Project in Science Class	0.7979
SCIEXPLAN	Listened to the Teacher Explain Something About Science	0.8119
SCIGROUP	Worked on Labs in Pairs or Small Groups in Science Class	0.7434
SCIPROJ	Completed a Science Project Outside the Classroom	0.5819
SCIREAD	Read About Science in Books, Magazines, or Articles in Science Class	0.7179
SCITEXPER	Observed the Teacher Performing an Experiment in Science Class	0.6939
SCITIME1	Spent the *Most* Time in Science Class Reading About Science	0.5191
SCITIME2	Spent the *Most* Time in Science Class Completing Science Project	0.4692
SCITIME3	Spent the *Most* Time in Science Class Observing the Teacher Performing an Experiment	0.2827
SCITIME4	Spent the *Most* Time in Science Class Listening to the Teacher Explain Something About Science	0.6557
SCITIME5	Spent the *Most* Time in Science Class Working on Labs in Pairs or Groups	0.4524
SCITIME6	Spent the *Most* Time in Science Class Completing a Project Outside the Classroom	0.1501
SCITIME7	Spent the *Most* Time in Science Class Using Computers, Calculators, or Other Machines to Learn About Science	0.2507

Math Items

Variable Name	Question (1 = Yes, 0 = No, unless indicated otherwise)	Factor loading
MATHPROB	Student Explains Solution to Math Problems in Class A = A teacher never asks me to explain how I solve math problems B = About once or twice a month teacher asks me to how I solve math problems C = About once a week D = Almost every day	B: -0.1159 C: -0.2135 D: 0.5811
MATHNOTES	Student Listened and Took Notes in Math Class	0.4628
MATHMACH	Student Used Computers, Calculators, or Other Machines to Learn Math in Math Class	0.5584
MATHGROUP	Student Worked in Groups to Learn an Idea or Solve a Problem in Math Class	0.5477
MATHREAD	Student Read About Math in Books, Magazines, or Articles in Math Class	0.4054
MATHTALK	Student Talked About How Math is Used in Other Subjects in Math Class	0.4940
MATHEXPL	Student Took Tests and Had to Explain His/Her Answers in Math Class	0.5172
MATHDISC	Student Discussed How Math is Used in Everyday Life in Math Class	0.5204