



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*



High School Course  
Access and Postsecondary  
STEM  
Enrollment and  
Attainment

Rajeev Darolia  
Cory Koedel  
Joyce B. Main  
Felix Ndashimye  
Junpeng Yan

---

## High School Course Access and Postsecondary STEM Enrollment and Attainment

Rajeev Darolia  
*University of Missouri*

Cory Koedel  
*University of Missouri/CALDER*

Joyce B. Main  
*Cornell Higher Education Research Institute*

Felix Ndashimye  
*University of Missouri*

Junpeng Yan  
*University of Missouri*

## Contents

---

Acknowledgements.....	ii
Abstract.....	iii
1. Introduction.....	1
2. Empirical Approach.....	5
3. Data and Context.....	10
4. Results.....	17
5. Sensitivity.....	22
6. Effect Heterogeneity.....	25
7. Conclusion.....	28
References.....	31
Tables and Figures.....	34

## Acknowledgements

---

We gratefully acknowledge research support from the National Science Foundation (awards 1532015 and 1531920) and CALDER, which is funded by a consortium of foundations (for more information about CALDER funders, see [www.caldercenter.org/about-calder](http://www.caldercenter.org/about-calder)). We thank the Missouri Department of Higher Education and Missouri Department of Elementary and Secondary Education for data access, Mark Ehlert for research support, and Kevin Stange and participants at the 2017 APPAM conference, 2018 CALDER conference, and Vanderbilt University seminar series for useful comments. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the author(s) are affiliated.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders.

CALDER • American Institutes for Research  
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007  
202-403-5796 • [www.caldercenter.org](http://www.caldercenter.org)

## **High School Course Access and Postsecondary STEM Enrollment and Attainment**

Rajeev Darolia, Cory Koedel, Joyce B. Main, Felix Ndashimye, Junpeng Yan

CALDER Working Paper No. 186

February 2018

### **Abstract**

We study the effects of access to high school math and science courses on postsecondary STEM enrollment and degree attainment using administrative microdata from Missouri. Our data panel includes over 140,000 students from 14 cohorts entering the 4-year public university system. The effects of high school course access are identified by exploiting plausibly exogenous variation in course offerings within high schools over time. We find that differential access to high school courses does not affect postsecondary STEM enrollment or degree attainment. Our null results are estimated precisely enough to rule out moderate impacts.

Keywords: STEM, College sorting; High school curricula

## 1. Introduction

Increased human capital production in science, technology, engineering, and mathematics (STEM) fields is a prominent policy goal of the United States. Workers with STEM backgrounds earn more on average than other workers, labor demand in STEM fields is projected to be strong, and expanding the STEM workforce has been identified as an important objective in promoting the long-term economic prosperity of the United States (Bureau of Labor Statistics, 2014; Committee on Prospering in the Global Economy of the 21st Century, 2007; Fayer, Lacey, & Watson, 2017; National Research Council, 2011). In addition to increasing the scope of the STEM workforce, diversity in STEM fields has received prominent attention in recent research and public policy discussions (Carnevale, Fasules, Porter, & Landis-Santos, 2016; Sass, 2015; White House, 2016).

Improved access to STEM courses in high school has been advocated as a lever by which the STEM workforce can be expanded and diversified. Postsecondary STEM outcomes are intermediary – the idea is that exposure to more, and more-advanced, STEM courses in high school will lead to more interest and success in STEM in college, which in turn will translate to a more robust STEM workforce. Calls for improved access to STEM coursework in high school, and especially improved access at schools that primarily serve under-represented minorities, have come from policy and advocacy groups, journalists, and the highest levels of government. For example, the Obama Administration’s “STEM for All” campaign argued, among other things, that “For high-school students, access to core and advanced STEM coursework is an essential part of preparing to enter the workforce equipped with relevant skills for a broad range of jobs, and to successfully pursue STEM degrees and courses in college” (White House, 2016).<sup>1</sup>

---

<sup>1</sup> Also see guidance from the President’s Council of Advisors on Science and Technology (2010), which recommends expanding the availability of advanced STEM courses in high school. Two other recent examples are,

The academic literature has devoted considerable attention to studying the effects of STEM courses in high school, measured in terms of both access and direct course-taking, on later-life outcomes. A well-established empirical regularity is that the more STEM courses a student takes during high school, the higher her likelihood of STEM enrollment and degree attainment in college (e.g., see Long, Conger & Iatarola, 2012; Maltese and Tai, 2011; Sadler & Tai, 2007). However, the endogeneity of students' own course-taking behaviors makes causal inference difficult – unobserved preference or endowment heterogeneity may lead to both the pursuit of technical courses in high school and college STEM outcomes.

Economists working in this area have typically skipped over intermediary postsecondary outcomes, such as STEM degree attainment, focusing instead on understanding how high school curricula influence labor market outcomes. Notable studies include Altonji (1995), Levine and Zimmerman (1995), and Rose and Betts (2004). These studies face the same fundamental concern about the endogeneity of students' own course-taking decisions. In recognition of this concern, the authors favor models that link variation in high school course offerings, irrespective of the courses that individual students take, to longer-term outcomes. While helpful, a remaining and well-understood endogeneity concern is that the course offerings of a high school may be related to student sorting to schools, and possibly other resources and opportunities, which can also affect college outcomes.

While both of these endogeneity concerns have been well-articulated previously, to the best of our knowledge they have not been simultaneously addressed. To be more specific, while previous studies mitigate the threat of endogeneity from individual student choices over courses, they are unable to fully address the recognized threat of endogenous course offerings across high

---

among policy and advocacy groups: Randazzo (2017); and in the media: Deruy (2016), which is motivated by a report from the U.S. Department of Education's Office of Civil Rights.

schools because they rely on cross-sectional data. An innovation of our study is the construction of a 14-year data panel of entrants into Missouri's 4-year public university system, merged with administrative records on course offerings at the high school level. These data facilitate a high school fixed effects identification strategy that leverages variation within high schools over time in course offerings, allowing us to address both endogeneity threats.

Our within-high-school identification strategy improves on available research but raises two issues. First, we sacrifice statistical power by isolating within-high-school variance in course access for identification. However, this concern is of limited practical importance in our study because of our large sample and the non-negligible within-high-school variance share of course offerings.<sup>2</sup> Our standard errors are small enough to permit meaningful inference. The second issue is the potential for endogenous changes in course offerings within high schools over time. Over the full 14 years of our data panel such changes seem plausible – e.g., a compositional shift in a neighborhood might induce a change in the high school curriculum driven by shifting student interests. However, over shorter time intervals, variation in course offerings within a high school is more likely to be driven by idiosyncratic shocks. Examples include changes to personnel and rigidities in the functions that map course offerings to enrollment within schools (i.e., rules, implicit or explicit, governing how many sections of a course are offered based on projected enrollment and class sizes). Although we lack data on the programmatic details that drive curriculum changes to isolate specific channels, we test indirectly for evidence of bias from endogenous changes within high schools over time by partitioning our long data panel into a series of shorter panels. Within the shorter panels systematic, endogenous shifts within high schools are

---

<sup>2</sup> As noted below, 17 percent of the variance in course access in our data panel occurs within high schools over time.



less likely. This exercise provides no indication that our findings are affected by bias from this type of endogeneity.

We show that expanded access to STEM courses in high school does not increase postsecondary STEM enrollment or degree attainment. Our estimates are substantively small and precise. Point estimates from our preferred models imply effect sizes of a one-standard-deviation increase in high school course access on postsecondary STEM enrollment and attainment of just 0.10-0.15 percentage points. With 95 percent confidence, we can rule out effects larger than 3-5 percent of the sample means for these outcomes. Our null findings are robust to numerous measurement modifications. They persist if we separately estimate the effects of access to math and science courses, and access to math courses that differ by the content level (regular or advanced).

We also show that there is no detectable effect heterogeneity across high schools that differ by the racial/ethnic minority share of the student body. Policy proposals to expand STEM course access at high schools with high minority shares reflect the concern that a lack of access affects students who attend these schools specifically. However, our results indicate that access to courses alone is not the problem. There *is* evidence of modest effect heterogeneity by gender and race within high schools, but the heterogeneity is not in a direction that suggests increases in course access will reduce postsecondary STEM outcome gaps – male and white students are marginally more responsive to changes in course access relative to women and underrepresented minority students. On the whole, these results indicate that simply increasing the number of math and science courses offered in high school is unlikely to change the demographic distribution of college STEM degrees in the direction intended by policy.

Finally, as a complement to our reduced-form analysis of course access, we explore models that aim to identify the causal effect of course-taking in high school. We use course access as an instrument for courses taken by individual students – a more assumptive modeling structure that we discuss in detail below. An insight from the instrumental variables models is that while increases in course access correspond positively to increases in course-taking for individual students, the mapping is substantively weak. The weak link between course access and course-taking, which is presumably a first-order pathway by which increased access would be expected to affect postsecondary STEM outcomes, is not consistent with the presence of widespread excess demand for math and science courses in high school.<sup>3</sup>

## 2. Empirical approach

We build on the methodological structure used by Altonji (1995), Levine and Zimmerman (1995), and Rose and Betts (2004). First, consider the following cross-sectional regression linking student course-taking in high school to subsequent outcomes:

$$Y_{is} = \beta_0 + X_i\beta_1 + Z_s\beta_2 + C_{is}\beta_3 + \varepsilon_{is} \quad (1)$$

In our application  $Y_{is}$  is a postsecondary STEM outcome – either enrollment in a STEM major or the completion of a STEM degree – for student  $i$  who attended high school  $s$ .  $X_i$  and  $Z_s$  are vectors of observed individual and high-school variables, respectively, from which  $C_{is}$  is separated for presentational convenience as it denotes the treatment of interest: the number (and sometimes type)

---

<sup>3</sup> Because variation in course access is such a weak predictor of course-taking, our study is ultimately uninformative about policies that require additional course-taking explicitly. Evidence on the effects of mandatory course-taking is mixed. Studies suggest short-term academic benefits but evidence on longer-term outcomes is less promising since such initiatives can induce dropout (e.g., Allensworth, Lee, Montgomery, & Nomi, 2009; DiCicca & Lillard, 2001; Cortes, Goodman, & Nomi, 2015; Jacob, Dynarski, Frank, & Schneider, 2017; a related literature examines high school exit exams and similarly finds negative effects on graduation: e.g., Jacob, 2001; Jenkins, Kulick, & Warren, 2006; Papay, Murnane, & Willett, 2010 ). The negative effects documented in some studies of course mandates make policies that expand course access without mandatory course-taking appealing.

of STEM courses taken in high school by student  $i$ . The elements of the  $X$ -vector include student-level race/ethnicity and gender, ACT math and English scores, and high school class rank.<sup>4</sup> The  $Z$ -vector contains time-varying high school characteristics including total enrollment, the share of the student body that identifies as a minority race/ethnicity, and the share of the student body that is free or reduced-price lunch eligible.  $\varepsilon_{is}$  is an error term.

We would like to interpret  $\beta_3$  as the causal effect of STEM course-taking in high school on STEM outcomes in college. However, causal inference is problematic because  $C_{is}$  is likely endogenous. As noted above, key sources of endogeneity include (a) within a high school, variation in individual student course-taking behavior will be driven by unobserved factors that also influence college outcomes, and (b) across high schools, variation in course offerings is likely correlated with factors that are difficult to measure and also affect STEM outcomes in college, such as unobserved student attributes (e.g., from Tiebout sorting) and school resources (e.g., teacher quality, facilities).

We address the first issue by substituting a measure of the courses offered by the high school for actual courses taken by each student:

$$Y_{is} = \delta_0 + X_i \delta_1 + Z_s \delta_2 + CA_s \delta_3 + u_{is} \quad (2)$$

Equation (2) is the same as Equation (1) except for the substitution of  $CA_s$  for  $C_{is}$ , where  $CA_s$  measures the *courses available* at high school  $s$  during student  $i$ 's high school career. In the absence of access to administrative data on course offerings from schools, Altonji (1995), Levine

---

<sup>4</sup> Students' class ranks and ACT scores are determined during the treatment window (high school). A concern is that including these variables could dull the estimated coefficients of course access and course-taking. In recognition of this concern, we have estimated our models that exclude these control variables and confirmed that the results we show below are robust (results available upon request). We prefer the models that include the full suite of control variables for students because they improve precision with no indication that they substantively influence the parameters of interest.

and Zimmerman (1995), and Rose and Betts (2004) construct proxy measures of  $CA_s$  based on the course-taking behaviors of a student's peers within the same school. Our study offers a modest data improvement in that we have access to administrative records on annual course offerings of high schools from the Missouri Department of Elementary and Secondary Education (see below for more details about the data).<sup>5</sup>

The advantage of the model in Equation (2) is that  $CA_s$  does not incorporate variation from student  $i$ 's own course-taking behavior to identify  $\delta_3$ . The substitution of  $CA_s$  for  $C_{is}$  also implies a shift in the interpretation of the focal parameter. Assuming other problems away,  $\delta_3$  can be interpreted as the effect of exposure to course offerings at the high school level, rather than actual courses taken. Thus,  $\delta_3$  has an “intent-to-treat” (ITT) interpretation. In addition to being informative about the treatment effect subject to the complier rate, the ITT parameter is of direct policy interest given that prominent proposals to increase the scope and diversity of the STEM workforce have focused on expanding course access in high school.

While identification is improved in Equation (2) relative to Equation (1), Equation (2) relies on variation across high schools in course offerings for identification. Altonji (1995), Levine and Zimmerman (1995), and Rose and Betts (2004) recognized this limitation but had access only to cross-sectional data and thus were unable to address it fully. A straightforward but important innovation of our study is the construction of a long data panel of high schools, which allows for an improved methodological approach (as advocated by Altonji, Blom, and Meghir, 2012).

---

<sup>5</sup>  $CA_s$  will be measured with error for mobile students during the late high school years because we cannot link individual students in the postsecondary and K-12 data systems, and thus cannot track individual mobility during high school. High school assignments are determined by the high school from which students graduated as coded in the higher education data system. This limitation is not unique to our study – it is also relevant for aforementioned prior studies that measure course access using peers' course-taking.

Specifically, our preferred models are panel-data versions of Equation (2) that leverage our ability to observe multiple cohorts of students graduating from high schools over time:

$$Y_{ist} = \gamma_0 + X_{it}\gamma_1 + Z_{st}\gamma_2 + CA_{st}\gamma_3 + \theta_s + \tau_t + e_{ist} \quad (3)$$

Equation (3) builds on Equation (2) with the addition of a time dimension, indexed for cohorts of high school graduates by  $t$ . Correspondingly, we can include high school and year fixed effects in the model,  $\theta_s$  and  $\tau_t$ , respectively. The parameter of interest in Equation (3) is  $\gamma_3$ , which is identified using variation in course offerings over time within high schools conditional on the sample-wide time trend captured by  $\tau_t$ . Our standard errors are clustered by high school throughout.

The concern that endogenous course offerings across high schools will bias the results is mitigated by Equation (3). As noted previously, the remaining threat to identification is endogenous changes to course offerings within high schools over time. Below we show results from a test designed to detect bias from such changes and we do not find evidence of bias. Equation (3) is our preferred specification for estimating the causal effects of access to STEM courses in high school.

In addition to the policy relevance of our reduced-form ITT parameter,  $\gamma_3$ , another desirable feature is that it allows for multiple pathways by which course offerings in high school can affect postsecondary STEM outcomes. For example, in addition to the intuitive first-order pathway of inducing students to take more STEM courses, increasing the number of high school STEM courses could also affect students by affecting their peers, changing the culture of a high school, and/or affecting teacher retention and recruitment, among other possibilities. These types of indirect effects can influence students above and beyond the effect of changing their own course-taking behavior and will be captured by  $\gamma_3$ .

Still, models that aim to identify the direct effect of course-taking are also of interest. By imposing more restrictive assumptions on the causal pathway, we can recover treatment effects of course-taking using an instrumental variables (IV) approach. The IV approach uses variation in course offerings within a high school over time to instrument for courses taken by students, and in turn use the instrumented values to estimate the effects of courses taken, as follows:

$$C_{ist} = \pi_0 + X_{it}\pi_1 + Z_{st}\pi_2 + CA_{st}\pi_3 + \phi_s + \rho_t + e_{1ist} \quad (4)$$

$$Y_{ist} = \alpha_0 + X_{it}\alpha_1 + Z_{st}\alpha_2 + \hat{C}_{ist}\alpha_3 + \psi_s + \eta_t + e_{2ist} \quad (5)$$

To support a causal interpretation,  $CA_{st}$  must be excludable from Equation (5) conditional on the other controls. Beyond the exogeneity conditions required for Equation (3), this additionally requires we assume that there are no indirect effects of  $CA_{st}$  on high school students through channels other than course-taking. Given this strict requirement, the estimation framework in Equations (4) and (5) can be used to obtain what is likely an upper-bound estimate of the course-taking effect. This is because any indirect effects of course access would likely generate upward bias by attributing correlated indirect effects to the course-taking mechanism. As a specific example, if more available STEM courses encourage a student's peers to take more courses, and this in turn influences her own interest in STEM, the indirect effect will be embodied in  $\alpha_3$  along with any direct effect on her own course-taking behavior.

The first stage of the IV model also proves useful for contextualizing our findings. As we show below, exposure to additional STEM courses in high school is a statistically significant predictor of the number of STEM courses taken. However, the substantive mapping of course exposure to course-taking is weak. The weak link between course access and course-taking helps to explain our primary finding that access to more STEM courses in high school does not improve STEM outcomes in college.

### 3. Data and Context

Our student records are from administrative microdata provided by the Missouri Department of Higher Education (DHE). We focus on full-time, first-time students who graduated from a Missouri public high school and matriculated to a 4-year Missouri public university within two years of completing high school. Our data panel covers over 140,000 students from 14 cohorts of entrants into the public university system between 1996 and 2009.

By virtue of using students in the DHE data to define our sample, our analysis necessarily conditions on university enrollment. This means that our microdata are ill-suited to examine effects on the extensive margin of college (i.e., attendance), but well-suited to examine compositional shifts in major choice and attainment conditional on entry into the 4-year public university system. Given that individuals who initially enroll in and complete STEM majors are positively selected among high school students, the latter margin is arguably the most important.<sup>6</sup> Still, the potential for effects of STEM course access on college enrollment potentially complicates the interpretation of our estimates. To get a sense of the importance of this concern, in supplementary models shown in Appendix Table A1 we regress high schools' annual 4-year college matriculation counts on high school fixed effects and STEM course access, conditional on high school enrollment – these regressions match the basic structure of Equation (3). There is no indication that changes to STEM course access within high schools over time affect the college matriculation rate. We therefore conclude that as a practical matter this is not an important concern, supporting our focus on the compositional shift between STEM and non-STEM fields among college entrants.<sup>7</sup>

---

<sup>6</sup> The average class rank of university entrants in our sample is in the 70th percentile; among STEM entrants the average class rank is in the 77th percentile.

<sup>7</sup> The estimated effect on matriculation is not statistically significant and the magnitude of the implied effect is trivial (Appendix Table A1). Moreover, given that we ultimately find null effects of expanded access to STEM courses in high school on college STEM outcomes, the most prominent biasing concern is that expanded high school STEM access spurs college enrollment, but not in STEM, and an effect of this particular nature – on college attendance but not in STEM – is unintuitive.

There are 13 public 4-year universities in Missouri as listed in Appendix Table A2. STEM education is highly concentrated within the system, with nearly 60 percent of STEM graduates coming from just two universities: the state flagship University of Missouri-Columbia (35 percent of all STEM graduates) and the engineering-focused Missouri University of Science and Technology (24 percent of all STEM graduates, despite accounting for just five percent of system enrollees). Three other universities – the highly selective Truman State University and moderately selective Missouri Southern State University and University of Central Missouri – produce about seven percent of STEM graduates each; all other universities produce fewer than five percent of STEM graduates.<sup>8</sup>

We track each student in the Missouri system to determine whether she graduated within six years of entry (from any system school), and if so, her final major.<sup>9</sup> The DHE data also include detailed information about students' academic ability that we incorporate into our models – i.e., ACT math and English scores (following Bettinger, Evans & Pope, 2013) and high school class ranks. Moreover, the number of courses students take in each of several subjects during high school, including math and science, are taken from the DHE data. A “course” is defined as one year of course-taking. During our analysis period, students needed to complete two math courses and two science courses to meet minimum high school graduation requirements.<sup>10</sup> These are minimum requirements established by the State Board of Education; local board policies may include additional requirements. For students who intend to enroll in a public university in Missouri, the Coordinating Board for Higher Education (CBHE) recommends four mathematics

---

<sup>8</sup> Selectivity designations are based on the 2015 Carnegie Classifications of Higher Education. See <http://carnegieclassifications.iu.edu>.

<sup>9</sup> Some students will graduate after the six-year window, but we follow convention in the literature of using six years for our primary analysis. Results are qualitatively similar when using graduation rates as measured over seven or eight years (omitted for brevity).

<sup>10</sup> The state increased requirements to three math and science courses starting in 2010, after the timespan of our data panel.



courses.

We identify initial majors and degrees based on the Classification of Instructional Programs (CIP) taxonomy developed by the US Department of Education for college majors. The initial major is an “intended” major; there are no requirements or formal system rules that govern the initial selection. We classify each major as either STEM or non-STEM, with STEM including the following fields: engineering (7% of initial majors), biological science (6%), computer science (3%), physical science (2%), engineering technology (1%), agricultural and animal science (1%), mathematics (1%), and other STEM (1%).<sup>11</sup>

Figure 1 shows trends in STEM majors and degrees for the high school cohorts in our sample from 1996-2009. Initial interest in STEM remained relatively flat over most of the data panel but increased some in the later years; in total, initial STEM enrollment increased 20 percent, or roughly 5 percentage points, between 1996 and 2009. Similarly, STEM attainment increased by about 23 percent. The growth in women declaring STEM majors was similar to the overall rate, but the growth in attainment among women was slightly below the sample average. The trend in initial STEM enrollment among underrepresented minority students (defined here as black and Hispanic students) is noisier but grew at a similar rate to the overall trend. However, STEM degree attainment actually declined by 13 percent among these students.

We supplement the DHE data with a data panel of high school course offerings assembled using administrative records from the Missouri Department of Elementary and Secondary Education (DESE). In our preferred measure of course availability, each course section is treated as a separate course. For example, if a high school offers three sections of algebra-I in a single year, this counts as three math courses. We also adjust by high school enrollment to get measures

---

<sup>11</sup> Other STEM includes technical subfields of education; military technologies; psychology; social sciences, health professions, and management sciences.

of “courses available per 100 students.” For students in each cohort at each high school, we average the total number of (enrollment adjusted) courses offered in the high school from the year of the student’s graduation and the two years prior to construct our measure of exposure to math and science courses during high school. We use three years because some high schools span grades 9-12 and others span grades 10-12.

Figure 2 shows trends over time in access to high school STEM courses in Missouri. The black solid line represents all math and science courses available per 100 students – the trend is relatively flat, with a slight uptick by the end of the period (overall growth of about six percent from 1996 to 2009). We also separately plot advanced math, high school level math, and science courses. Dividing math and science courses is straightforward; to differentiate the content of math courses we coded each math course in the high-school data as either “advanced” (i.e., college prep and college level courses) or “standard” (i.e., high school-level math courses). The coding is based on the course title available in administrative records (we were unable to follow a similar process to classify science courses because of inconsistent reporting across high schools and years).<sup>12</sup> The overall growth in math and science courses is driven primarily by increases in advanced math offerings, with the average advanced math offering increasing nearly 14 percent. Over the analysis period, standard high school-level math course access stayed flat, while average science course access actually declines by about seven percent.

In sensitivity checks, we also use two other measures of course availability. The first is the total number of course offerings unadjusted for student enrollment. A larger number of course offerings may provide more access in an absolute sense, in a way that is missed by our enrollment-

---

<sup>12</sup> Specifically, we coded courses as either high school level, college preparatory level, or college level based on administrative course numbers, course grade-level (a standardized reporting of the year in school in which students typically take the class), sequence number (identifies content of courses that are taught at more than one level), and delivery system.

adjusted measures. The second is a measure of “topic availability,” which we also measure per 100 students. The “topic availability” measure does not count each course section of a course separately. For example, if a high school offers three sections of algebra-I in a single year, this counts as just one topic. The value of this alternative measure is best articulated by noting that our primary measure captures variation in access to STEM courses along two dimensions: (1) increased availability of seats holding the topic set fixed, and (2) increased availability of topics. The topic availability measure isolates variation along the latter dimension only.

We additionally collect data on high school characteristics from the Common Core of Data made available by the National Center for Educational Statistics (NCES). We merge the information about high schools to student records in the DHE data by high school and year.<sup>13</sup> The final merged dataset includes over 140,000 students who attended 498 public high schools and matriculated to one of the 4-year public universities in Missouri.

Summary statistics for students and high schools are reported in Table 1. The sample is 56 percent female and 85 percent white. Black students comprise 8 percent of the sample, and Hispanic and Asian students account for 2 percent each. High schools in Missouri are disproportionately rural, although student representation is more balanced across high school types than is implied by the high school characteristics because rural schools are small (see columns 3 and 4).

Table 2 shows that approximately 20 percent of initial and completed degrees at Missouri universities are in STEM fields.<sup>14</sup> Forty-three percent of students who declare a STEM major upon

---

<sup>13</sup> A notable variable is high school enrollment, which we use as a covariate in our fully specified models and to adjust our preferred course-availability measures. Given that our course-availability measures cover courses offered in grades 10-12, we use enrollment in grades 10-12 for consistency.

<sup>14</sup> Although student transfers out of STEM are higher than transfers into STEM, the STEM enrollment and attainment shares end up being similar because initial STEM majors graduate at higher rates.

entry graduate with a STEM degree, while just four percent of students who do not initially declare a STEM major complete a STEM degree. These simple statistics highlight the strong link between initial STEM enrollment and completion.

Anticipated differences by race/ethnicity and gender in STEM enrollment and attainment are also on display in Table 2. For example, the first two columns show that men are more than twice as likely as women to declare a STEM major and earn a STEM degree. Among races/ethnicities, Asian students are the most likely to initially declare a STEM major and complete a degree (31 and 21 percent, respectively), while black students are least likely (18 and 6 percent, respectively). Among those that declare STEM degrees; male, white, and Asian students are more likely (46, 44, and 51 percent, respectively) than female, black, and Hispanic students (38, 24, and 38 percent, respectively) to earn one.

Table 3 focuses on STEM course-taking and course exposure in high school. The first takeaway from Table 3 is that there is substantial variation in STEM courses taken, course availability, course availability per 100 students, and topic availability per 100 students in our data. The most relevant measure for our primary models is course availability per 100 students, where row-1, column-3 of the table shows that the mean value is 10.8 with a standard deviation of 3.6. Thus, the range of course exposure within one standard deviation of the mean is 7.2-14.4 STEM courses per 100 students on average during grades 10-12. Variation in course access within high schools over time – the variation we isolate for identification in our preferred models – accounts for 17 percent of the total variance of enrollment-adjusted course availability.<sup>15</sup>

---

<sup>15</sup> We decompose the variance in course availability per 100 students by regressing this variable on the vector of high school indicator variables. One minus the R-squared from the regression gives the share of the variance that occurs within high schools.

The remainder of Table 3 shows splits for the course-taking and course access measures by (a) STEM attainment status and (b) demographics. The large differences across demographic groups in terms of postsecondary STEM outcomes, shown in Table 2, are not apparent at nearly the same level when we focus on high school STEM exposure. It is the case that students who ultimately complete STEM degrees take more math and science courses in high school, but they have only a very slight advantage in course access (columns 2—5). Unsurprisingly, female and male students have similar course access (differentials would only be unexpected in the presence of gender segregation in high school, or substantial gendered selection into our sample from some high schools); perhaps more surprising is that female students take about the same number of math and science courses in high school as male students despite being much less likely to enroll in or complete a STEM degree in college. The splits by race/ethnicity show that Asian students take the most math and science courses relative to other groups and black students take the least. In terms of course exposure, black, Hispanic and Asian students attend high schools that offer more math and science courses than white students (column 2), but this is due to the overrepresentation of white students in small, rural schools. This can be seen in column 3, where racial differences in access largely disappear when measured by courses available per 100 students.<sup>16</sup> There are small differences by race/ethnicity in absolute topic availability, and more-pronounced differences when we adjust for enrollment. Column 5 shows that black, Hispanic, and Asian students have fewer topics available in enrollment-adjusted terms than white students.

---

<sup>16</sup> Average high school enrollment for white students is lower than that for black, Hispanic, and Asian students.

## 4. Results

### 4.1. Primary Findings

Table 4 presents results from linear probability models building up to our primary specification as shown in Equation (3). The outcome in the first vertical panel is an indicator variable for whether initial postsecondary enrollment is in STEM (columns 1-3) and the outcome in the second panel is an indicator for STEM degree completion (columns 4-6). Within each panel, the first row of estimates uses actual courses taken as the independent variable of interest and the second row uses courses available. Results using our preferred specification are reported in the second row of columns 3 and 6.

Starting with the first row of the table, we see strong relationships between courses taken and postsecondary STEM outcomes. The relationships are fairly stable across models. Recall that the baseline rates of initially choosing a STEM major and completing a STEM degree are 21 and 12 percent in our sample, respectively (per Table 2). Noting that a one-standard-deviation change in high school STEM course-taking in our sample is 2.9 courses (Table 3), these estimates imply a strong link between course-taking in high school and postsecondary STEM outcomes. However, given the concern about endogenous course selection by individual students, it is ill-advised to interpret the estimates in the first row of Table 4 as causal.

The second row shows results after replacing the courses-taken variable with courses available. Per above, our preferred specifications use courses per 100 students to measure access, but our findings are not qualitatively sensitive to using alternative measures (see Section 5 below). The estimates moderate substantially when we move to the models that use course access. This is attributable to two factors: (1) the removal of bias from endogenous student choices, and (2) the shift in interpretation to the ITT parameter. The general takeaway reading across the columns of

the second row is that the underspecified models indicate positive and sometimes statistically significant “effects” of course access in high school on postsecondary STEM outcomes. However, estimates from the full specification with high school fixed effects provide no such indication.

While our standard errors rise some when we move to the full specification, which is expected because we leverage less identifying variation, the null results are not driven by an increase in our standard errors. The estimates themselves are quite small in magnitude. Specifically, the point estimates for initial STEM enrollment and degree attainment, taken at face value, imply effects of one more course per 100 students on postsecondary STEM outcomes of 0.03-0.04 percentage points. These equate to about 0.1-0.2 percent of the baseline rates of STEM enrollment and completion of 21 and 12 percent, respectively.

Moreover, even at their upper bounds, the implied effects are modest at best. The upper bound of the 95-percent confidence interval for the course-access coefficient in the STEM enrollment model is about 0.20 percentage points; in the STEM attainment model it is 0.17 percentage points. These estimates correspond to one-unit increases in courses available per 100 students. Per Table 3, the standard deviation of this variable is 3.6. Multiplying the upper-bound point estimates by 3.6 gives upper bounds of a one-standard deviation increase in course access per 100 students, which are roughly 0.72 ( $3.6 \times 0.20$ ) and 0.61 ( $3.6 \times 0.17$ ) percentage points for STEM enrollment and attainment, respectively. These correspond to just 3.4 and 5.1 percent of the sample means of these outcomes.

The models in Table 4 use all high school math and science courses to measure STEM exposure. We next use separate measures to explore the potential for effect heterogeneity of exposure to math and science courses, and to math courses that differ by the level of content covered. Access to different types of courses, and in particular differential access to advanced

courses across demographic and socioeconomic groups, has received significant attention in research (e.g., see Conger, Long and Iatarola, 2009; Klopfenstein, 2004).

Table 5 shows results from models that permit effect heterogeneity between math and science courses, and between math courses by level. All results are from our full specification. There is no evidence that differential exposure to math or science courses separately in high school affects postsecondary STEM enrollment or attainment. Similarly, there are no differential effects of access to regular versus advanced math courses. The point estimates throughout Table 5 are small, fluctuate in sign, and none are close to statistically significant at conventional levels.

#### **4.2. Instrumental Variables Extension**

We now turn to the instrumental variables (IV) models described in Section 2. Under the more restrictive assumption that the only pathway by which increased course access in high school affects postsecondary STEM outcomes is by directly affecting students' own course-taking behaviors, the IV estimates can be interpreted as causal effects of course-taking. While the effect of course availability on students' own course-taking is a plausible first-order pathway for effect, we again note that to the extent that the exclusion restriction is violated, we would expect the IV estimates to be biased upward due to other positive benefits associated with more STEM course availability in high school, such as effects on peers (and vice versa for reduced availability).

A nice illustrative feature of the IV estimation is the first-stage analysis, where we regress students' own course-taking behaviors on course availability at the high school. If increased course-taking is the main pathway for effect, the strength of this mapping is critical to the overall effect of increasing course availability. Table 6 shows the first stage results for two versions of Equation (4), with results from the full version shown in column 2. High school course availability is a statistically significant predictor of individual student course-taking. However, it is not a strong



instrument. In column 1 the F-statistic is below the Stock and Yogo (2005) weak identification threshold value of 16 (10% maximal IV size). In our preferred model in column 2 it is even smaller, well below the conventional threshold for a weak instrument, raising concerns about bias and precision of the IV estimates.

Substantively, the first-stage results indicate that for every one-unit increase in courses available per 100 students on average per year of high school, a student's own cumulative course-taking increases by just 0.02-0.04 courses. To put this number in context we can perform a rough back-of-the-envelope calculation of the "conversion rate" of courses available, as measured in our models, to total courses taken during high school. If we assume that each class has a capacity of 20 students (around the average for math and science classes in our data), students are distributed across classes at random, classes are filled to capacity, and expansion into an extra STEM course does not crowd out any other STEM course for individual students (the latter two assumptions are essentially that there is excess demand for STEM courses), then a one unit increase in our measure of course availability during high school would be expected to increase the total number of STEM courses taken for an individual student by up to 0.60 courses.<sup>17</sup>

Our estimates in Table 6 fall well short of this level and in fact, they imply that expanded course access does very little to increase STEM course-taking. Put another way, our estimates suggest what is very close to a pure substitution with other STEM courses when STEM course offerings increase. Moreover, our sample conditions on 4-year public university enrollees, who are

---

<sup>17</sup> To elaborate briefly, at the upper bound with a course capacity of 20 students, if 20/100 students take each offered course and each course is accessible and not redundant, the simple expected increase in total courses taken during high school for a student who is exposed to one more course per year on average for three years is 0.60. A simple calculation of the lower bound is more difficult because pass-through can be affected by additional constraints, such as whether marginal courses fit into students' course sequences, students are otherwise eligible for courses, and whether new courses are on new topics. That said, if we use our "topic availability" measure of course access it is fairly easy to arrive at a lower bound of 0.20, and the appendix shows that our results are similar (and even weaker) using that measure in the first stage (see Appendix Table A4).

positively selected among high school students in Missouri (per Table 1, the average class rank in our sample is over the 70<sup>th</sup> percentile). If students in our positively selected sample are more interested in STEM coursework than the average high school student, the expected conversion rate would be higher in our sample than is implied by our simple back-of-the-envelope calculation.

Thus, while course-taking is technically responsive to course availability as indicated by the statistically significant estimates in Table 6, the level of responsiveness is modest and not consistent with the presence of widespread excess demand for math and science courses in high school. This result could reflect a lack of demand for more STEM coursework in high school unconditionally, and/or the effects of other constraints faced by high school students, such as requirements to take courses across many fields for high school graduation and college admittance. Regardless of the source, the weak first-stage estimates help to explain our null reduced-form results in Table 4.

The first-stage estimates also have implications for the interpretation of the reduced-form findings. While our investigation was initially motivated by an interest in what is best described as an extensive margin intervention, the lack of a behavioral response of students on the extensive margin means that the primary treatment experienced by most students is on the intensive margin, in the form of smaller STEM classes. This is not to say that our analysis isn't informative about the extensive margin, as the pass-through result is critical to understanding policies that aim to expand course access, but ex post it is useful context that the way that most students are affected by expanded course access is in the form of smaller STEM classes. Inadvertently, our reduced-form findings speak to the potential for policies aimed at reducing STEM class sizes in high school to affect postsecondary STEM outcomes.

For completeness we briefly present results from the second-stage IV regressions in Table 7. Given that our instrument is weak we can glean little insight from the findings. Even under the strict IV assumptions, there is not clear evidence that additional course-taking in high school improves postsecondary STEM outcomes, but large effects (positive or negative) cannot be ruled out. While the first-stage regressions are informative about our investigation of course-access effects, our study is ultimately not informative about the effects of high school course-taking.

## **5. Sensitivity**

### **5.1 Period Subgroups**

The key identification threat in our models is the potential for endogenous changes to course offerings within high schools over time. Because our results are primarily null, the main concern is negative bias, which might come about if, for example, high schools where STEM training or interest is trending downward respond by offering more courses, and vice versa. This would induce a negative correlation between courses available within high schools over time and subsequent STEM outcomes, which in turn could generate null results from our specifications even if STEM access in high school positively affects postsecondary STEM outcomes, all else equal. We do not view this type of biasing scenario as likely. Instead, it seems more likely that our estimates, if anything, would be biased upward because changes to STEM course offerings within high schools over time are likely positively correlated with changes to the quality of STEM training and/or STEM interest within a high school. Nonetheless, the general biasing threat merits attention; if for no other reason than from a mechanical standpoint, over a 14-year span many factors within a high school can change and we rely critically on the high school fixed effects for identification.

We test indirectly for the influence of potential bias from endogenous changes to course offerings over time by replicating our primary results using partitions of the full data panel. We

hypothesize that if bias from endogenous changes within high schools is present, model replications based on data that cover a shorter timespan will be less biased because there is less time for major changes. We would view substantial differences in our estimates when we go from using the full panel, to using just a portion of the panel, as a likely symptom of endogenous changes to course offerings within high schools over time.

Table 8 shows results from replications of our main model estimated on datasets that cut the data panel in half (columns 2 and 3) and into thirds (columns 4-6). For ease of comparison we re-produce our main estimates from Table 4 in column 1. The findings are generally consistent across the various partitions of the full data panel. The point estimates are small and statistically insignificant, with one exception (the coefficient in column 6 for the degree-attainment model is statistically significant at the 10 percent level), and they nominally flip sign in one case (initial-major model, column 5). Taken as a whole, we interpret the results in Table 8 as suggesting that endogenous changes to course offerings within high schools over time are unlikely to drive our null findings.

We also briefly mention a related test for this type of bias, in which we estimate models that include high-school specific linear time trends. This narrows the identifying variation further by isolating deviations from the trend for each high school over the timespan of the data panel. Given that our results even without the high school specific time trends are null, and that these models are more demanding from a statistical power perspective (i.e., our standard errors are larger), it is unsurprising that these models do not overturn our null findings (results omitted for brevity).

## 5.2 Alternative Measures of Course Access

We re-estimate our models using two other measures of course availability. The first is analogous to our preferred courses-per-100-students measure, but is unadjusted for student enrollment at the high school. This allows for the possibility that absolute course access is important regardless of the size of the student body. The second measure is the above-described topical measure – like our primary measure, it is adjusted into per-100-student units, but it does not double-count repeat courses as expanding STEM access. Results using these alternative measures are shown in Appendix Table A3. They are substantively very similar to our primary findings in Table 4.

We also estimate “first stage” regressions using the alternative measures of access, analogous to the models we report on in Table 6. The results are shown in Appendix Tables A4 and A5. The strength of the unadjusted course-availability instrument is similar to the enrollment-adjusted version. The first-stage regression for the topical availability measure shows an even weaker and statistically insignificant relationship between topical exposure and course-taking in high school. This leads us to believe that our measures that count repeat courses are preferable for measuring course access.

An explanation for the weak predictive power of our topical-availability instrument is that math and science courses on different topics may be viewed as substitutes by students attempting to satisfy various high school graduation and college requirements. For example, if a new science topic is offered in geology, but a student has already satisfied her science requirements by taking biology and chemistry, she may have limited interest in the course, and/or limited capacity to take it given other requirements that must be satisfied. In such a scenario, measures that privilege non-

repeat courses at the expense of fully measuring capacity will be less predictive of students' course-taking behaviors.

## **6. Effect Heterogeneity**

Next we consider the possibility that the effects of course access vary by the racial/ethnic composition of the high school. This might be expected if, for example, high schools with higher percentages of minority students offer less access to STEM courses, in which case we might expect greater response elasticities to changes in course offerings at these schools. The descriptive statistics in Table 3 provide no *prima facie* indication of this, but some heterogeneity in course access – particularly in narrow pockets of the distribution such as among very high minority-share high schools – could be obscured in Table 3.

We focus on minority students that are underrepresented in STEM fields as a group (black and Hispanic students) because of their importance to policy and due to sample size considerations in Missouri. We estimate separate models for three overlapping subsets of schools: those with underrepresented minority student shares above 25 percent, above 50 percent, and above 75 percent. The former group subsumes the latter groups, but not the reverse. The reason for the overlapping samples is that only a small fraction of Missouri high schools contain substantial minority student shares – e.g., the 75<sup>th</sup> percentile high school in the state distribution has a minority share of just 18 percent. The structure of our investigation allows us to balance our interest in examining effect heterogeneity across high schools that differ as much as possible along this dimension against the loss of statistical power as the sample shrinks.

Table 9 shows results from our courses-taken and courses-available models akin to Table 4. For brevity, we only report findings from the fully specified models. As shown in the top panel of Table 9, like with the estimates from the full sample, we estimate a strong positive relationship

between STEM courses taken in high school and initial enrollment in a STEM field. The magnitudes of the estimates are somewhat smaller in Table 9 than in Table 4, but lead to a similar conclusion. In contrast, the results from the degree attainment models in the last three columns, even when we use courses taken as the independent variable of interest, are much weaker than what we show for the full sample in Table 4 and not statistically significant.

High attrition rates from STEM fields have been well documented, as have differential attrition rates by race/ethnicity (e.g., National Science Foundation, 2012). A potential explanation for the racial/ethnic attrition gaps suggested by previous research is that different groups are differentially prepared to succeed in STEM (e.g., Arcidiacono, Aucejo, & Spenner, 2012). Among students in high schools with large proportions of minority students, our results suggest that variation along at least this one dimension of preparation – high school STEM coursework – does not positively map to STEM success in college, even in models that embody endogeneity owing to students’ own course choices in high school.

Moving to the models of course access in the bottom panel of the table, where we have more causal purchase, there is no evidence that increased exposure to STEM courses in high school corresponds to improved STEM outcomes in college among students who attend high schools with a high proportion of minority students. If anything, the reverse is weakly suggested by the mostly negative point estimates, several of which are statistically significant or on the margin of being so. Inference is similar when using our other measures of course access and when we break out science and advanced/regular math courses (results available upon request).

Next, in Table 10 we examine effect heterogeneity by race/ethnicity and gender at the individual student level, within high schools. Following Table 9, for race/ethnicity heterogeneity we focus on comparing white students to black and Hispanic students. The models interact our

primary measure of course availability with indicators for students' genders and race/ethnicities. Male and white students are the omitted comparison groups, and thus effects for all other groups are relative to them. For brevity, we show results only for the fully-specified models of course access.

Column (1) shows results when the outcome is initial STEM enrollment. There is statistical evidence of effect heterogeneity by gender and race, but the magnitude is small to moderate. The coefficient for women of -0.13 percentage points, statistically significant at the 10 percent level, implies that a one-standard-deviation increase in course access during high school has an effect on STEM enrollment that is 0.47 percentage points lower relative to white men. For the race/ethnicity comparison, the -0.32 percentage point effect for underrepresented minority students relative to white men is somewhat larger and translates to a differential effect size of 1.2 percentage points, or 5.5 percent of the sample mean, for a one standard deviation increase in course access. When we turn to the model of degree attainment the race/ethnicity and gender gaps moderate and become statistically insignificant.

For both women and underrepresented minorities, and in both models, the overall effects of increased course access, inclusive of the main coefficient, are small and statistically insignificant. Moreover, the differential effects relative to white men are best described as small to moderate. Still, the direction of the findings is not encouraging about the prospects for using high school STEM access as a policy lever to promote STEM diversity. The results suggest that expanded course access in high school could modestly widen postsecondary STEM enrollment gaps by race and gender.<sup>18</sup>

---

<sup>18</sup> This is consistent with findings from Conger, Long and Iatarola (2009).



## 7. Conclusion

We use administrative microdata from Missouri covering 14 cohorts of entering postsecondary students to examine the effects of access to STEM courses in high school on STEM outcomes in college. STEM interest and success in college are intermediary outcomes on the path toward a larger and more-diverse STEM workforce. Using multiple measures of STEM course access in high school, including measures that separate exposure to advanced coursework in math, we consistently show that changes in course access do not causally affect postsecondary STEM outcomes.

Our preferred specifications focus on the reduced-form effects of course access. These models are conceptually appealing because they allow course access to improve student outcomes through additional pathways beyond direct course-taking. They are of interest from a policy perspective because a lack of available STEM courses in high schools has been postulated as a barrier to STEM entry and success in college (e.g., Deruy, 2016; President's Council of Advisors on Science and Technology, 2010; Randazzo, 2017; White House, 2016). Moreover, policies that modify access to STEM coursework would be fairly straightforward to implement by state and local education agencies, making them appealing in terms of feasibility, and create less risk for adverse unintended consequences than course-mandate policies (Allensworth, Lee, Montgomery, & Nomi, 2009; DiCicca & Lillard, 2001; Jacob, Dynarski, Frank, & Schneider, 2017).

We also instrument for high school course-taking using variation in course access. While our first stage is statistically significant in a technical sense, course access is a weak instrument for course-taking. The weak predictive power of course access over course-taking is inconsistent with pent-up demand for STEM course-taking in high school. It implies that when afforded more

access to STEM courses, high school students mostly substitute between other STEM courses. This helps to explain our null reduced-form findings for course access.

We explore the potential for effect heterogeneity across high schools that differ by the share of underrepresented minority students, and within high schools by student race and gender. Our analysis of effect heterogeneity across high schools is motivated by the concern that access to STEM coursework is more restricted in high-minority high schools, in which case we might expect students to be more responsive to changes. However, we find no evidence of effect heterogeneity along this dimension. We also examine effect heterogeneity by race and gender within high schools. Our large data panel allows for a well-powered analysis in which we find some statistically significant differences, but they are modest in magnitude. The estimates suggest that postsecondary STEM outcomes for female and underrepresented minority students are less affected by access to STEM courses in high school than white male students. The implication is that broad, untargeted efforts to expand STEM access in high school may modestly exacerbate current race- and gender-based imbalances in STEM fields.

We caution that our results may not be informative about changes in STEM course access outside of the range of observed values in our data. As an extreme example, our findings should not be taken to imply that reducing STEM access in high school to zero would have no effect on postsecondary STEM outcomes. And while our reliance on natural variation in “business as usual” course offerings within high schools over time for identification is appealing in some ways, as discussed above, it also limits the range of research questions we can answer. For example, interventions to improve the quality of high school STEM education on the intensive margin may offer more promise. It is not clear what characteristics of intensive-margin interventions would drive change (again, our results imply class-size reductions alone will likely be ineffectual), but

one possibility is the development of a deeper, more stable STEM curriculum, including a pipeline of STEM training that pre-dates high school enrollment. Variability in “stable” STEM curricula would occur mostly across high schools, making our estimation strategy ill-suited to speak to the potential effects. That said, evidence to date on more substantial STEM interventions, like STEM high schools, is not particularly promising (e.g., Wiswall et al., 2014). More broadly, changes on the intensive margin can be effective if the standard approach to STEM education in high school can be improved. Margins for improvement might include recruiting better teachers, changing student and teacher incentives, and improving STEM facilities and instructional materials.<sup>19</sup> But if such improvements were obvious and feasible, they would likely already be implemented. Moreover, efforts to improve STEM education will crowd out resources targeted toward other types of learning given educational budget constraints. Noting these challenges, the lack of effects of simple expansions in course access that we document here suggest that for high school STEM policies to be effective at promoting postsecondary STEM interest and success, the norm of high school STEM instruction will need to change.

---

<sup>19</sup> Evidence from Jackson (2010, 2014) suggests the use of incentives for teachers and students to spur advanced course-taking may be a promising option for improvement. However, the program Jackson studies is not targeted at STEM fields in high school and he does not focus on STEM outcomes in college, so the applicability of his findings to our context is uncertain.

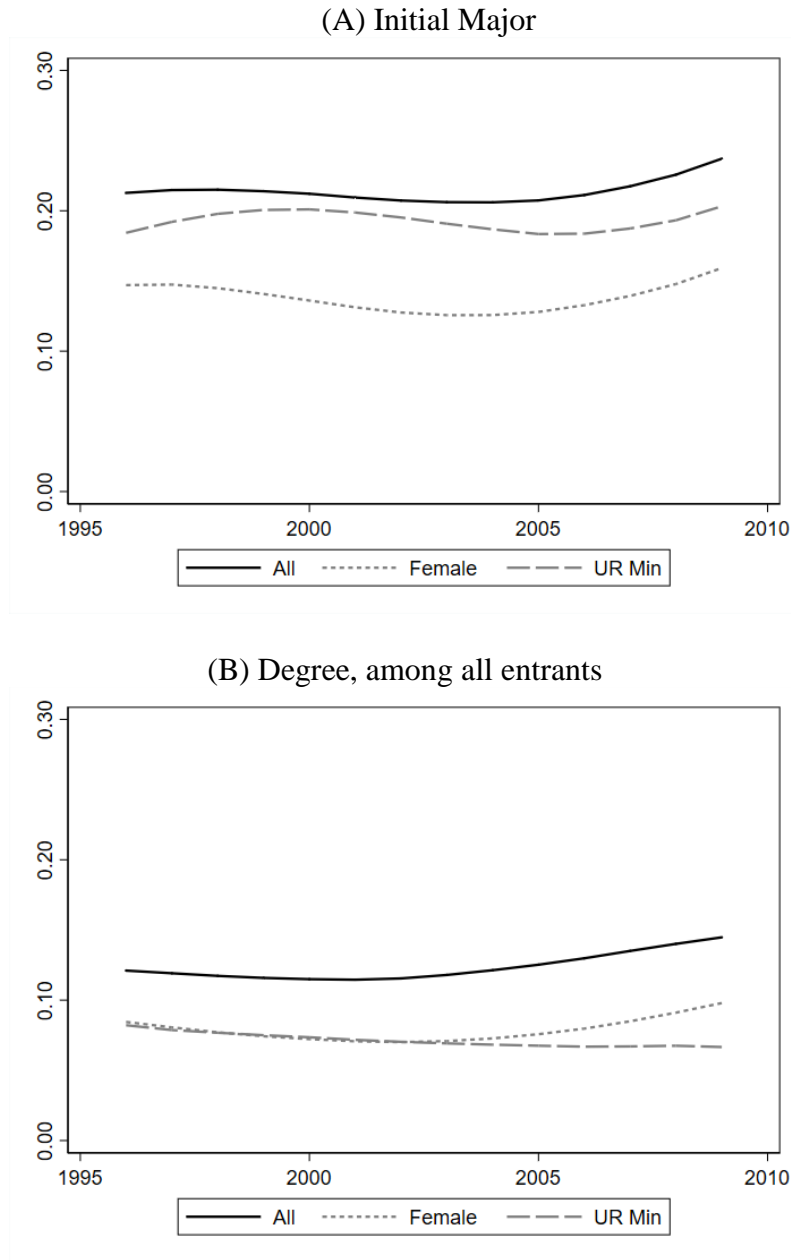
## References

- Allensworth, E., Lee, V. E., Montgomery, N., & Nomi, T. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis*, 31, 367-391.
- Altonji, J. G. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources*, 30(3), 409-438.
- Altonji, J.G., Blom, E., Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1), 185-223.
- Arcidiacono, P., Aucejo, E. & Spenner, K. (2012). What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice. *IZA Journal of Labor Economics*, 1(1), 1-24.
- Bettinger, E., Evans, B., & Pope, D. 2013. Improving College Performance and Retention the Easy Way: Unpacking the ACT Exam. *American Economic Journal: Economic Policy*, 5(2), 26-52.
- Bureau of Labor Statistics. 2014. STEM 101: Intro to Tomorrow's Jobs. *Occupational Outlook Quarterly* (Spring 2014). (www.bls.gov/ooq)
- Carnevale, A.P., Fasules, M.L., Porter, A., & Landis-Santos, J. (2016). African Americans: College majors and earnings. Washington, DC: Georgetown University Center on Education and the Workforce.
- Committee on Prospering in the Global Economy of the 21st Century (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington DC: The National Academies Press.
- Conger, D., Long, M.C., and Iatarola, P. (2009). Explaining race, poverty and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management*, 28(4), 555-576.
- Cortes, K. E., Goodman, J., & Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra. *Journal of Human Resources*, 50, 108-158.
- DeCicca, P. P., & Lillard, D. R. (2001). Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, 20, 459-473.
- Deruy, E. (2016). Where calculus class isn't an option. *The Atlantic* (06.07.2016). Retrieved on 10.23.2017 at <https://www.theatlantic.com/education/archive/2016/06/where-calculus-class-isnt-an-option/485987/>

- Fayer, S., Lacey, A., & Watson, A. (2017). *STEM occupations: Past, present, and future*. Washington, DC: U.S. Bureau of Labor Statistics.
- Jackson, K. (2014). Do College-Preparatory Programs Improve Long-Term Outcomes? *Economic Inquiry* 52(1), 72-99.
- Jackson, K. (2010). A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program. *Journal of Human Resources* 45(3), 591-639.
- Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23, 99-121.
- Jacob, B., Dynarski, S., Frank, K., & Schneider, B. (2017). Are expectations enough? Estimating the effect of a mandatory college-prep curriculum in Michigan. *Educational Evaluation and Policy Analysis*, 39(2), 333-360.
- Jenkins, K. N., Kulick, R. B., & Warren, J. R. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis*, 28, 131-152.
- Klopfenstein, K. (2004). Advanced placement: Do minorities have equal opportunity? *Economics of Education Review*, 23, 115-131.
- Levine, P.B. & Zimmerman, D.J. (1995). The benefit of additional math and science classes for young men and women. *Journal of Business and Economic Statistics*, 13(2), 137-149.
- Long, M., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, 49(2), 285-322.
- Maltese, A.V. & Tai, R.H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM Among U.S. students. *Science Education*, 95(5), 877-907.
- Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, 32, 5-23.
- President's Council of Advisors on Science and Technology. (2010). *Prepare and inspire: K-12 education in science, technology, engineering, and math (STEM) education for America's future*. Washington, DC: Executive Office of the President of the United States.
- National Research Council (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Washington, DC: The National Academies Press.

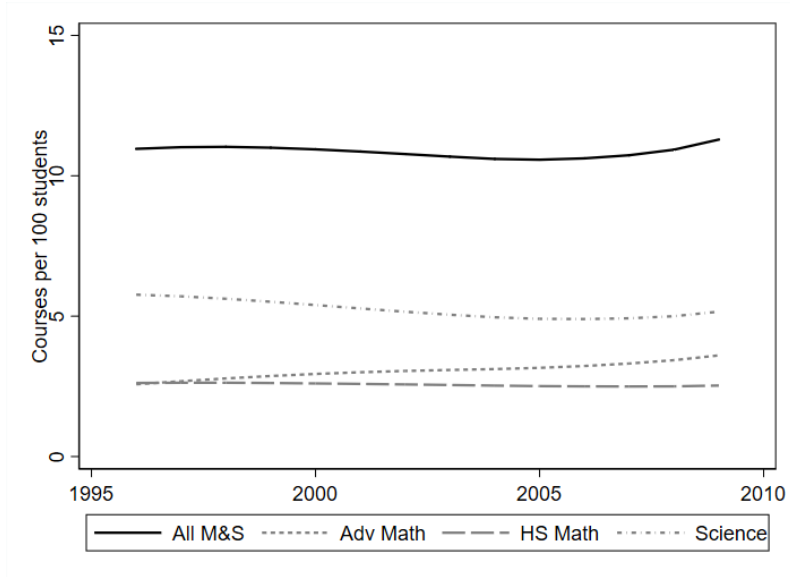
- Randazzo, M. (2017). Students shouldn't live in STEM deserts. *U.S. News & World Report* (05.10.2017). Retrieved on 10.23.2017 at: <https://www.usnews.com/opinion/knowledge-bank/articles/2017-05-10/the-us-must-address-disparities-in-access-to-stem-education>
- Rose, H. & Betts, J. R. (2004). The effect of high school courses on earnings. *The Review of Economics and Statistics*, 86, 497-513.
- Sadler, P. M., & Tai, R. H. (2007). The two high school pillars supporting college science. *Science*, 317, 457-458.
- Sass, T. (2015). Understanding the STEM Pipeline. CALDER Working Paper No. 125. Washington DC: CALDER.
- White House. (2016, February 11). STEM for all. [Blog post]. Retrieved from <https://obamawhitehouse.archives.gov/blog/2016/02/11/stem-all>.
- 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.

Figure 1: Trends in STEM Initial Major and Degrees by HS Cohort.



Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: Locally weighted smoothed line (lowess) line. X-axis is year of high school graduation. UR Min = Underrepresented minority student (black or Hispanic).

Figure 2: Trends in High School STEM Course Access by HS Cohort.



Source: Administrative data on Missouri public high school course offerings on average annually in grades 10-12., per 100 Students. Notes: Locally weighted smoothed line (lowess) line. X-axis is year of high school graduation.



Table 1: Sample Summary Statistics.

	Mean	SD	Mean	SD
<b>A. Students</b>				
Male	44%		--	--
Female	56%		--	--
White	85%		--	--
Black	8%		--	--
Hispanic	2%		--	--
Asian	2%		--	--
Other race/ethnicity	3%		--	--
Age at entry	18.1	0.4	--	--
HS Class Rank (Percentile)	70.8	22.5	--	--
ACT English	23.3	5.1	--	--
ACT Math	22.7	4.7	--	--
Number of Students	141,579			
<b>B. High Schools</b>				
	<u>School-year weighted</u>		<u>Student weighted</u>	
Graduates	102.2	108.6	261.7	160.2
Enrollment	366.7	386.3	889.5	537.5
Minority %	10%	20%	14%	19%
Free and Reduced Price Lunch	29%	16%	22%	15%
Urban	8%		18%	
Suburban	13%		32%	
Rural	80%		50%	
Number of High Schools	498			

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: All numbers are annual.

Table 2: STEM Initial Majors and Degrees.

	Initial STEM major, all entrants	STEM Degree, all entrants	STEM Degree, all graduates	STEM Degree, Initial STEM majors	STEM Degree, Initial non-STEM majors
All Students	21%	12%	20%	43%	4%
Male	31%	18%	31%	46%	6%
Female	14%	8%	12%	38%	3%
White	21%	13%	20%	44%	4%
Black	18%	6%	16%	24%	2%
Hispanic	22%	11%	20%	39%	3%
Asian	31%	21%	34%	51%	8%

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: Degree reflects degree acquisition in six years.

Table 3: Math and Science Courses Taken and Course Availability in High School.

	Courses Taken	Course Availability	Course Availability Per 100 Students	Topic Availability	Topic Availability Per 100 Students
All students	7.1 (2.9)	85.2 (47.4)	10.8 (3.6)	24.5 (8.3)	4.5 (4.2)
STEM degree recipients	8.3 (3.1)	88.1 (47.7)	10.7 (3.5)	24.9 (8.2)	4.3 (4.1)
Non-STEM degree recipients	7.3 (3.1)	87.9 (48.0)	10.7 (3.5)	24.8 (8.2)	4.4 (4.1)
Male	7.2 (2.8)	86.2 (47.1)	10.7 (3.4)	24.6 (8.2)	4.4 (4.1)
Female	7.1 (2.9)	84.4 (47.6)	10.9 (3.6)	24.3 (8.4)	4.6 (4.3)
White	7.1 (2.9)	83.3 (47.7)	10.9 (3.6)	24.3 (8.1)	4.7 (4.3)
Black	6.6 (2.2)	97.2 (43.0)	10.7 (3.8)	24.6 (9.9)	3.1 (2.6)
Hispanic	7.1 (2.7)	95.7 (43.7)	10.3 (2.8)	26.3 (8.0)	3.6 (3.1)
Asian	8.2 (2.9)	103.2 (42.1)	10.0 (2.5)	26.6 (8.2)	3.1 (2.4)

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: Standard deviation in parentheses.

Table 4: STEM Major and Degree Attainment Models.

	Initial Major			Degree Attainment		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>A. Courses Taken</i>						
Courses Taken	0.0137	0.0142	0.0142	0.0060	0.0060	0.0057
	(0.0008)***	(0.0008)***	(0.0007)***	(0.0005)***	(0.0005)***	(0.0005)***
<i>B. Course Availability</i>						
CA Per 100	0.0012	0.0002	0.0004	0.0002	0.0011	0.0003
	(0.0005)**	(0.0006)	(0.0008)	(0.0003)	(0.0003)***	(0.0007)
Individual controls & Year FE	X	X	X	X	X	X
HS Controls		X	X		X	X
HS FE			X			X
N	141,579	141,579	141,579	141,579	141,579	141,579

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: CA Per 100 = courses available per 100 students. Each coefficient is from a separate regression. All models control for high school graduation year (year fixed effects). Student controls are race/ethnicity, ACT math and English scores, and high school class rank. High school controls include location (urban, suburban, or rural; this factor drops out with the inclusion of HS fixed effects), enrollment, percent of the student body that identifies as a minority race/ethnicity, and percent of the student body which is free or reduced price lunch eligible. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Table 5: STEM Major and Degree Attainment Models, with Course-Type Heterogeneity. Courses Available Only.

	Initial Major				Degree Attainment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Advanced Math CA Per 100	-0.0014 (0.0023)	-0.0013 (0.0023)		-0.0019 (0.0024)	0.0006 (0.0017)	0.0004 (0.0017)		-0.0000 (0.0018)
Standard Math CA Per 100		0.0007 (0.0021)		0.0003 (0.0022)		-0.0011 (0.0018)		-0.0014 (0.0019)
Science CA Per 100			0.0009 (0.0011)	0.0011 (0.0012)			0.0007 (0.0010)	0.0009 (0.0011)
Individual controls & Year FE	X	X	X	X	X	X	X	X
HS Controls	X	X	X	X	X	X	X	X
HS FE	X	X	X	X	X	X	X	X
N	141,579	141,579	141,579	141,579	141,579	141,579	141,579	141,579

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: CA Per 100 = courses available per 100 students. All models control for high school fixed effects, student race/ethnicity, student ACT math and English scores, student high school class rank, enrollment in the high school, percent minority in the school, percent free/reduced price lunch in the school, and high school graduation year (year fixed effects). Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Table 6: Results from First Stage Regressions of Course Taking on Course Availability.

	(1)	(2)
CA Per 100	0.0411	0.0208
	(0.0108)***	(0.0091)**
Kleibergen-Paap LM statistic p-value	0.00	0.03
Kleibergen-Paap Wald F-statistic	14.42	5.22
Individual & HS controls & Year FE	X	X
HS FE		X
N	141,579	141,579

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: CA Per 100 = courses available per 100 students. All models control for high school graduation year (year fixed effects). Student controls are race/ethnicity, ACT math and English scores, and high school class rank. High school controls include location (urban, suburban, or rural; this factor drops out with the inclusion of HS fixed effects), enrollment, percent of the student body that identifies as a minority race/ethnicity, and percent of the student body which is free or reduced price lunch eligible. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Table 7: STEM Major and Degree Attainment Models, 2SLS Estimates.

	Initial Major		Degree	
	(1)	(2)	(3)	(4)
Instrumented Courses Taken	0.0049	0.0205	0.0274	0.0129
	(0.0142)	(0.0379)	(0.0095)***	(0.0322)
Individual & HS controls & Year FE	X	X	X	X
HS FE		X		X
N	141,579	141,579	141,579	141,579

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: CA Per 100 = courses available per 100 students. All models control for high school graduation year (year fixed effects). Student controls are race/ethnicity, ACT math and English scores, and high school class rank. High school controls include location (urban, suburban, or rural; this factor drops out with the inclusion of HS fixed effects), enrollment, percent of the student body that identifies as a minority race/ethnicity, and percent of the student body which is free or reduced price lunch eligible. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Table 8: STEM Major and Degree Attainment Models, Various Time Periods. Courses Available Only.

	All Years	Split Panel in Half		Split Panel in Thirds		
		1996-2002	2003-2009	1996-2000	2001-2005	2006-2009
	(1)	(2)	(3)	(4)	(5)	(6)
<i>A. Initial Major</i>						
CA Per 100	0.0004	0.0008	0.0012	0.0012	-0.0025	0.0005
	(0.0008)	(0.0014)	(0.0012)	(0.0018)	(0.0021)	(0.0020)
<i>B. Degree</i>						
CA Per 100	0.0003	0.0006	0.0014	0.0004	0.0027	0.0022
	(0.0007)	(0.0010)	(0.0009)	(0.0015)	(0.0017)	(0.0013)*
Indiv. & HS controls & Year FE	X	X	X	X	X	X
HS FE	X	X	X	X	X	X
N	141,579	69,166	72,413	48,411	51,370	41,798

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: CA Per 100 = courses available per 100 students All models control for high school fixed effects, student race/ethnicity, student ACT math and English scores, student high school class rank, enrollment in the high school, percent minority in the school, percent free/reduced price lunch in the school, and high school graduation year (year fixed effects). Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10



Table 9: STEM Major and Degree Attainment Models, by High School Racial/Ethnic Composition.

	Initial Major			Degree Attainment		
	Minority > 25%	Minority > 50%	Minority > 75%	Minority > 25%	Minority > 50%	Minority > 75%
<i>A. Courses Taken</i>						
Courses Taken	0.0084	0.0082	0.0114	0.0014	0.0008	0.0015
	(0.0012)***	(0.0013)***	(0.0018)***	(0.0010)	(0.0012)	(0.0017)
<i>B. Course Availability</i>						
CA Per 100	-0.0022	-0.0061	-0.0061	-0.0018	-0.0005	0.0005
	(0.0014)	(0.0036)*	(0.0044)	(0.0007)**	(0.0014)	(0.0014)
Indiv. & HS controls & Year FE	X	X	X	X	X	X
HS FE	X	X	X	X	X	X
N	17206	8959	4069	17206	8959	4069

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: The high school underrepresented minority shares are calculated as the sample average enrollment shares of black plus Hispanic students from NCES data, covering all students and over our full data panel. CA Per 100 = courses available per 100 students All models control for high school fixed effects, student race/ethnicity, student ACT math and English scores, student high school class rank, enrollment in the high school, percent minority in the school, percent free/reduced price lunch in the school, and high school graduation year (year fixed effects). Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Table 10: STEM Major and Degree Attainment Models, with Race/Ethnicity Heterogeneity. Courses Available Only.

	Initial Major	Degree
	(1)	(2)
CA per 100	0.0017	0.0006
	(0.0011)	(0.0008)
CA per 100 X Female	-0.0013	-0.0001
	(0.0008)*	(0.0007)
CA per 100 X Underrepresented Minority	-0.0032	-0.0011
	(0.0011)***	(0.0008)
Indiv. & HS controls & Year FE	X	X
HS FE	X	X
N	141,579	141,579

Source: Administrative data on white, black, and Hispanic Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: CA Per 100 = courses available per 100 students All models control for high school fixed effects, student race/ethnicity, student ACT math and English scores, student high school class rank, enrollment in the high school, percent minority in the school, percent free/reduced price lunch in the school, and high school graduation year (year fixed effects). Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

## Appendix: Supplementary Tables

Appendix Table A1: The Effect of STEM Course Access on the Number of Students Matriculating to the 4-year Public University System.

	Number of Matriculating Students
CA Per 100	0.024
	(0.034)
HS controls & Year FE	X
HS FE	X
Dependent Variable Mean (Standard Deviation)	21 (30)
N (high-school-by-year)	6644

Notes: This model is estimated at the level of the high school cohort (i.e., high-school and graduation year). The dependent variable is the number of students from who matriculated to a 4-year public university in Missouri. CA Per 100 = high school STEM courses available per 100 students, which is the treatment variable of interest in the main text. The model includes high school and year fixed effects and controls for enrollment in the high school, percent minority in the high school, and percent free/reduced price lunch in the school. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Appendix Table A2. Public 4-Year Universities in Missouri

University	Enrollment Entry Share	Graduation Rate	STEM % of Entrants	STEM % of Graduates
Overall	1.00	0.61	1.00	1.00
Univ of Missouri-Columbia	0.27	0.73	0.33	0.35
Univ of Missouri -Rolla	0.05	0.70	0.21	0.24
University of Central Missouri	0.09	0.58	0.07	0.07
Southeast Missouri State Univ	0.09	0.55	0.05	0.05
Western Missouri State Univ	0.06	0.35	0.04	0.02
Missouri Southern State Univ	0.17	0.60	0.08	0.07
Northwest Missouri State Univ	0.06	0.58	0.03	0.04
Univ of Missouri -St. Louis	0.03	0.48	0.02	0.02
Univ of Missouri -Kansas City	0.04	0.51	0.03	0.04
Lincoln Univ	0.03	0.30	0.02	0.01
Truman State Univ	0.08	0.78	0.09	0.07
Missouri State Univ	0.04	0.41	0.02	0.02
Harris Stowe State Univ	0.01	0.15	0.00	0.00

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: Ordered by the number of STEM graduates.

Appendix Table A3: STEM Major and Degree Attainment Models. Courses Available Only; Alternative Measures of Course Availability in High School.

	(1)	(2)	(3)	(4)	(5)	(6)
<i>A. Courses Availability</i>						
Unadjusted course availability	-0.0002	0.0000	0.0001	0.0000	0.0001	0.0000
	(0.0000)***	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
<i>B. Topic Availability</i>						
Topic availability Per 100 students	0.0012	-0.0001	0.0010	-0.0001	0.0008	0.0001
	(0.0004)***	(0.0005)	(0.0012)	(0.0003)	(0.0003)**	(0.0010)
Individual controls & Year FE	X	X	X	X	X	X
HS controls		X	X		X	X
HS FE			X			X
N	141,579	141,579	141,579	141,579	141,579	141,579

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: Each coefficient is from a separate regression. The course-availability measure in panel A is substantively the same as in our main models, but not adjusted for student enrollment – e.g., it captures the raw number of courses to which students have access in high school. The topic availability measure in panel B is enrollment-adjusted, but does not count additional sections of the same course as new courses, as described in the text. All models control for high school graduation year (year fixed effects). Student controls are race/ethnicity, ACT math and English scores, and high school class rank. High school controls include location (urban, suburban, or rural; this factor drops out with the inclusion of HS fixed effects), enrollment, percent of the student body that identifies as a minority race/ethnicity, and percent of the student body which is free or reduced price lunch eligible. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Appendix Table A4: Results from Separate First Stage Regressions of Course Taking on Course Availability and Topic Availability.

	(1)	(2)
<i>A. Courses Availability</i>		
Unadjusted Course Availability	0.0078	0.0045
	(0.0024)***	(0.0017)***
Kleibergen-Paap LM statistic p-value	0.00	0.01
Kleibergen-Paap Wald F-statistic	10.52	6.86
<i>B. Topic Availability</i>		
Topic Availability Per 100 students	0.0165	0.0097
	(0.0099)*	(0.0124)
Kleibergen-Paap LM statistic p-value	0.08	0.44
Kleibergen-Paap Wald F-statistic	3.17	0.61
Individual & HS controls & Year FE	X	X
HS FE		X

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: Each coefficient is from a separate regression. All models control for high school graduation year (year fixed effects). Student controls are race/ethnicity, ACT math and English scores, and high school class rank. High school controls include location (urban, suburban, or rural; this factor drops out with the inclusion of HS fixed effects), enrollment, percent of the student body that identifies as a minority race/ethnicity, and percent of the student body which is free or reduced price lunch eligible. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10

Appendix Table A5: Results from Combined First Stage Regressions of Course Taking on Unadjusted Course Availability and Topic Availability.

	(1)	(2)
Unadjusted Course Availability	0.0076 (0.0024)***	0.0044 (0.0017)**
Topic Availability Per 100 students	0.0146 (0.0101)	0.0022 (0.0127)
Kleibergen-Paap LM statistic p-value	0.01	0.03
Kleibergen-Paap Wald F-statistic	6.07	3.52
Individual & HS controls & Year FE	X	X
HS FE		X

Source: Administrative data on Missouri public HS students who matriculate into a Missouri public 4-year university. Notes: All models control for high school graduation year (year fixed effects). Student controls are race/ethnicity, ACT math and English scores, and high school class rank. High school controls include location (urban, suburban, or rural; this factor drops out with the inclusion of HS fixed effects), enrollment, percent of the student body that identifies as a minority race/ethnicity, and percent of the student body which is free or reduced price lunch eligible. Standard errors clustered by high school included in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10