Examining the Relationship between STEM Coursetaking in High School and Grade 12 NAEP Mathematics Performance

AIR - NAEP Working Paper 2021-05

Darrick Yee

Burhan Ogut

George Bohrnstedt

Markus Broer

Ruhan Circi

DECEMBER 2021

The research contained in this working paper was commissioned by the National Center for Education Statistics (NCES). It was conducted by the American Institutes for Research (AIR) in the framework of the Education Statistics Services Institute Network (ESSIN) Task Order 14: Assessment Division Support (Contract No. ED-IES-12-D-0002/0004) which supports NCES with expert advice and technical assistance on issues related to the National Assessment of Educational Progress (NAEP). AIR is responsible for any error that this report may contain. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Advancing Evidence. Improving Lives.

This page is intentionally left blank.

Examining the Relationship between STEM Coursetaking in High School and Grade 12 NAEP Mathematics Performance

AIR - NAEP Working Paper 2021-05

DECEMBER 2021

Authors:
Darrick Yee, Burhan Ogut, George Bohrnstedt, Markus Broer, Ruhan Circi

American Institutes for Research®
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
202.403.5000 | **AIR.ORG**

AIR

Established in 1946, with headquarters in Arlington, VA, the American Institutes for Research (AIR) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of health, education, and workforce productivity. For more information, visit www.air.org.

NCES

The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education in the U.S. and other nations. NCES is located within the U.S. Department of Education and the Institute of Education Sciences. NCES fulfills a Congressional mandate to collect, collate, analyze, and report complete statistics on the condition of American education; conduct and publish reports; and review and report on education activities internationally.

ESSIN

The Education Statistics Services Institute Network (ESSIN) is a network of companies that provide the National Center for Education Statistics (NCES) with expert advice and technical assistance, for example in areas such as statistical methodology; research, analysis and reporting; and Survey development. This AIR-NAEP working paper is based on research conducted under the Research, Analysis and Psychometric Support sub-component of ESSIN Task Order 14 for which AIR is the prime contractor.

William Tirre, a Program Director in the NCES Assessment Division, oversees the Research, Analysis and Psychometric Support sub-component of ESSIN Task Order 14.

Suggested citation:

Yee, D., Ogut, B., Bohrnstedt, G., Broer, M., and Circi, R. (2021). *Examining the Relationship Between STEM Coursetaking in High School and Grade 12 NAEP Mathematics Performance.* [AIR-NAEP Working Paper #2021-25]. Washington, DC: American Institutes for Research.

For inquiries, contact:

Darrick Yee, Senior Researcher: dyee@air.org

Markus Broer, Project Director for Research under ESSIN Task 14: mbroer@air.org

Mary Ann Fox, AIR Vice President and Project Director of ESSIN Task 14: mafox@air.org

Executive summary

This study examined high school coursetaking in science, technology, engineering, and mathematics (STEM) courses as a precursor to end-of-high school mathematics proficiency, as measured by the 2013 grade 12 mathematics assessment of the National Assessment of Educational Progress (NAEP). Using a special "overlap sample" of students from the National High School Longitudinal Study of 2009 (HSLS:09) who were also selected to take the NAEP 2013 Grade 12 assessments, we linked school-reported transcript information to NAEP response data to answer the following questions.

How strongly does STEM coursetaking in high school relate to end-of-high school mathematics proficiency?

There were strong relationships between coursetaking indicators and grade 12 NAEP mathematics performance. While this may not be surprising, the magnitude of the relationships was notable. For example, students who took calculus scored nearly 46 NAEP scale score points higher, on average, than students with a similar number of total mathematics course credits whose most advanced mathematics course was Algebra II. This difference was equivalent to roughly 1.4 standard deviations on the grade 12 NAEP mathematics assessment, approximately 53% larger than the overall Black-White score gap.

How does the relationship between STEM coursetaking and end-of-high school mathematics proficiency change when controlling for measures of prior mathematics achievement and student background characteristics?

After accounting for pre-high school characteristics, such as race/ethnicity, gender, socioeconomic status, and prior mathematics achievement, as well as science and engineering courses and STEM course grade point average (GPA), the above relationships were attenuated but still substantial. The estimated average difference between students who took calculus and similar students who stopped at Algebra II declined to 11 points, or about one-third of a standard deviation. Additionally, students who took one additional Advanced Placement (AP) or International Baccalaureate (IB) mathematics course scored about 6 points higher, all else equal. Non-AP/IB credits in mathematics and science were not significantly associated with NAEP scores after accounting for the above factors.

While these results underscore the fact that high school coursetaking in mathematics is strongly related to other factors, including pre-high school factors (such as math proficiency at the

beginning of grade 9), they also suggest that more advanced coursetaking may lead to meaningfully higher mathematics proficiency.

Are there distinct STEM coursetaking patterns in high school for students who score at or above NAEP's college preparedness benchmark in mathematics?

Using results from our statistical models, we identified students who, based on their coursetaking and background characteristics, would be predicted to score at or above the benchmark score of 163 on the grade 12 NAEP mathematics assessment, which has been identified by the National Assessment Governing Board as indicating academic preparedness for college. We found that these students could be partitioned into four groups, each containing more than 1,100 students, that were distinctive with respect to key coursetaking and background characteristics, with two groups of particular interest.

The first group was notable due to its relatively low levels of advanced coursetaking. Only 4% of the students in this group took courses more than one level beyond Algebra II, and almost none of the students earned AP/IB credits in mathematics. The existence of this group suggests that, for many students, advanced coursetaking may not be necessary to obtain high proficiency in mathematics.

The second group had background characteristics typically associated with low mathematics proficiency, including the lowest average grade 9 algebra scores, by far the lowest average socioeconomic status (SES), and the highest percentage of minority students among the four groups. However, this group also took advanced courses at high rates, with more than 95% of its students taking precalculus or above and more than 40% taking calculus. This group provides evidence of a student subpopulation with somewhat disadvantaged background characteristics but advanced coursework and high end-of-high-school mathematics proficiency.

These results are observational in nature, and this study was not designed to identify causal effects of coursetaking. However, taken together, the results are consistent with the idea that, while end-of-high school mathematics achievement is dependent in large part on pre-high school factors, meaningful improvement may be achieved via more advanced coursetaking.

Contents

Executive summary iii

Introduction 1

Background 1

Study Overview 2

Methods 3
- Data 3
- Analysis 9

Results 11
- Mathematics Courses 11
- Science Courses 14
- Correlations 16

Conclusions 29

References 30

Appendix A A-1

List of Tables

Table 1. Summary Statistics for HSLS-NAEP overlap and full HSLS samples, by student demographic characteristics 5

Table 2. Summary statistics for HSLS-NAEP overlap and full HSLS samples, by STEM coursetaking 6

Table 3. Estimated mean grade 12 NAEP mathematics scores and percentage of students, by highest mathematics course taken, 2013 12

Table 4. Mean NAEP grade 12 mathematics scores, by science course completion status, 2013 15

Table 5. Bivariate correlations between HSLS:09 variables and estimated NAEP 2013 grade 12 mathematics scores for the HSLS-NAEP overlap sample 17

Table 6. Results from marginal maximum-likelihood regressions of NAEP 2013 grade 12 mathematics scored on STEM coursetaking variables and student background characteristics 18

Table 7. Group means for clusters obtained by *k*-means cluster analysis of students with predicted NAEP 2013 grade 12 mathematics scores at or above 163 22

Table A-1. Pairwise correlations between coursetaking variables A-1

List of Figures

Figure 1. Distribution of NAEP grade 12 mathematics scores, by highest mathematics course taken: 2013 14

Figure 2. Distribution of NAEP 2013 grade 12 mathematics scores, by science course completion status 16

Figure 3. Scatter plot of HSLS:09 SES index vs. predicted NAEP 2013 grade 12 mathematics score for students with predicted scores of 163 or higher, by cluster 26

Introduction

The relationship between high school coursework and future success has been extensively studied in the literature. Studies show that students who have taken more challenging courses in high school score higher on academic tests and are more likely to enroll in college (Long et al., 2012; Ma & Wilkins, 2007; Adelman, 2006; Noble, Roberts, & Sawyer, 2006; Madigan, 1997). The purpose of this study is to expand the current knowledge on the relationship between high school STEM coursetaking and mathematics proficiency by examining a special "overlap sample" of students from the National High School Longitudinal Study of 2009 (HSLS:09) who were selected to take the 2013 grade 12 mathematics assessment from the National Assessment of Educational Progress (NAEP). This unique dataset provides a means for linking student-level NAEP responses to data collected earlier in students' high school careers, including measures of grade 9 mathematics proficiency and official high school transcripts.

This study aims to extend the prior research in a number of ways. First, the relatively recent, nationally representative nature of our datasets provides for improved generalizability of results, compared to studies investigating narrower populations (ACT test takers, students in a single state, etc.) or students who graduated before recent educational initiatives (such as Algebra for All or Race to the Top) or widespread adoption of the Common Core State Standards. Second, the use of NAEP as our primary outcome helps to mitigate potential distortions caused by high-stakes testing incentives, such as score inflation or test-specific coaching, while retaining the interpretability of a widely recognized assessment. Finally, HSLS:09 obtained course data based on high school transcripts, rather than student self-reports, so that our coursetaking predictors are likely to be more reliable than in prior studies that rely on self-reported transcript information.

Background

Prior research has shown that coursework in high school is a strong predictor of students' end-of-high school proficiency in mathematics and related subjects. Controlling for prior achievement, Noble and Schnelker (2007) examined relationships between course patterns and students' ACT performance in English, mathematics, and science. The authors found that taking trigonometry, calculus, and other advanced mathematics courses, in addition to Algebra I, geometry, and Algebra II, was associated with the highest ACT mathematics scores. Similarly, in a recent technical brief, Allen (2015) concluded that achievement in high school courses in core subjects (English, mathematics, social studies, and natural science) was predictive of subject-specific ACT performance (e.g., the strongest predictor of performance in ACT science was end-of-course chemistry achievement).

Studies have also examined the relationship between the level of courses taken and future performance. Mathematics, perhaps more than any other subject, is regarded as having an ordinal sequence of less-advanced to more-advanced courses, where entry into the next course usually requires completion of courses at previous levels. In an early study, Lee et al. (1997) identified eight mathematics coursetaking patterns and showed that higher course difficulty levels were closely correlated with higher mathematics test scores at the end of high school. Studies have also applied coding schemes to define mathematics course sequences, typically with general math at the lower end and calculus at the other (Schiller & Hunt, 2003; Adelman, 1999; Schneider, Swanson, & Riegle-Crumb, 1998). Using one such categorization, Horn and Nunez (2000) found that students who took advanced mathematics courses in high school were very likely to enroll in a 4-year college, while Ma and Wilkins (2007) found that upper-level mathematics courses (trigonometry, precalculus, and calculus) were associated with increased student growth as measured by mathematics tests, showing that the content of the courses taken may regulate growth in mathematics proficiency.

These observational findings have largely been consistent with results from causal studies. Analyzing data from the Trends in International Mathematics and Science Study (TIMSS) using a propensity score-matching method, Leow et al. (2004) concluded that advanced coursework leads to higher TIMSS performance. In another study, Byun, Irvin, and Bell (2015) investigated the effect of mathematics coursetaking beyond Algebra II on achievement and college enrollment using longitudinal and nationally representative data from the Education Longitudinal Study of 2002 (ELS:2002). These researchers found that advanced coursetaking has a positive effect on both mathematics proficiency and college enrollment. Similarly, Long, Conger, and Iatarola (2012), using longitudinal data from a census of Florida public high school students, estimated that taking at least one “rigorous” mathematics course early in high school can increase 10th-grade mathematics proficiency by as much as one-quarter of a standard deviation and substantially increase the probability of graduating high school and attending a 4-year college. Both of the latter studies also found that course effects appeared to be stronger for disadvantaged students.

Study Overview

The goal of the current study is to investigate high school coursetaking in STEM subjects as a predictor of performance on the NAEP grade 12 mathematics assessment. The focus on mathematics is important because it is often seen as foundational to proficiency and achievement in the other STEM subjects (Cribbs et al., 2015; Wang, 2013).

Specifically, we seek to address the following research questions:

1. How strongly does STEM coursetaking in high school relate to end-of-high school mathematics proficiency?

2. How does the relationship between STEM coursetaking and mathematics proficiency change when controlling for measures of prior mathematics achievement and student background characteristics?

3. Are there distinct STEM coursetaking patterns in high school for students who score at or above NAEP's college preparedness benchmark in mathematics?

Research question 1 (RQ1) is aimed at estimating differences in NAEP performance among students with different levels of high school coursework in STEM subjects, both unconditionally and conditional on other course-related variables. RQ2 is intended to examine the degree to which these differences might be explained by pre-high school background characteristics, such as prior mathematics achievement, race/ethnicity, and socioeconomic status (SES). The motivation underlying RQ3 is to investigate the extent to which combinations of student characteristics that, based on findings from RQ2, would hypothetically lead to high proficiency actually exist in the student population. For this purpose, we define "high proficiency" as meeting or exceeding the NAEP score of 163, which has been identified by the National Assessment Governing Board (NAGB) as the benchmark for academic preparedness for college (Fields, 2014, p. 8).

Methods

Data

Data and Sample

The data for this study were drawn from the 2013 NAEP grade 12 mathematics assessment and HSLS:09.

NAEP data

Student-level item responses and item response theory (IRT) parameters were taken from the 2013 administration of the NAEP grade 12 mathematics test. NAEP is a congressionally mandated assessment program administered by the National Center for Education Statistics (NCES), within the U.S. Department of Education's Institute of Education Sciences, and is the only U.S.-based assessment that measures student achievement in various subjects at the national level. NAEP mathematics assessments are administered at regular intervals to nationally representative samples of students in grades 4, 8, and 12, and the results provide a broad view of students' knowledge, skills, and performance in mathematics (NCES, n.d.). Scores on the grade 12 mathematics assessment are reported on a scale of 0 to 300

HSLS: 09 data

The remaining variables for this study were drawn from HSLS:09. HSLS:09 is a nationally representative, longitudinal study of approximately 24,000 9th-graders in 944 schools, including both public and private school students. The initial, base-year HSLS:09 collection included an algebra assessment and survey component in the fall of 9th grade in 2009. A follow-up data collection was administered in spring 2012, when most students were in 11th grade, and student transcript data were collected from schools in the 2013 update, after most students had completed high school. The analyses for this study use variables from the base-year and student transcript collections, which are described in more detail in the *Variables* section below.

Analysis samples

Two analysis samples were used for this study. The "overlap sample," employed to address research questions 1 and 2, consists of a subset of the HSLS:09 cohort that was selected to take the NAEP grade 12 mathematics assessment in 2013. A crosswalk of HSLS:09 and NAEP student identifiers provided by NCES enables the linking of the two datasets. The "full sample," used to address research question 3, consists of all HSLS:09 students whose transcripts were included in the 2013 update and had complete data on the predictors used in the analyses for research questions 1 and 2.

Not all students in the HSLS:09 cohort were eligible for NAEP. While the full set of eligibility criteria were not available, one condition for eligibility was that the student needed to be in grade 12 in the spring of 2013; the overlap sample was therefore limited to students who were not retained in grade and did not graduate early.

Summary statistics for the overlap and full samples are presented in Tables 1 and 2, by student demographic characteristics and STEM coursetaking, respectively. Statistics for the 2013 NAEP grade 12 mathematics sample are provided for comparison.

Table 1. Summary Statistics for HSLS-NAEP overlap and full HSLS samples, by student demographic characteristics

	HSLS-NAEP Overlap Sample	Full HSLS Sample	NAEP Sample
Observations	2,710	15,330	46,450
		Percentage	
Gender			
Female	50.2	50	49.8
	(1.4)	(0.7)	(0.4)
		Percentage	
Race/Ethnicity			
White	53	52.8	56.4
	(2.2)	(1.1)	(0.7)
Black	11.2	12.9	14.8
	(1.2)	(0.8)	(0.5)
Hispanic	20.8	21.7	20.5
	(1.9)	(0.9)	(0.6)
Asian	6.3	4	5.5
	(1.1)	(0.4)	(0.3)
Other	8.7	8.6	2.6
	(1.6)	(0.5)	(0.2)
		Mean	
Socioeconomic Status	0.03	-0.03	N/A[1]
SES index	(0.03)	(0.02)	

[1] SES index is not available for the full NAEP sample.

NOTE: Standard errors in parentheses. HSLS estimates were obtained using HSLS second follow-up weights and linearized Taylor series standard errors. The HSLS SES index ranges from -1.75 to 2.28, with an overall standard deviation of 0.72. NAEP estimates are based on public school students with complete data on gender and race/ethnicity, with jackknife replication used to calculate standard errors. NAEP race/ethnicity is based on school records and may not agree with HSLS-reported race/ethnicity.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

While the two samples exhibit similar percentages of female students, they appear to differ slightly with respect to race/ethnicity and SES. The percentage of Black students is slightly lower in the overlap sample (11.2%) than in the full sample (12.9%), while the percentage of Asian students is somewhat higher (6.3% vs. 4.0%). Students in the overlap sample also appear to come from families with higher average socioeconomic status, as measured by the HSLS SES index (0.03 vs. -0.03).

The overlap sample also differs from the full sample in terms of STEM coursetaking and achievement, as shown in Table 2. In general, students in the overlap sample were more likely to have enrolled in higher-level mathematics courses, as might be expected due to the exclusion of dropouts and students who were retained in grade. For example, the percentages of students with Algebra I (or below) and Geometry as their highest-level course were 1.6% and 7.3%, respectively, in the overlap sample, while the corresponding percentages in the full sample were 7.7% and 10.4%, indicating that a substantially larger proportion of students in the full sample did not take courses beyond Geometry. Students in the overlap sample were also more likely to have taken AP/IB calculus (14.6 vs. 10.6%) or precalculus (23.4 vs. 19.6%).

Similarly, students in the overlap sample were more likely to have earned at least one credit in biology (94.6 vs. 88.3%), chemistry (73.6 vs. 63.9%), and physics (39.9 vs. 34.6%), earned slightly more credits in mathematics and science (including AP/IB credits), and on average achieved slightly higher overall GPAs in STEM courses (2.58 vs. 2.41) compared to the full sample. In short, the overlap sample appears to be somewhat higher-achieving, where STEM courses are concerned, than the full HSLS sample, likely because of its restriction to students who completed high school within 4 years

Table 2. Summary statistics for HSLS-NAEP overlap and full HSLS samples, by STEM coursetaking

	HSLS-NAEP overlap sample	Full HSLS sample
Observations	2,710	15,330
Highest math course taken	Percentage	
Algebra I or below	1.6	7.7
	(0.5)	(0.5)
Geometry	7.3	10.4
	(0.9)	(0.5)
Algebra II	24.2	22.9
	(1.8)	(0.8)

	HSLS-NAEP overlap sample	Full HSLS sample
Trigonometry, probability/statistics, or other advanced math course	22.5	22.3
	(1.6)	(0.9)
Other AP/IB	2.4	1.9
	(0.6)	(0.3)
Precalculus	23.4	19.6
	(1.2)	(0.6)
Calculus	4	4.6
	(0.6)	(0.3)
AP/IB calculus	14.6	10.6
	(1.4)	(0.6)
Science courses completed (minimum 1 credit earned)	Percentage	
Biology	94.6	88.3
	(1.0)	(0.8)
Chemistry	73.6	63.9
	(1.9)	(1.2)
Geology/Earth Science	67.4	65.7
	(2.6)	(1.5)
Physics	39.9	34.6
	(2.1)	(1.0)
Credits earned in STEM courses	Mean	
Math	3.89	3.65
	(0.04)	(0.02)
AP/IB math	0.26	0.18
	(0.02)	(0.01)
Science	3.6	3.33
	(0.05)	(0.03)
AP/IB science	0.27	0.19
	(0.02)	(0.01)

	HSLS-NAEP overlap sample	Full HSLS sample
Engineering	0.2	0.19
	(0.02)	(0.01)
All STEM	8.2	7.67
	(0.09)	(0.05)
GPA in STEM courses	Mean	
GPA	2.58	2.41
	(0.03)	(0.02)

NOTE: Standard errors in parentheses. Estimates obtained using HSLS second follow-up weights and linearized Taylor series standard errors.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Variables

Grade 12 NAEP mathematics proficiency. Grade 12 mathematics proficiency, the dependent variable for RQ1 and RQ2, was measured using results from the 2013 grade 12 NAEP mathematics assessment. Student item responses and IRT parameters from the NAEP assessment were used to impute scale scores for these analyses. Nationally, NAEP grade 12 mathematics scores have a range of 0 to 300, with a mean of 153 and a standard deviation of 33.

Highest mathematics course taken (HIMATH). The HSLS dataset provides an ordinal variable for the "highest" mathematics course appearing on each student's final high school transcript, regardless of whether or not the student was awarded credit for the course. The original variable contained 14 values, which we collapsed into 8, ranging from "Algebra I or below" to "AP/IB Calculus," in order to mitigate estimation difficulties resulting from small frequencies.

Science courses (SCIENCE). Indicators were included for course completion in each of four science subjects: biology, geology/Earth sciences, chemistry, and physics. For each subject, students were assigned a value of 1 if they earned at least one credit in any course in the subject; otherwise, they were assigned a value of 0.

Credits earned (CREDITS). A set of predictors was included measuring the number of credits earned in each of the following six course categories: mathematics (any), AP/IB mathematics, science (any), AP/IB science, engineering, and all STEM courses. For mathematics and science, the overall credits earned included AP/IB credits.

GPA in STEM courses (GPA). Grade point averages (GPAs) in individual subjects were strongly correlated in the sample. In order to mitigate potential problems related to highly collinear predictors, overall STEM GPA was employed as the primary measure of student course performance, rather than individual course or overall GPA.

Several student background characteristics were also included in the analyses: gender (coded 0 for male and 1 for female); race/ethnicity (Black, Hispanic, Asian, or Other, all coded as dummy variables, with White as the omitted category); and the SES index provided in the HSLS:09 dataset, which incorporates information on the income, educational attainment, and occupational status of each student's parents.

Analysis

MML Regression Analysis

To examine the relationship between STEM coursetaking and performance on the NAEP mathematics assessment, NAEP scores were modeled for student i using the linear equation,

$$Y_i = \beta_0 + \beta'_1 HIMATH_i + \beta'_2 SCIENCE_i + \beta'_3 CREDITS_i + \beta'_4 GPA_i + \beta'_5 X_i + \varepsilon_i \qquad (1)$$

where Y_i denotes the student's scale score on the 2013 grade 12 mathematics assessment, X_i is the vector of student background characteristics summarized in Table 1 (gender, race/ethnicity, and socioeconomic status), and ε_i is an error term. The parameters β_1 through β_4 capture the primary relationships of interest between course-related variables and NAEP achievement.

Because the predictors for this model were taken from the HSLS:09 dataset and were not included in NAEP's population structure model, scale scores from the NAEP dataset may lead to biased estimates if used as outcomes for this study (Jacob & Rothstein, 2016; Briggs, 2008; Mislevy, 1991). For this reason, the model's parameters were estimated directly using marginal maximum likelihood (MML), which makes use of student item responses, IRT parameter estimates for each item, and the predictor variables from equation (1) (Cohen & Jiang, 1999). The analyses were conducted using AM software (American Institutes for Research, 2011). Because the primary predictors of interest were collected during the second HSLS:09 follow-up in 2013, student weights from that collection were used in all of the analyses.

The HSLS:09 dataset provides two methods for estimating the standard errors for statistics computed using its variables: an analytical method using Taylor series linear approximation and a resampling method using balanced repeated replication (BRR).[1] BRR estimation was not feasible for the overlap sample, as the reduced sample size resulted in replications in which

[1] See https://nces.ed.gov/pubs2014/2014361.pdf (p. 122).

weights were unchanged or an insufficient number of observations remained in the sample. For this reason, we report linearized Taylor series standard errors for all analyses.

Cluster Analysis

To address RQ3, “Are there distinct STEM coursetaking patterns in high school for students who score at or above NAEP’s college readiness preparedness benchmark in mathematics?” we applied a series of cluster analyses to students in the full HSLS:09 sample. The goal of these analyses was to identify groups of college-prepared students who were distinct in terms of their coursetaking and background characteristics in order to illustrate possible “paths” to college-preparedness in mathematics and the coursetaking combinations associated with them.

In order to restrict the analysis sample to students with high mathematics achievement, it was first necessary to estimate mean NAEP scores for each student, including those who did not participate in NAEP. To obtain these means, we imputed NAEP scores for each student in the full sample who had complete data on the predictors employed in our preferred model from RQ2, using the regression-based relationships between the variables of interest and NAEP performance.

The sample was restricted to students whose predicted mean NAEP score was at least 163, the cutoff score for college preparedness proposed by NAGB.[2] Analyses were then carried out on this “college-prepared” sample using k-means cluster analysis (MacQueen, 1967; Everitt et al. 2011). This approach attempts to partition observations into k groups, where k is a number specified by the researcher, so that each observation’s values on a specified variable vector are closer to its group’s mean than to any other group’s mean. In order to focus on characteristics that were of substantive interest, the following variables were included in these analyses:

- Continuous/ordinal: *GPA*, *CREDITS*, *HIMATH*, and *SES*
- Dichotomous: *SCIENCE*, *FEMALE*, and race/ethnicity indicators

Cluster analyses were conducted using the cluster *kmeans* command in Stata, which follows an iterative procedure in which initial clusters are formed, observations are reassigned to the cluster whose mean is the shortest “distance” away, cluster means are recomputed, and the process is repeated until no observations change clusters between two consecutive iterations (StataCorp, 2017).

[2] NAGB defines college preparedness as “…qualify[ing] for placement into entry-level college credit courses that meet general education requirements, without the need for remedial coursework in mathematics or reading.” For details on methodology, see https://www.nagb.gov/focus-areas/reports/preparedness-research.html.

Dichotomous and continuous variables were both included in the analyses, and the latter also exhibited scale differences that were not substantively meaningful (e.g., a one-point difference in GPA may be more consequential than a one-point difference in total mathematics credits). For this reason, conventional measures of dissimilarity, such as Euclidean distance, were not likely to produce useful results. We instead employed Gower's (1971) dissimilarity coefficient, which is often considered appropriate for combinations of dichotomous and continuous variables. For student i and cluster k, Gower's coefficient can be expressed as:

$$\frac{\sum_{v=1}^{V} d_{ikv}}{V},$$

where V is the total number of variables, and for each variable x_{iv},

$$d_{ikv} = \frac{|x_{iv} - \bar{x}_{kv}|}{\max(x_{\cdot v}) - \min(x_{\cdot v})}$$

where $\bar{x}_{kv}$ denotes the mean of x_{iv} for students currently in cluster k, and the range in the denominator is taken across all students in the sample (StataCorp, 2013; Everitt et al., 2011).

Results

Findings for each research question are presented below.

RQ1: How strongly does STEM coursetaking in high school relate to end-of-high school mathematics proficiency?

We address RQ1 by estimating the relationships between NAEP performance and coursetaking separately for mathematics and science courses. We also present zero-order correlations between selected variables and estimated NAEP performance.

Mathematics Courses

For mathematics courses, NAEP score means and standard errors were computed via MML for each level of *HIMATH*. These results are presented in Table 3.

Table 3. Estimated mean grade 12 NAEP mathematics scores and percentage of students, by highest mathematics course taken, 2013

Highest mathematics course	Percentage of students	NAEP mean score
Algebra I or below	1.6	103
	(0.5)	(13.0)
Geometry	7.3	132
	(0.9)	(4.5)
Algebra II	24.2	138
	(1.8)	(2.9)
Trigonometry, probability/statistics, or other advanced math course	22.5	148
	(1.6)	(2.2)
Other AP/IB	2.4	169
	(0.6)	(5.6)
Precalculus	23.4	170
	(1.2)	(1.9)
Calculus	4.0	183.0
	(0.6)	(4.1)
AP/IB calculus	14.6	199
	(1.4)	(2.8)

NOTE: *N*≈2,710. Standard errors in parentheses. Estimates obtained using direct estimation in AM software.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Table 3 shows that NAEP performance was strongly associated with *HIMATH*. The estimated mean NAEP score for students in the lowest category (Algebra I or below) was 103 scale score points, while the estimate for the second-lowest category (Geometry) was 132. Nationally, NAEP scores have a standard deviation of about 33 points; thus, the 29-point difference between these point estimates should be considered large.

Even larger differences appeared between other categories. Students who enrolled in AP/IB calculus had an estimated mean of 199 points, while those who did not enroll in courses higher than Algebra II had an estimated mean of only 138. This difference of 61 points is nearly twice as large as the White-Black achievement gap in typical NAEP mathematics assessments.[3] The overall range of means, between the lowest category of Algebra I or below and the highest category of AP/IB calculus, was also quite large, at 96 points, or roughly three standard deviations.

To examine the dispersion of scores within each category, we also produced 20 NAEP score plausible values (PVs) for each student in the overlap sample using an MML regression model containing all predictors, following Ogut et al. (2015), and these PVs were employed to estimate the 5th, 25th, 50th, 75th, and 95th percentiles of the score distributions within each *HIMATH* category. The results are displayed in Figure 1, along with lines indicating the cutoff scores for NAEP achievement levels in grade 12 mathematics (*Basic*, *Proficient*, and *Advanced*).

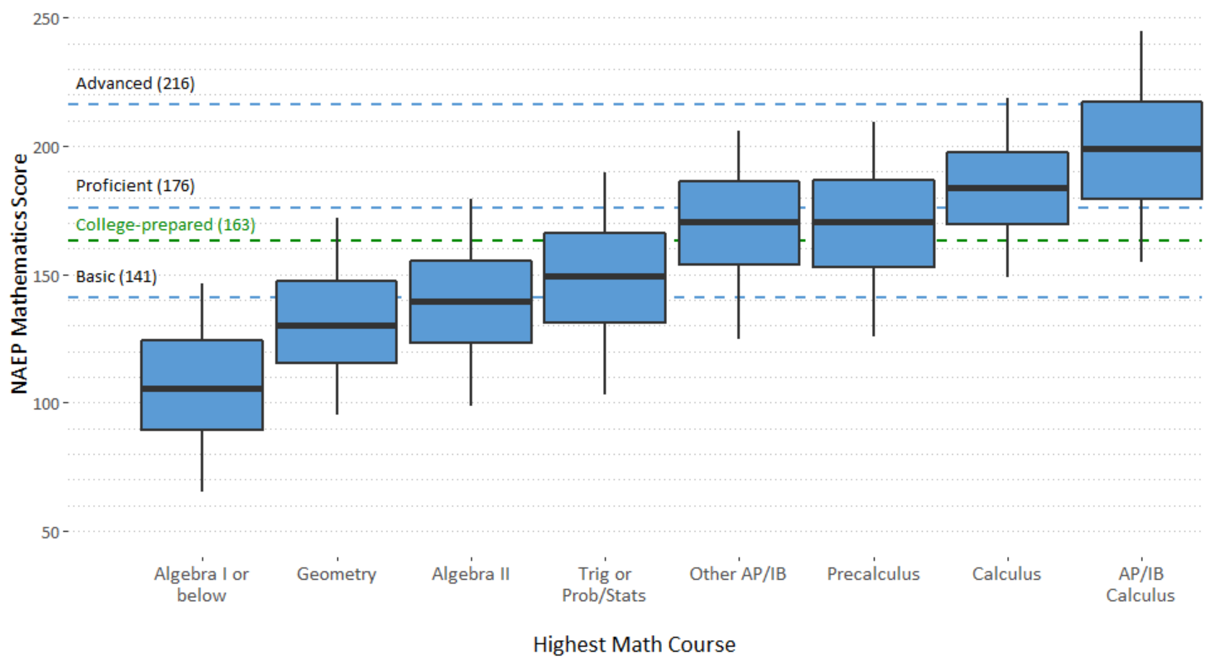
Figure 1 shows that calculus and AP/IB calculus were the only categories in which a majority of students performed at the *NAEP Proficient* level (176) or higher: the estimated median scores for students who took calculus or AP/IB calculus both exceeded 176, while the medians for students whose highest course taken was below calculus were all below 176. In addition, among the 1.6% of students who did not take courses beyond Algebra I, a large majority scored below the *NAEP Basic* level of 141, as did a majority of students whose highest course was Geometry or Algebra II.

Another important NAEP cutoff score on the grade 12 mathematics scales is 163, the score identified by NAGB as indicating preparedness for college. As Figure 1 indicates, most students who took mathematics courses at the precalculus level or above exceeded this benchmark, while a majority of students in the other categories did not. This finding calls into question the adage that Geometry and Algebra II are the gatekeeper mathematics courses for preparedness for college work (Adelman, 2006).

---

[3] For example, the Black-White gaps for the 2009 and 2013 NAEP grade 12 mathematics assessment were 30 points, and the 2011 and 2013 NAEP grade 8 mathematics Black-White gaps were both 31 points.

Figure 1. Distribution of NAEP grade 12 mathematics scores, by highest mathematics course taken: 2013



NOTE: Whiskers extend to 5th and 95th percentiles. Medians indicated by horizontal black lines within boxes. NAEP scores estimated using 20 plausible values (PVs) per student, with PV parameters based on marginal maximum-likelihood regression.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Science Courses

We next examined the relationship between science coursetaking and NAEP performance by estimating mean NAEP scores for students who did and did not earn at least one credit in each of the four subject areas—biology, chemistry, geology/Earth science, and physics—included in the dataset. The results are shown in Table 4.

Table 4. Mean NAEP grade 12 mathematics scores, by science course completion status, 2013

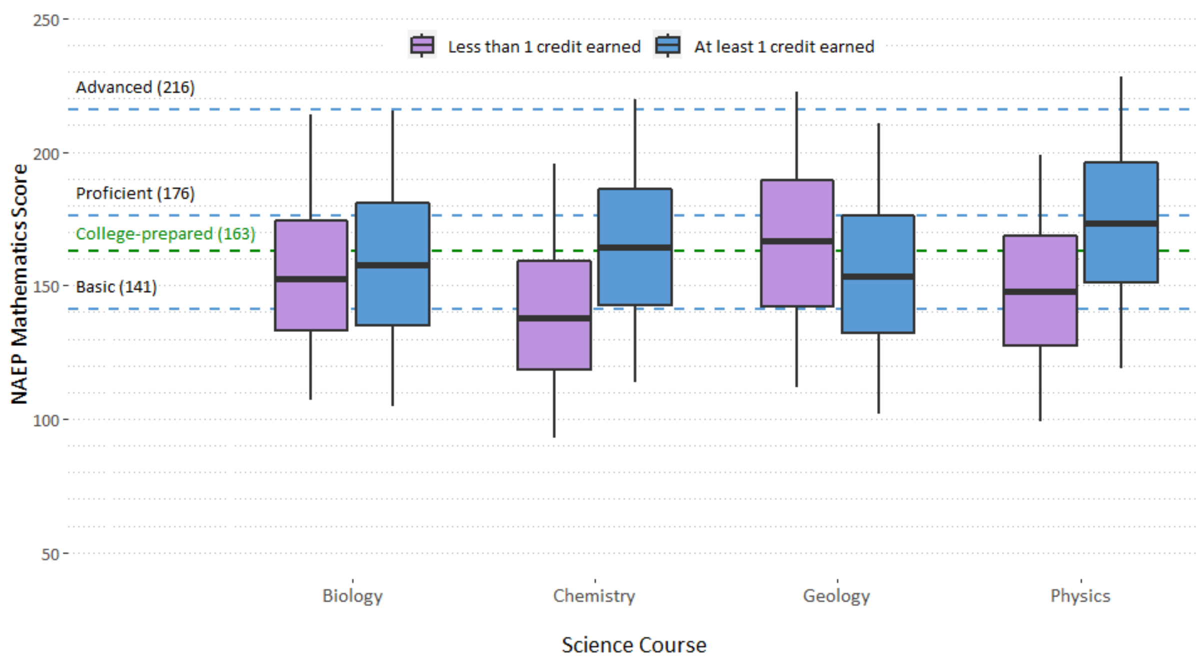
Column 1	At least one credit earned		
Subject	Yes	No	Percent
Biology	158.0	155.0	94.6
	(1.7)	(4.9)	
Chemistry	165.0	139.0	73.6
	(1.7)	(3.4)	
Geology/Earth Science	154.0	166.0	67.4
	(1.8)	(3.0)	
Physics	174.0	147.0	39.9
	(2.2)	(1.9)	

NOTE: *N*≈2,710. Standard errors in parentheses. Estimates obtained using direct estimation in AM software.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Students who earned at least one credit ("completers") in chemistry or physics scored higher, on average, than those who did not. The differences between completers and noncompleters in chemistry and physics (26 and 27 points, respectively) were large, although not as large as for many of the *HIMATH* course categories. Students who completed biology did not appear to differ substantially from those who did not, in terms of average NAEP mathematics performance, while geology/Earth science completers appeared to have somewhat lower average scores than noncompleters (-12 points). This finding, which may seem counterintuitive, might be explained by students with strong mathematics skills foregoing these courses in favor of biology, chemistry, or physics.

Score distributions for science course completers and noncompleters are displayed in Figure 2. About half of the students who earned at least one credit in physics performed at the *NAEP Proficient* level or higher in mathematics; for all other categories, this percentage was noticeably less than 50%. The lowest-performing category consisted of students who did not earn at least one credit in chemistry, who constituted about one-third of the analysis sample (see Table 4). Most of these students scored below the *NAEP Basic* level of 141 points. However, slightly more than half of the students who earned credits in chemistry or physics scored above the college-prepared level, while only a minority of students in the other categories did.

Figure 2. Distribution of NAEP 2013 grade 12 mathematics scores, by science course completion status



NOTE: *N*≈2,710. Whiskers extend to 5th and 95th percentiles. NAEP scores estimated using 20 plausible values (PVs) per student, with PV parameters based on marginal maximum-likelihood regression.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Correlations

Pairwise correlations between estimated NAEP scores and each of the continuous variables in the analysis are presented in Table 5. When treated as continuous, the highest mathematics course indicator was strongly correlated with NAEP grade 12 mathematics performance ($\hat{\rho}$ = .64) and was substantially more strongly correlated than any variable except overall STEM GPA performance ($\hat{\rho}$ = .64). Notably, the correlation between NAEP grade 12 mathematics and total mathematics credits ($\hat{\rho}$ = .15) was among the weakest. The HSLS:09 grade 9 algebra score was also strongly correlated with NAEP grade 12 mathematics performance ($\hat{\rho}$ = .54), indicating that prior achievement is strongly associated with performance at grade 12. Finally, the correlations between NAEP and AP/IB credits in mathematics and science were 0.51 and 0.45 respectively, nearly as high as the correlation between prior achievement and NAEP performance.

Table 5. Bivariate correlations between HSLS:09 variables and estimated NAEP 2013 grade 12 mathematics scores for the HSLS-NAEP overlap sample

Variable	Correlation
STEM GPA	0.64
Math credits	0.15
AP/IB math credits	0.51
Science credits	0.36
AP/IB Science credits	0.45
Engineering credits	0.06
STEM credits	0.26
Highest math course	0.64
SES index	0.33
Grade 9 algebra score	0.54

NOTE: *N*≈2,710. Correlations are means estimated from 20 NAEP plausible values.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

RQ2: How does the relationship between STEM coursetaking and end-of-high school mathematics proficiency change when controlling for measures of prior mathematics achievement and student background characteristics?

We fit a series of multivariate MML regression models to address this question, with the results presented in Table 6. We present our results starting with mathematics coursetaking variables in Model 1. Science courses are added in Model 2, STEM GPA in Model 3, pre-high school background characteristics in Model 4, and grade 9 algebra scores in Model 5.

Model 1 mirrors the results for RQ1, with strong relationships between mathematics coursetaking and NAEP performance. Conditional on mathematics credits earned, students who took no courses beyond Algebra I performed more than 37 points lower, on average, than students whose highest mathematics course was Algebra II (the reference category). At the other end of the *HIMATH* range, students who took some version of calculus scored more than 40 points higher than Algebra II students. Each of these differences was greater than the national standard deviation of 33 points. Furthermore, conditional on *HIMATH*, each additional credit in AP/IB mathematics courses was associated with a nearly 17-point higher NAEP mathematics score, representing about half of a standard deviation. Notably, the pseudo-R^2 for Model 1 was 0.58, indicating that the highest mathematics course and mathematics credit variables explained more than half of the total variation in NAEP performance among the HSLS:09 overlap sample.

In Model 2, the addition of science coursetaking indicators, as well as credits in engineering and credits summed across all STEM courses, attenuates the mathematics-related coefficients and shows that science courses and credits were related to mathematics performance even after accounting for mathematics coursetaking. Students who earned at least one credit in physics or chemistry scored significantly higher than those who did not, all else equal, and each additional credit in AP/IB science courses was associated with a 7.5-point increase in performance.

The addition of GPA, background characteristics, and grade 9 algebra scores in models 3 through 5 continues to attenuate the parameter estimates for mathematics and science coursetaking, a result consistent with the idea that pre-high school characteristics cause variations in both coursetaking and end-of-high school mathematics performance. Nevertheless, Model 5, our preferred model, shows that AP/IB credits in both mathematics and science were significantly related to mathematics performance even after accounting for race/ethnicity, SES, gender, grade 9 algebra scores, and the other coursetaking variables, while regular credits were not. Similarly, students who took precalculus or calculus performed significantly better than those who stopped at Algebra II, with calculus takers performing about one-third of a standard deviation better, on average. Additionally, students who earned credits in physics and chemistry performed better than those who did not, all else equal. Importantly, the pseudo-R^2 for Model 5 was nearly 0.86, indicating that a large majority of the overall variation in NAEP mathematics performance was explained by the model.

In short, while the relationships between STEM coursework and NAEP grade 12 mathematics performance were much weaker after controlling for background and pre-high school characteristics, there were still statistically significant and substantively meaningful relationships between NAEP performance and *HIMATH*, AP/IB credits in both mathematics and science, physics and chemistry credits, and overall STEM GPA.

Table 6. Results from marginal maximum-likelihood regressions of NAEP 2013 grade 12 mathematics scored on STEM coursetaking variables and student background characteristics

Outcome: NAEP 2013 grade 12 mathematics score	Model 1	Model 2	Model 3	Model 4	Model 5
Highest math course taken (vs. Algebra II)					
Algebra I or below	37.6*	-31.3*	-22.1*	-21.6*	-16.3
	-12.4	-12.1	-11.2	-10.7	-11.8

Outcome: NAEP 2013 grade 12 mathematics score	Model 1	Model 2	Model 3	Model 4	Model 5
Geometry	-9.2	-3.2	1.0	-0.9	0.4
	-5	-5.1	-4.7	-4.5	-4
Trigonometry, probability/statistics, or other advanced math course	10.1*	8.5*	5.9	6.2*	3.1
	-3.4	-3.3	-3.0	-2.8	-2.5
Other AP/IB	15.7*	16.3*	8.2	8.7	4.0
	-6.4	-5.9	-5.3	-5.2	-4.6
Precalculus	29.7*	24.4*	14.7*	14.5*	7.5*
	-3.3	-3.3	-3.0	-2.9	-2.6
Calculus	45.7*	40.5*	23.8*	21.6*	11.1*
	-5.3	-5.1	-4.8	-4.6	-4.0
AP/IB calculus	40.1*	32.1*	18.0*	18.5*	11.6*
	-5.2	-5.1	-5.0	-4.7	-4.1
Science courses completed (at least one credit earned)					
Biology		-9.0*	-4.1	-3.9	-3.3
		-4.1	-4.0	-3.7	-3.0
Geology/Earth Science		0.1	1.5	1.5	0.8
		-2.6	-2.2	-2.1	-1.7
Chemistry		9.5*	6.4*	6.2*	5.1*
		-2.9	-2.6	-2.4	-2.1
Physics		6.3*	8.7*	7.7*	4.7*
		-2.3	-2.1	-1.9	-1.6
Credits earned					
Math	-3.7*	-3.7*	-3.0	-2.0	-1.0
	-1.2	-1.8	-1.6	-1.6	-1.4
AP/IB Math	16.9*	13.4*	11.6*	9.8*	5.9*
	-2.8	-2.7	-2.4	-2.3	-2.0
Science		0.7	-0.3	0.1	0.6
		-1.8	-1.6	-1.6	-1.3
AP/IB Science		7.5*	6.2*	5.9*	3.0*

Outcome: NAEP 2013 grade 12 mathematics score	Model 1	Model 2	Model 3	Model 4	Model 5
		-1.6	-1.4	-1.4	-1.2
Engineering		2.0	2.1	0.7	0.0
		-1.8	-1.7	-1.6	-1.4
All STEM		-0.3	-1.1	-1.4	-0.9
		-1.3	-1.2	-1.2	-1.0
GPA in STEM courses			16.7*	15.7*	9.4*
			-1.5	-1.5	-1.2
Female				-8.8*	-6.7*
				-1.7	-1.5
Race/ethnicity					
Black				-12.6*	-9.8*
				-3.3	-2.8
Hispanic				-0.2	-1.4
				-2.8	-2.5
Asian				1.4	-1.1
				-3.8	-3.4
Other				-1.9	-2.6
				-3.2	-2.8
SES index				3.4*	1.4
				-1.3	-1.1
Grade 9 algebra score					16.5*
					-1.2
Constant	152.3*	152.5*	118.0*	125.5*	135.7*
	-5	-6.4	-6.6	-6.3	-5.1
RMSE	22.0	20.8	18.0	16.9	13.2
Pseudo-R^2	0.579	0.636	0.728	0.768	0.859

* $p < .05$.

NOTE: $N \approx 2{,}710$. Standard errors in parentheses. Sample drawn from “overlap sample” of students from the High School Longitudinal Study of 2009 who were administered the NAEP 2013 grade 12 mathematics assessment.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

The results above suggest that many “paths” to high proficiency and college readiness might exist, in the sense that many combinations of predictors could result in predicted scores that meet or exceed the NAEP “college-prepared” score of 163. However, the practical implications of these results may be limited if such predictor combinations are not found among the student population. For example, the models suggest that students with very low SES and grade 9 algebra scores might still achieve high mathematics proficiency by the end of high school if they also complete many AP/IB mathematics and science courses, but such students may be extremely rare in practice, if they exist at all. This concern motivates our third research question.

RQ3: Are there distinct STEM coursetaking patterns in high school for students who score at or above NAEP’s college preparedness benchmark in mathematics?

To address this question, we sought to partition college-prepared students into a reasonably small number of clusters, such that each cluster (a) contains a nontrivial proportion of students; and (b) distinguishes itself from the others in ways that are easily interpretable. As described in the *Analysis* section above, we employed k-means clustering for these analyses to limit the number of clusters to an acceptable range. After examining solutions ranging from three to eight clusters, a k-means solution with four clusters emerged as one that satisfied our requirements well.

Summary statistics for each of these clusters, as well as for the overall “college-prepared” sample, are presented in Table 7. Clusters are presented in descending order by their mean estimated NAEP scores, from 204 for cluster A to 173 for cluster D. To aid in interpretation, the SES index and grade 9 algebra scores from HSLS:09 (which were not among the clustering variables) are standardized across the full sample, so that values above 0 indicate that students have higher SES or prior achievement than the mean for the HSLS:09 full sample.

Table 7. Group means for clusters obtained by *k*-means cluster analysis of students with predicted NAEP 2013 grade 12 mathematics scores at or above 163

	Cluster				
Variables	A	B	C	D	Overall
GPA and credits					
STEM GPA	3.4	3.2	3.0	3.0	3.2
Math credits	4.4	4.0	4.1	3.5	4.0
AP/IB math credits	1.4	0.2	0.4	#	0.5
Science credits	4.4	4.0	3.9	3.5	4.0
AP/IB science credits	1.2	0.3	0.3	0.2	0.5
Highest math course taken					
Algebra I or below	#	#	#	0.1	#
Geometry	#	#	#	4.9	1.1
Algebra II	#	#	#	26.4	5.9
Trigonometry, probability/statistics, or other advanced math course	#	#	#	65	14.4
Other AP/IB	0.6	4.9	4.8	1.7	3.1
Precalculus	2.2	70.2	53.6	2.0	34.4
Calculus	2.0	24.0	17.4	#	11.7
AP/IB calculus	95.2	1.0	24.2	#	29.4
Science courses (at least 1 credit)					
Biology	94.7	94.1	95.4	90.3	93.7
Geology/Earth Science	47.8	59.0	56.0	62.9	56.4
Chemistry	92.3	90.3	91.1	82.1	89.2
Physics	75.5	61.8	62.5	34.5	59.3
Background characteristics					
Female	46.1	50.9	47.1	40.0	46.5
White	66.5	76.7	55.7	69.8	68.1
Black	3.1	2.6	5.4	3.6	3.6
Hispanic	9.7	9.4	25.1	14.7	14

	Cluster				
Asian	14.2	5.3	7.4	3.3	7.5
Other	6.4	6.1	6.5	8.5	6.8
SES index (standardized)	0.92	1.13	-0.52	0.33	0.55
Grade 9 algebra score (standardized)	1.32	0.86	0.82	0.89	0.97
NAEP score (predicted)	204	182	179	173	185
N	1,840	2,160	1,120	1,230	6,340

Rounds to zero.

NOTE: Cluster analysis sample includes only students with predicted scores of at least 163. Variables labeled as "standardized" are standardized across the full sample of approximately 15,330 students. Grade 9 algebra scores and NAEP scores were not included in the cluster analysis.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Our analysis reveals meaningful heterogeneity in coursework and demographic characteristics within this sample and, more importantly, may point to "paths" to college-readiness that vary across students with different backgrounds. The smallest cluster contained about 1,120 observations, comprising roughly 18% of the college-prepared sample; each cluster thus represents a nontrivial proportion of the college-prepared population.

Cluster A, which was characterized by the highest average NAEP scores, presented a pattern that was primarily distinguished by its extremely high prior achievement scores—1.32, compared to 0.97 for the overall sample—and advanced coursework, with nearly all students (95%) taking AP/IB calculus as their highest mathematics course and students more likely to have completed physics than the overall college-prepared sample (76% vs. 59%). The average student in this cluster also earned more than one credit in both AP/IB mathematics (1.38) and AP/IB science (1.17), while no other cluster exceeded an average of 0.38 in either subject. Cluster A students, on average, took more science and mathematics credits than those in any other cluster and also had the highest overall GPA in STEM courses. The average SES of this group was also 0.92 standard deviation (SD) higher than the overall HSLS:09 population, and the group had the highest percentage of Asian students (14%) of all four clusters, as well as the second-lowest percentages of Black and Hispanic students (3.1% and 10%, respectively).

Cluster B's average NAEP score of 182, while the second highest of the four groups, was much closer to those of Clusters C (179) and D (173) than to that of Cluster A, which it lagged by more than 22 points, despite being more demographically similar to Cluster A than to the other two clusters. This may be attributable, in part, to its relatively low average grade 9 algebra score (0.86 SD above the full-sample mean, compared to 0.97 for all four clusters combined) and differences in coursetaking: students in cluster B, on average, earned less than one-third of a credit in AP/IB science and even fewer in AP/IB mathematics (0.16), and only 25% took calculus as their highest math course, with about 1% taking AP/IB calculus. The highest math course taken by a student in this cluster was typically either precalculus (70%) or calculus (24%). Cluster B had the highest percentage of White students (77), the lowest percentage of Black students (2.6), and the highest average SES (1.13) of the four clusters and, therefore, might be thought of as consisting of students who, on average, were the “least disadvantaged” by traditional background measures.

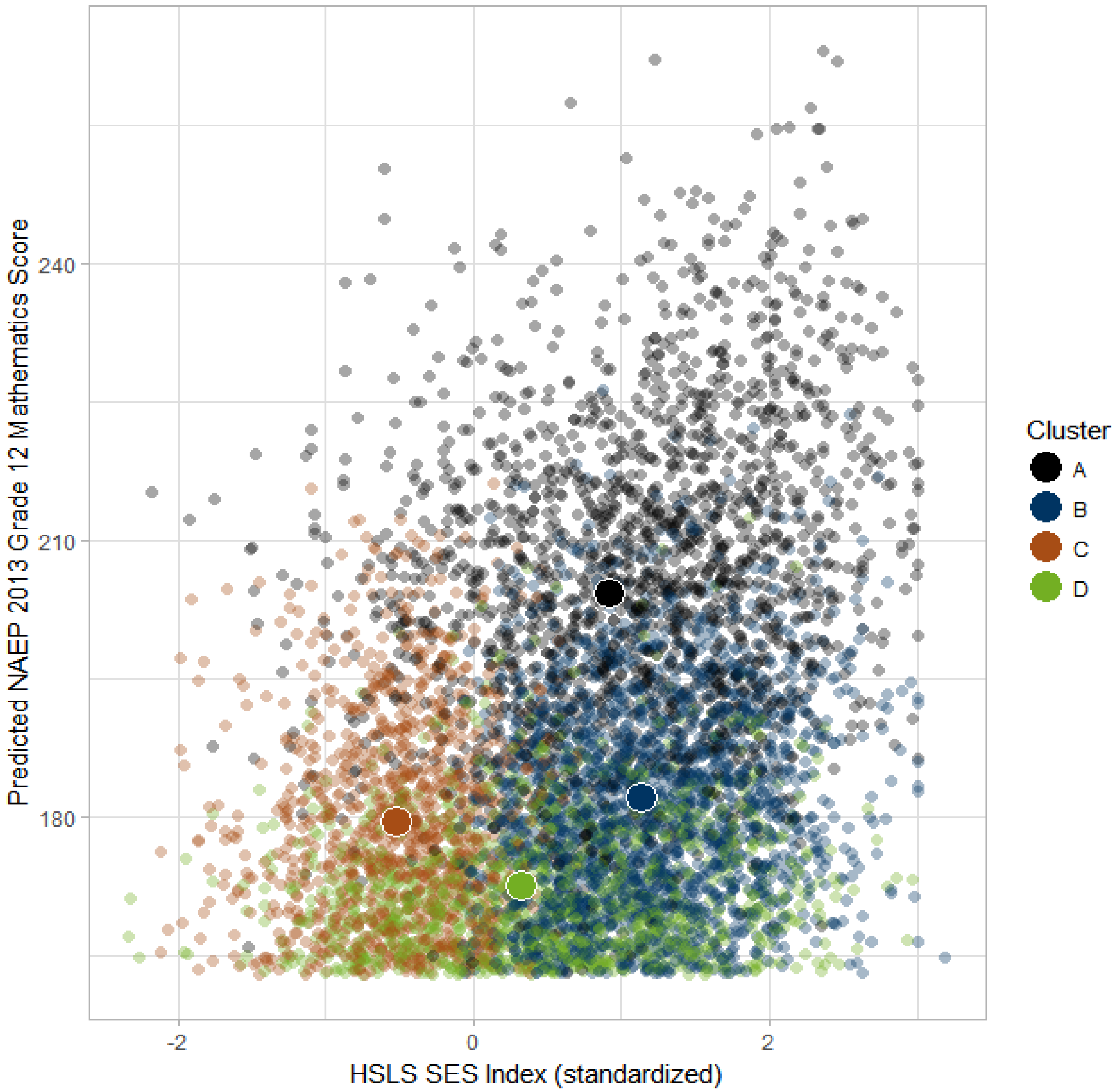
Cluster C is of particular interest for this study and provides confirmatory evidence of a possible effect of coursetaking on mathematics proficiency. Specifically, students in Cluster C were more likely to take either calculus or AP/IB calculus (42%) as their highest math course than were those in Clusters B (25%) and D (rounds to zero) and, on average, earned more than twice as many AP/IB mathematics credits (0.38 vs. 0.16 and 0.02) and were more likely to complete chemistry or physics, especially compared to Cluster D. However, students in this cluster also had the lowest average prior achievement scores (0.82) and by far the lowest SES (-0.52), more than a full standard deviation lower than the overall college-prepared average of 0.55. Compared to the others, this cluster also had the lowest percentage of White students (56%) and the highest percentage of Black (5%) and Hispanic (25%) students. Based on these background characteristics alone, this cluster might be expected to perform worse on NAEP than the other clusters. However, the average NAEP score for this cluster was 6 points higher than for cluster D—a group with higher prior achievement and a much higher average SES—and within 3 points of cluster B. These results suggest that students in Cluster C may have been able to mitigate the negative effects of what might be considered disadvantageous background characteristics to obtain high proficiency in mathematics by taking a rigorous set of courses. At a minimum, they provide evidence that many low-SES and minority students are able to complete advanced STEM coursework and achieve high mathematics proficiency in high school.

Cluster D students, compared with the other three clusters, had the lowest average predicted NAEP score (173), the lowest percentage completing chemistry (82%), by far the lowest percentage to take physics (35%), and the lowest percentage to take any form of calculus (rounds to zero) as their highest mathematics course, despite having the second-highest average prior mathematics performance (0.89). Instead, nearly all of these students took either Algebra II (26%) or one course beyond (65%) as their highest math course. While Cluster D had the lowest average predicted NAEP score of any of the four clusters, its mean performance was still 10 full points, or roughly one-third of a standard deviation, above the college-prepared benchmark of 163. The students in this cluster suggest that, other things being equal, neither calculus nor physics is required for a student to attain college preparedness in mathematics.

To provide a more detailed look at differences between clusters with respect to their performance on NAEP mathematics and SES, we present a scatter plot of NAEP scores vs. SES values for all students with scores of 163 or greater in Figure 3. Clusters are distinguished by color, and students are represented as semitransparent circles. Cluster means are presented as larger, opaque circles.

The overall upward slope of points in Figure 3 reflects the positive marginal relationship between SES and NAEP performance. Closer examination of the clusters, however, reveals the potential role that coursetaking may play in moderating this relationship. While Cluster A has a noticeably higher mean SES than Cluster D, there is substantial overlap between the two groups in terms of SES; however, the two clusters exhibit little overlap in NAEP scores, with students in the more advanced coursetaking Cluster A generally scoring much higher at all SES levels. Meanwhile, Clusters B and C are quite distinct in terms of SES, with B having the highest mean (1.13) of the four groups and C the lowest (-0.52), with little overlap between the two. Given the marginal relationship between SES and NAEP, one might therefore expect Cluster B to score well below Cluster C; however, the two clusters are nearly indistinguishable in terms of NAEP scores. In both the A vs. D and B vs. C comparisons, the cluster with stronger coursetaking appears to “overperform” relative to the marginal relationship.

Figure 3. Scatter plot of HSLS:09 SES index vs. predicted NAEP 2013 grade 12 mathematics score for students with predicted scores of 163 or higher, by cluster



NOTE: *N*≈6,340. Predicted NAEP scores generated using parameter estimates from marginal maximum-likelihood regression on NAEP-HSLS overlap sample. Clusters obtained via *k*-means cluster analysis using variables related to high school coursetaking and performance in STEM subjects and student background characteristics.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) and National Assessment of Educational Progress (NAEP) 2013 Grade 12 Mathematics Assessment.

Discussion

The results for RQ1 and RQ2 in this study are broadly in line with the findings from a host of prior studies demonstrating positive associations between higher-level STEM coursetaking and mathematics proficiency (e.g., Allen, 2015; Byun et al., 2015; Ma & Wilkins, 2007; Noble & Schnelker, 2007; Noble et al., 2006; Lee et al., 1997; Madigan, 1997). A key contribution of this paper is to confirm that such findings extend to a large, recent, and nationally representative student sample using proficiency measures from a well-established assessment program with low stakes.

Although this study was not designed to identify causal effects of coursetaking, the findings are also consistent with previous research showing positive causal effects of advanced coursetaking on achievement, including studies such as Leow et al. (2004), Long et al. (2012), and Byun et al. (2015). Among these, the Byun et al. study is the most similar to this one, in that it employs a nationally representative, longitudinal dataset—the Education Longitudinal Study of 2002 (ELS:2002)—and focuses on an indicator for having taken mathematics courses beyond Algebra I as its main predictor of interest. Although the authors use an assessment specific to ELS:2002 as the outcome rather than the grade 12 NAEP assessment, their results are remarkably similar to our own. Using propensity-score matching methods, they estimate an average effect of roughly two-thirds of a standard deviation (0.63-0.74) for courses beyond Algebra I; the analogous weighted mean of coefficients from Model 5 in Table 6 yields an estimate of about 21 points, or 0.64 standard deviation. This may suggest that, at least where the “highest mathematics course” predictors are concerned, our estimates may only be mildly biased when interpreted as causal effects, perhaps due to the inclusion of highly explanatory covariates and the high pseudo-R^2 of Model 5 (0.86).

Another major contribution of the current study is the coursetaking profiles of students with high mathematics proficiency. We found that these students could be partitioned into four large, distinct clusters with differing background characteristics and coursetaking patterns. One of the clusters was identified as having disadvantaged background characteristics, including very low SES and the highest percentage of Black and Hispanic students. However, this group may have been able to mitigate some of these disadvantages via strong coursetaking: specifically, students in this cluster (Cluster C) were much more likely to take some version of calculus (41.6%) as their highest math course than were those in Clusters B (25.0%) and D (approximately 0%). On average, they also earned more than twice as many AP/IB Mathematics credits and were slightly or substantially more likely to complete chemistry or physics.

Another interesting result from the cluster analyses was that, for some students, advanced coursetaking was not required to perform at a college-prepared level. Students from Cluster D had the lowest average number of AP/IB credits in mathematics and science and were the least likely to take chemistry or physics; moreover, almost none of these students took calculus, and 98% did not reach precalculus. Given that these students tended to come from high-SES backgrounds, one explanation for this result might be that they were more likely than others to receive out-of-school instruction, such as instruction from private instructors or highly educated parents.

We also note that all of the clusters had mean grade 9 algebra scores that were well above the full-sample average, indicating that the students in these clusters were highly proficient in mathematics relative to their peers even before completing any high school courses.

Limitations

Although, as noted above, Model 5 explains a large proportion of NAEP score variance and includes results consistent with previous causal studies, this remains a purely observational study and, therefore, causal inferences should not be drawn. In particular, noncognitive traits and environmental factors, such as motivation, persistence, school safety, academic expectations, and instructional quality, have not been accounted for. For example, if highly motivated students are both higher-achieving and more likely to pursue advanced coursework than others, and these motivational differences are not fully captured by covariates, then the effects of higher-level mathematics courses may be overestimated. Similarly, if Black students tend to enroll in schools or courses with lower academic standards or instructional quality, then Black students with the same observed coursetaking and GPA as others would likely score lower, leading to a negative coefficient for Black students.

Because the overlap sample appears somewhat higher-achieving than the full HSLS sample, likely because it was restricted to students who were not retained in grade and did not drop out, the results may not fully generalize to the entire population of U.S. high school students. For example, if students who are likely to be retained or drop out tend to attend schools with less effective instruction, then higher-level courses may not improve their mathematics ability, and our estimates would be larger than those that would be obtained from the full population. On the other hand, our estimates would be smaller if advanced coursetaking has a more pronounced effect for lower-SES students, as suggested by some studies (Long et al., 2012; Byun et al., 2015).

Course availability is also not accounted for by our models. If substantial numbers of high-ability students would like to take higher-level courses but are unable to due to lack of availability, then our parameter estimates for lower-level courses may be inflated, which may lead to attenuated estimates of the differences between lower- and higher-level courses. In addition, the standard errors may be larger than those that would be obtained from a dataset where all courses were available to all students. Differences in course availability may affect our estimates in more complex ways, as well; for example, while about 12% of our overlap sample enrolled in some version of calculus, the corresponding percentage for Black students was only 7%, while the percentages for White and Asian students were 23% and 36%, respectively. Parameter estimates for course-related variables and race/ethnicity may both be affected in unanticipated ways if such differences are due in large part to differences in course availability between student subgroups.

Conclusions

Our findings provide evidence that coursetaking in mathematics and science is significantly and meaningfully related to mathematics proficiency among U.S. students, even after accounting for pre-high school factors, which explain much of the overall variation. Our results, while observational, are in line with those from prior causal studies, suggesting that advanced coursework in mathematics and science may help to improve mathematics proficiency, although the effects are likely much smaller than the observed differences among students with different levels of coursework.

Additionally, our cluster analyses revealed that high-performing students exhibit substantial diversity with respect to coursetaking and background characteristics. While high-performing students from low-SES backgrounds tend to have advanced coursework, many students from higher-SES backgrounds do not. These results suggest multiple paths to high end-of-high school proficiency in mathematics, some of which may not include advanced courses.

References

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.

Adelman, C. (2006). *The Toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

Allen, J. (2015). *Influence of achievement in core high school courses on ACT scores* (ACT Technical Brief). https://www.act.org/content/dam/act/unsecured/documents/2015-Tech-Brief-Influence-of-Achievement.pdf

Altonji, J. G., Blom, E., & Maghir, C. (2012). *Heterogeneity in human capital investments: High school curriculum, college major, and careers* (NBER Working Paper #17985). Cambridge, MA: National Bureau of Economic Research.

American Institutes for Research (2011). AM Statistical Software. https://am.air.org/

Briggs, D. C. (2008). Using explanatory item response models to analyze group differences in science achievement. *Applied Measurement in Education*, *21*(2), 89–118.

Byun, S. Y., Irvin, M. J., & Bell, B. A. (2015). Advanced math course taking: Effects on math achievement and college enrollment. *Journal of Experimental Education*, *83*(4), 439–468.

Chen, X. (2009). *Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education* (NCES 2009-161). Washington, DC: National Center for Education Statistics.

Cohen, J. D., & Jiang, T. (1999). Comparison of Partially Measured Latent Traits across Nominal Subgroups. *Journal of the American Statistical Association*, *94*(448), 1035–1044. https://doi.org/10.1080/01621459.1999.10473856

Cribbs, J., Hazari, Z., Sonnert, G., & Sadler, P. (2015). Establishing an explanatory model for mathematics identity. *Child Development*, *86*(4), 1048–1062.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th Edition). Chichester, United Kingdom: Wiley & Sons, 54–56, 111–130.

Fields, R. (2014). *Towards the National Assessment of Educational Progress (NAEP) as an indicator of academic preparedness for college and job training*. Washington, DC: National Assessment Governing Board. https://www.nagb.gov/content/nagb/assets/documents/what-we-do/preparedness-research/NAGB-indicator-of-preparedness-report.pdf.

Friedman, L. & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple regression. *The American Statistician*, *59*(4), 127–136.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*, 857–871.

Horn, L., & Nuñez, A. M. (2000). *Mapping the road to college: First-generation students' math track, planning strategies, and context of support* (NCES 2000-153). Washington, DC: National Center for Education Statistics.

Jacob, B., & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives*, *30*(3), 85–108.

Lee, V., Burkham, D., Smerdon, B., Chow-Hoy, T., & Geverdt, D. (1997). *High school curriculum structure: Effects on course-taking and achievement in mathematics for high school graduates. An examination of data from the National Education Longitudinal Study of 1988* (Working Paper Series). Washington, DC: National Center for Education Statistics.

Leow, C., Marcus, S., Zanutto, E., & Boruch, R. (2004). Effects of advanced course-taking on math and science achievement: Addressing selection bias using propensity scores. *American Journal of Evaluation*, *25*(4), 461–478.

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, *49*(2), 285–322.

Ma, X., and Wilkins, J. L. M. (2007). Mathematics coursework regulates growth in mathematics achievement. *Journal for Research in Mathematics Education, 38*(3), 230–257.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*(14), 281–297.

Madigan, T. (1997). S*cience proficiency and course-taking in high school: The relationship of science course-taking patterns to increases in science proficiency between 8th and 12th grades*. Washington, DC: National Center for Education Statistics; Suitland, MD: Bureau of the Census.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*(2), 177–196.

National Center for Education Statistics. (n.d.). National Assessment of Educational Progress. https://nces.ed.gov/nationsreportcard/

National Center for Education Statistics. (2014). *National Assessment of Educational Progress (NAEP) 2013 mathematics and reading grades 4 and 8 assessments restricted-use data files data companion*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics. (2015). *High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and high school transcript data file documentation*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Noble, J., Roberts, B., & Sawyer, R. L. (2006). *Student achievement, behavior, perceptions, and other factors affecting ACT scores* (ACT Research Report 2006-1). Iowa City, IA: ACT, Inc.

Noble, J., & Schnelker, D. (2007). *Using Hierarchical Modeling to Examine Course Work and ACT Score Relationships cross High Schools* (ACT Research Report 2007-2). Iowa City, IA: ACT, Inc.

Schiller, K., & Hunt, D. (2003). *Accumulated disadvantages: Racial and social class differences in high school mathematics course trajectories*. Paper presented at the annual meeting of the American Education Research Association, Chicago, IL.

Schneider, B., Swanson, C. B., & Riegle-Crumb, C. (1998). Opportunities for learning: Course sequences and positional advantages. *Social Psychology of Education*, *2*(1), 25–53.

StataCorp. (2017). *Stata Multivariate Statistics Reference Manual, Release 15*, 134–141. College Station, TX: StataCorp LLC.

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, *50*(5), 1081–1121.

Appendix A

Table A-1. Pairwise correlations between coursetaking variables

Variable	STEM GPA	Highest math course	Math credits	AP/IB Math credits	Science credits	AP/IB Science credits	Engineering credits	STEM credits	SES index	Grade 9 Algebra Score
STEM GPA	1.00									
Highest mathematics course (ordinal)	0.60	1.00								
Math credits	0.23	0.31	1.00							
AP/IB Math credits	0.40	0.63	0.30	1.00						
Science credits	0.37	0.44	0.39	0.31	1.00					
AP/IB Science credits	0.36	0.45	0.12	0.51	0.47	1.00				
Engineering credits	0.06	0.05	0.01	0.000	0.03	0.04	1.00			
STEM credits	0.34	0.39	0.73	0.30	0.75	0.31	0.32	1.00		
SES index	0.30	0.29	0.07	0.22	0.22	0.24	0.02	0.13	1.00	
Grade 9 Algebra score	0.58	0.58	0.10	0.45	0.29	0.41	0.08	0.21	0.33	1.00

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit AIR.ORG.

American Institutes for Research®
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
202.403.5000 | **AIR.ORG**