ACT Working Paper 2017-5

An Empirically-Derived Index of High School Academic Rigor

Jeff Allen, PhD
Edwin Ndum, PhD
Krista Mattern, PhD

June, 2017

Jeff Allen is a statistician in the Research division at ACT. He specializes in longitudinal research linking test scores to educational outcomes and student growth models.

Edwin Ndum is a research scientist in Statistical and Applied Research. He does research on reliability and validity of test scores, differential prediction, and mediating effects of psychosocial attributes.

Krista Mattern is a director in Statistical and Applied Research specializing in the validity and fairness of assessment scores as well as more general issues in higher education such as enrollment, persistence, and graduation.

An Empirically-Derived Index of High School Academic Rigor

Abstract

We derived an index of high school academic rigor by optimizing the prediction of first-year college GPA based on high school courses taken, grades, and indicators of advanced coursework. Using a large data set (n~108,000) and nominal parameterization of high school course outcomes, the high school academic rigor (HSAR) index capitalizes on differential contributions across courses and nonlinear relationships between course grades and first-year college GPA (FYGPA). Test scores from 8th grade were incorporated in the model to isolate the effect of HSAR. High school courses with the largest contributions to FYGPA were English 11, English 10, Chemistry, and Algebra 2. Participation in AP, accelerated, or honors courses increased HSAR. The correlation of the HSAR index and FYGPA was 0.50 and 0.49 in two cross-validation samples. While the HSAR index was the strongest predictor of FYGPA, it only led to a modest improvement in overall prediction when combined with high school GPA (HSGPA) and ACT Composite score. The predictive strength of the HSAR index was consistent across different types of high schools and colleges, and subgroup differences in the HSAR index were smaller than subgroup differences in ACT Composite score. Implications for high school counselors, researchers, and postsecondary student service personnel are discussed.

Keywords: academic rigor, weighted high school GPA, college GPA, college readiness, high school courses, ACT test scores

1. INTRODUCTION

In response to global competition in commerce, industry, science, and innovation, a principal recommendation from *A Nation at Risk* was that students take “The New Basics” high school curriculum to become prepared for college or work after high school (National Commission on Excellence in Education, 1983). The New Basics specified the number of years of high school courses that should be devoted to English, mathematics and computer science, social studies, science, and foreign languages. Following reforms in high school curriculum, the concept of academic rigor gained traction in the 1990s as researchers and educational practitioners recognized variation across high school courses in the meaning of course grades and their ability to signal readiness for college-level courses. Longitudinal research on high school graduates of 1982 established that academic intensity and quality of high school curriculum are the strongest predictors of bachelor’s degree completion, more so than unweighted HSGPA and test scores (Adelman, 1999). It became apparent that simply taking the New Basics was not enough, and that the nature and quality of courses taken determine if students are ready for college and work (ACT, 2004).

Rigor is multi-faceted and can be influenced by the type and sequence of courses taken, intensity and difficulty of the courses, alignment to postsecondary expectations, quality of instruction, and level of student engagement and effort. Through this conceptualization, rigor is an unobservable construct that can vary across schools and classrooms. Further, because it can be influenced by student engagement and effort, two students in the same classroom can experience different levels of rigor. Prior attempts at measuring rigor attend to the facets that can be reasonably measured given the data available, including the labels and special designations of courses taken, such as whether the course is designated as honors, dual enrollment, Advanced

Placement (AP), or International Baccalaureate. Such courses are designed to enhance rigor, but designations do not guarantee rigor. There is also a threat of AP and dual enrollment courses becoming watered down as they have become part of the mainstream curriculum (Dougherty, Mellor, and Jian, 2006; Mangan, 2016).

Academic rigor can be operationalized through differential weighting of grades from high school courses or by awarding bonus points for taking certain types of courses. While unweighted high school GPA is predictive of first-year college GPA (FYGPA), colleges will often employ weighting (or bonus points) for use in admissions or placement decisions (Sadler & Tai, 2007; Clinedinst, Koranteng, & Nicola, 2015). Presumably, the weighted HSGPA will be a better measure of readiness for first-year college courses, both in terms of better alignment to college course content and higher correlations with college grades. Use of weighted HSGPA or bonus points also encourages students to take challenging courses in high school (Klopfenstein and Lively, 2015), rather than choosing easier courses that might lead to a higher unweighted HSGPA. Next, we review some of the prior studies on measures of academic rigor. There is variation across studies in the facets of rigor that are attended to, if and how criterion data are used to establish the scoring method, the methods used to record coursework data, and if and how high school course grades are incorporated into the scoring method.

Prior studies using measures of academic rigor

Acknowledging the difficulties with measuring the unobservable rigor construct, a measure representing academic intensity and quality of high school curriculum was constructed (Adelman, 1999). Intensity was measured by credits earned in core subject areas, while quality was measured by number of advanced placement courses, highest level of mathematics attained, and whether mathematics coursework was mostly remedial. Of all the high school curricula variables studied, highest mathematics course taken had the strongest relationship with bachelor's degree completion, with students who completed a course beyond Algebra 2 experiencing much higher rates of success. The measure of intensity and quality of high school coursework was combined with overall HSGPA, class rank, and test scores to form an overall measure of academic resources. A follow-up study for high school graduates of 1992 reaffirmed the importance of academic intensity and quality of high school curriculum for predicting bachelor's degree attainment (Adelman, 2006). However, the gap between the predictive strength of academic intensity and quality of high school curriculum and high school GPA narrowed.

Some measures of rigor are constructed through empirical relationships between coursework and college outcomes. Sadler and Tai (2007) found that performance in college science courses increased with last high school science course grade, and whether the high school science course was considered honors or AP. Moreover, performance in the college science course was significantly higher when the high school AP exam score was 3 or higher (on a 1-5 scale). Their model suggested adding bonus points to the high school grade based on whether the student was in an honors course, AP course, or AP course with exam score of at least 3.

The academic rigor index (ARI) was developed by relating high school coursework indicators to FYGPA (Wyatt, Wiley, Camara, & Proestler, 2011). For each subject area (English, mathematics, social science, science, and foreign language), points are awarded based on taking courses that distinguish group mean FYGPA. Up to five points are awarded in each subject area, and the rules for awarding points vary across subject areas. Generally, a higher ARI is obtained by taking more courses; taking more honors, AP, or dual enrollment courses; and taking higher-level courses. The ARI is constrained by (1) using the same 0-5 point scale for each subject area, whereas some subject areas may contribute more to college readiness, (2) scoring based on whole number increments, whereas greater granularity might improve predictive performance, and (3) scoring decisions based on distinguishing mean FYGPA one variable at a time instead of maximizing the prediction of FYGPA using all variables simultaneously. Because the ARI does not depend on course grades earned, it is designed to be used in conjunction with HSGPA as a measure of college readiness. A later study found that the index was positively correlated with FYGPA (average r across institutions of 0.25), though less than SAT scores (average multiple r = 0.37) and HSGPA (average r = 0.38). The ARI did not explain variance in FYGPA beyond that already explained by SAT scores and HSGPA (Mattern & Wyatt, 2012).

A follow-up study examined less restrictive scoring procedures to determine the extent that the predictive performance of the ARI could be improved (Beatty, Sackett, Kuncel, Kiger, Rigdon, Shen, & Walmsley, 2012). Two empirically-based scoring methods were used: the vertical percent method (Strong, 1926) and multiple linear regression. The two methods were applied to 395 indicators of course rigor (capturing if each course was taken, when it was taken, and whether it was an AP, honors, or dual-enrollment course). The index based on the multiple linear regression model outperformed the index based on the vertical percent method, and offered some improvement over the ARI. The multiple regression-based index had a larger cross-validated correlation with FYGPA than the ARI (0.381 vs. 0.329) and larger increase in R^2 over a base model with SAT total score and HSGPA (0.007 vs. 0.000).

Other approaches for operationalizing rigor do not use criterion data (such as FYGPA) to determine the weights for course grades or the bonus points to be awarded. Recently, a graded response model (GRM) has been proposed to obtain an alternative weighting of HSGPA (Hansen, Sadler, & Sonnert, 2016). This model is less restrictive than unweighted HSGPA and some other formulations of weighted HSGPA because it allows for differences in difficulty across high school courses, allows for the difference between letter grades to vary for each course (e.g., the difference between A and B can be different than the difference between B and C), and allows the reliability of grades to vary across courses. The weighting system under this model is not determined by relationships with other criterion (such as college grades or test scores), and so may predict better across a larger set of outcomes. The GRM-weighted GPA was a stronger predictor of college calculus grade (r = .355) than unweighted GPA (r = 0.332).

It's also possible to account for differences in difficulty across both courses and high schools. Bassiri and Schulz (2003) used ACT test scores as "common items" across high schools to create a high school difficulty scale using a model based on Item Response Theory. They found that the relationship of unweighted HSGPA with FYGPA depended on the grading standards of the high school and that HSGPA adjusted for difficulty of courses and school outperformed unweighted HSGPA for predicting FYGPA (R^2 = 0.25 vs. 0.17).

Focus of current study

In this study, we develop a new index that is designed to isolate the effect of HSAR from prior academic achievement and maximize the prediction of FYGPA. Potentially, the new index can serve as an outcome of the high school academic experience, a complementary predictor (along with HSGPA and college admissions test scores) of college outcomes, and an early warning measure for academic distress. The index is developed using a fully empirical approach, rather than a hybrid rational-empirical approach used in some prior studies.

Our approach is similar to that of Beatty et al. (2012) in three important respects: (1) we use multiple linear regression with many predictor variables to optimize the prediction of FYGPA, (2) high school course-taking information is collected in a standardized fashion through the administration of a large-scale college admissions test, and (3) we do not attempt to measure certain aspects of rigor, including quality of instruction, course intensity, course difficulty, and alignment with college courses. Our approach also differs from Beatty et al. (2012) in three primary respects: (1) 8th grade test scores are included in the development model to better isolate the effects of HSAR, (2) high school grades, in addition to coursework indicators, can contribute to the index, and (3) the nature of the high school course data collected is quite different. The data available support a limited conceptualization of high school academic rigor (HSAR) because some determinants of rigor (e.g., quality of instruction, course intensity and difficulty) are not incorporated. We address the following research questions:

RQ1: Which high school courses relate most to FYGPA, either by simply taking the course or by earning higher grades in the course?

RQ2: Does the HSAR index provide incremental prediction of ACT Composite score above HSGPA and 8th grade test scores?

RQ3: Does the HSAR index provide incremental prediction of FYGPA above HSGPA and ACT Composite score?

RQ4: Does the relationship of the HSAR index and FYGPA vary by high school or college type?

RQ5: How much does the HSAR index vary across sociodemographic subgroups?

2. Methods

2.1. Sample

The sample included 108,381 students who took the ACT Explore test in 8th grade, the ACT test in 11th or 12th grade, and enrolled in college long enough to earn a FYGPA. Coursework and grades data was collected from students when they took the ACT test in high school, and first-year GPA was provided by 425 colleges and universities. To be included in the sample, students must have responded to a majority of the 30 coursework questions collected in high school and must have been enrolled at a high school in the United States when they took the ACT test. The students completed high school between 2006 and 2013.

The sample can be compared to the population of all 11th grade high school students in the United States (Table 1) on gender, race/ethnicity, geographic region, high school type, and high school locale. The sample is mostly female (60%) while the population is evenly split on gender. The sample contains a much higher percentage of White students (77%) than the population (53%), and lower percentages of African American (10% vs. 15%), Asian (2% vs. 6%), and Hispanic (4% vs. 23%) students. The sample is predominantly from the Midwest (51%) and South (47%), with very little representation from the northeast and western United States. The geographic disparity is due to (1) the ACT test being the predominant college admissions test used in the Midwest and South, and the SAT test being more common in the Northeast and West, and (2) greater participation in ACT postsecondary research services among postsecondary institutions from the Midwest and South regions. The sample includes students from Catholic schools (7%), private schools (3%) and public schools (90%). Students from public schools with higher poverty concentration (percent of students eligible for free or reduced lunch) are underrepresented. Relative to the population, the sample contains relatively more students attending schools in rural or town locales (40% vs. 30%) and fewer in suburban or urban rural locales (60% vs. 70%).

Table 1. Comparing the Sample to the Population of U.S. Students in Grade 11

Variable	Sample %	Population %
Gender		
Female	59.6	49.4
Male	40.4	50.6
Race/ethnicity		
African American	10.4	15.0
Asian	2.3	5.6
Hispanic/Latino	4.1	23.1
Other	2.8	3.4
White	77.3	52.9
Unknown	3.3	
Region[1]		
Midwest	51.1	22.3
Northeast	0.1	16.8
South	47.3	37.6
West	1.6	23.4
School type		
Catholic	6.6	3.2
Private	3.2	4.1
Other	0.5	3.1
Public, <20% FRL	23.8	15.8
Public, 20-40% FRL	30.5	23.8
Public, 40-60% FRL	23.1	23.0
Public, 60-80% FRL	8.6	15.8
Public, >80% FRL	2.9	10.4
Public, FRL unknown	0.8	0.8
School locale		
Rural	21.2	18.6
Town	18.4	11.5
Suburban	35.4	40.8
City	25.1	29.2

Note: Population percentages for gender, race/ethnicity, and school locale reflect public school students only

[1] Definition of geographic region used by U.S. Census Bureau

Among the 425 colleges and universities, 81 were more selective[2] 4-year institutions, 241 were less selective 4-year institutions, and 103 were 2-year institutions. The student sample size was 32,292 for selective 4-year, 66,208 for less selective 4-year, and 9,881 for 2-year.

2.2. Data

2.2.1 Middle school test scores

ACT Explore is a standardized achievement test typically taken in 8th or 9th grade, including tests of English, mathematics, reading, and science (ACT, 2013). Subject area scores range from 1-25 and the Composite score is calculated as the mean of the four subject area scores. Explore is intended for all students in grades 8 and 9 and focuses on the knowledge and skills that are usually attained by grade 8. The four subject area scores are used in the derivation of the HSAR index, and the Composite score is used in regression models addressing the research questions.

2.2.2 High school coursework, grades, and test scores

High school coursework and grades are collected when students register for the ACT test in high school. For 30 different courses, students are asked if they 1) have taken the course (or are currently taking the course), 2) have not taken the course but plan to later, or 3) have not taken the course and will not take it later. For this study, students were classified as having taken a course prior to college if they marked option 1 or 2. Students are also asked to report the grade they earned in each course already taken, with five options (A, B, C, D, or F). The treatment of missing course grade is discussed in Section 2.2.5. The data for each course was coded with up

[2] Selectivity is based on the high school class ranks of their accepted freshmen: The majority of freshmen at highly selective schools are in the top 10%, selective in the top 25%, traditional in the top 50%, and liberal in the top 75% of their high school class. Institutions with open admissions policies accept all high school graduates to limit of capacity. For our analysis, we classified 4-year institutions as more selective (selective or highly selective) or less selective.

to six categories: A) through F) depending on course grade earned, or X) did not take the course. For two courses (English 9 and English 10), virtually all students reported taking the courses and so five categories were used (A through F). For eleven courses,[3] very few students reported F grades, so grades D and F were combined into one category. Table 2 lists the 30 high school courses and provides the course participation rates, mean course grade, and standard deviation of course grades.

HSGPA was determined by averaging grades reported by students across up to 23 core high school courses. When students register for the ACT test, they are also asked whether they have taken advanced placement, accelerated, or honors courses in English, mathematics, social studies, natural sciences, or foreign languages. Binary indicators for each type of advanced coursework were used.

The ACT test is designed to measure academic skills necessary for education and work after high school, and the content of the tests is related to major curriculum areas (ACT, 2014). The ACT test focuses on the knowledge and skills attained as the cumulative effect of school experience and is oriented towards the general content areas of college and high school instructional programs. The ACT Composite score is the average of the four ACT subject area scores from the multiple choice portion of the test (English, mathematics, reading, and science), and each of these scores is reported on a 1-36 scale. The ACT Composite score is used in regression analyses addressing the research questions.

2.2.3 Sociodemographic variables and high school characteristics

Gender, family income level, race/ethnicity, and parent education level were collected when students registered for the ACT test in high school. Family income level was categorized in

[3] English 12, Other English, U.S. History, World History, Other History, American Government, Geography, German, Other Foreign Language, Music, and Drama.

Table 2. Course Participation Rates and Course Grade Statistics

Course	Participation rate	Grades	
		Mean	SD
English 9	1.000	3.50	0.68
English 10	1.000	3.47	0.69
English 11	0.999	3.43	0.72
English 12	0.985	3.57	0.58
Other English	0.289	3.72	0.51
Algebra 1	0.993	3.45	0.76
Algebra 2	0.982	3.29	0.82
Geometry	0.991	3.32	0.80
Trigonometry	0.591	3.36	0.75
Calculus	0.412	3.46	0.65
Other math beyond Algebra 2	0.653	3.42	0.68
Computer Math/ Science	0.173	3.75	0.50
Physical, Earth, General Science	0.842	3.56	0.66
Biology	0.982	3.44	0.72
Chemistry	0.924	3.30	0.80
Physics	0.566	3.43	0.69
U.S., American History	0.998	3.53	0.67
World History, Civilization	0.880	3.56	0.66
Other History	0.360	3.61	0.61
Government, Civics, Citizenship	0.875	3.60	0.62
Economics, Consumer Econ.	0.580	3.60	0.61
Geography	0.516	3.66	0.59
Psychology	0.445	3.62	0.58
Spanish	0.763	3.51	0.72
French	0.183	3.52	0.71
German	0.059	3.54	0.67
Other Language	0.079	3.60	0.63
Art	0.559	3.84	0.44
Music	0.483	3.94	0.28
Drama/Theater	0.205	3.88	0.38
High school advanced coursework	Participation rate		
English	0.671		
Mathematics	0.609		
Social Studies	0.544		
Natural Sciences	0.563		
Foreign Languages	0.258		

four levels: <$36K, $36-$60K, $60-$100K, and $>100K. Race/ethnicity was categorized as African American, Asian, Hispanic, White, and other (including Native American and two or more races). For students in the latest three cohorts (2011-2013), paternal and maternal education level was collected and the higher of the two values was used, with categories of high school or less, some college less than a bachelor’s degree, bachelor’s degree, and graduate study or higher.

Information about the high school the student attended was obtained from the National Center for Education Statistics (NCES) Common Core of Data (Glander, 2016) and an additional database of schools (Market Data Retrieval, http://schooldata.com/). Variables included school category (Catholic, private, public, or other), percent of students eligible for free or reduced lunch (FRL, available only for public schools), class size, and locale (rural, town, suburban, or city). Combining school category and school FRL%, high school type was coded as: Catholic, private, home, other (e.g., state or county-operated schools), public < 20% FRL, public 20-40% FRL, public 40-60% FRL, public 60-80% FRL, and public > 80% FRL.

2.2.4 First-year college GPA

Through research partnerships or participation in research services offered by ACT, postsecondary institutions provide ACT with FYGPA for first-year students. For students who persisted through the first year of college, FYGPA represents performance through the spring semester. For students who dropped out after the first semester of college, fall GPA is carried forward. FYGPA was collected from 425 colleges and universities and coded on the usual 4-point scale.

2.2.5 Missing data and imputation

High school coursework and grade data could be missing because: (1) students reported taking the course, but did not provide the grade they earned, (2) students reported currently taking the course, so did not provide the final grade they earned, (3) students hadn't taken the course yet but planned to, 4) students did not respond for the course in question. For cases (1) through (3), we classified students as having taken the course before college, and imputed their course grade based on their overall high school GPA and test scores using a multiple imputation procedure (Berglund, 2010).

For case (4), we know neither if they took the course nor their course grade. The prevalence of this type of missing information varied by course, with an overall rate of 14% averaged across the 30 courses. Course classification (A-F course grades, or X = course not taken) was imputed using logistic regression modeling. For each course, the probability of each classification was first determined based on test scores, overall HSGPA, and classifications from other courses. Then, course classification was imputed using the modeled probabilities and uniform random number generation. The course data was imputed sequentially, so that the imputation for each course was informed by the data from previously-imputed courses.

2.3 Derivation of high school academic rigor index

We sought an index of academic rigor that was a function of the high school coursework and grades data available to us. We envisioned the index being a weighted function of the coursework and grades data, such that courses and grades with greater importance would receive larger weights. We estimated the HSAR index by regressing FYGPA on the 30 high school course classification variables, as well as the 5 indicators for advanced coursework. By using nominal coding for the high school courses, the model makes no assumptions of the ordered importance of the coursework and grades. This allows, for example, for the difference between earning an A and B in Algebra 2 to be larger than the difference between earning a B and C in

Algebra 2, and for the difference between earning an A and B in Chemistry to be larger than the difference between earning an A and B in American Government.

Students with higher pre-high school academic achievement are more likely to take challenging courses in high school and earn higher grades. To isolate the effect of high school coursework and grades on FYGPA, the middle school test scores were also included in the regression model used to develop the HSAR index. The index is therefore designed to predict FYGPA based on high school data, net of the effects of pre-high school academic achievement.

A hierarchical linear model with random intercepts for college was used to account for nesting of students within colleges. By fitting the regression model, we obtain the weights needed for the HSAR index scoring. The HSAR index for the *i*th student is then calculated by Equation 1:

Equation 1: HSAR index

$$\text{HSAR}_i = \sum_{c=1}^{30}\left(\sum_{k=1}^{5}\beta_{ck}I_{ick}\right) + \sum_{v=1}^{5}\theta_v I_{iV}$$

where $\beta_{ck}$is the regression coefficient for the *c*th course (*c*=1,2,..,30) and *k*th grade category (*k*=1,2,3,4 or 5)[4]; I_{ick}=1 if the *i*th student's classification for course *c* is *k* and I_{ick}=0 otherwise; θ_v is the regression coefficient for the *v*th advanced coursework content indicator (*v*=1,2,3,4,5[5]); and I_{iv}=1 if the *i*th student took advanced courses in the *v*th subject and I_{iv}=0 otherwise. Note that, while the middle school test scores were used in the regression model used to estimate the regression weights needed for the HSAR index, they are not included in the HSAR index

[4] Course grades were coded as A = "1", B = "2", C = "3", D = "4", and F = "5". As discussed earlier, some courses have fewer than 6 response categories and the coding and equation is adjusted accordingly.
[5] Advanced coursework content indicator was coded as English = "1", mathematics = "2", social sciences = "3", natural sciences = "4", and foreign language = "5".

scoring. The HSAR index can be decomposed into subject-specific components (English, mathematics, social studies, natural science, foreign language, and the arts) by summing only the terms within each subject. Moreover, the contribution for each course is obtained by summing only the terms specific to that course.

We first estimated the HSAR index weights using the total sample. The total sample results reflect our best estimates of the weights that would be used in practice for future cohorts of students. RQ1 (*Which high school courses relate most to FYGPA, either by simply taking the course or by earning higher grades in the course?*) is addressed by comparing the weights across the high school courses. The predictive strength of a multiple linear regression model is summarized by R^2, which is the proportion of variance in the dependent variable explained by the model's predictors and can be written as the variance of the model's predicted values, divided by the variance of the dependent variable. The variance (or standard deviation) of each course's contribution to the HSAR index (Equation 1) can therefore serve as a proxy of each course's predictive strength. For each predictor, we used the standard deviation of the predictor's contribution as a summary measure of the course's contribution to FYGPA: Courses with the largest standard deviations have the strongest relationships with FYGPA. The total sample results are also examined for counterintuitive results, such as cases where a higher course grade leads to a lower HSAR score.

The HSAR index requires 147 weights (regression coefficients) – 142 for the high school coursework and grades classifications and 5 for the advanced coursework indicators. Large samples are needed to estimate the index, and the performance of the index could be sensitive to the sample used to estimate the weights. To test the predictive strength of the HSAR index, we used a cross-validation approach whereby the total sample was split in two according to students' cohort year. Sample 1 included students who completed high school from 2006 to 2009 (n=49,782), and sample 2 included students who completed high school from 2010 to 2013 (n=58,599). The HSAR index weights applied to sample 2 were estimated using sample 1, and the weights applied to sample 1 were estimated using sample 2. The courses that were removed from the model during the total sample analysis were also removed for the cross-validation analysis. This cross-validation approach protects from detecting artificially strong correlations between the HSAR index and FYGPA, and mimics how the HSAR index could be applied in practice.

2.4 Correlations and regression models

Total group means, standard deviations, and correlations of key variables (8th grade Explore Composite score, HSAR index total score, HSAR index subject-specific scores, HSGPA, ACT Composite score, and FYGPA) are presented. The correlations and results of the regression models are used to address research questions 2-4. Table 3 shows the relationship between the research questions and the regression models, and provides the specifications (dependent variable, predictor variables, and samples) for each model. When FYGPA is the dependent variable, random intercept models are used to account for clustering within colleges.

While the HSAR index is designed to predict FYGPA, RQ2 examines whether the HSAR index also predicts ACT Composite score, above and beyond the effects of pre-high school academic achievement and HSGPA. RQ3 examines how well the HSAR index predicts FYGPA, above and beyond the effects of pre-high school academic achievement and traditional predictors of FYGPA (HSGPA and ACT Composite score). These regression models are fit separately for sample 1 and sample 2, allowing us to assess the consistency of results across independent samples.

Table 3. Regression Model Specifications for Addressing Research Questions 2-4

Research question (RQ)	Dependent variable	Predictor variables	Sample
2. How well does the HSAR index predict ACT Composite score?	ACT Composite score	8th grade Explore Composite score, HSAR index, HSGPA	sample 1, sample 2
3. How well does the HSAR index predict FYGPA?	FYGPA	Model 1: HSGPA, ACT Composite score Model 2: HSGPA, ACT Composite score, HSAR index	sample 1, sample 2
4. Does the relationship of the HSAR index and FYGPA vary by high school or college type?	FYGPA	Model 1: HSAR index, high school type, HSAR index*high school type Model 2: HSAR index, college type, HSAR index*college type	Total

RQ4 examines the consistency of the HSAR index as a predictor of FYGPA across high school and college settings. To put the consistency of the HSAR index into context, we also examine the consistency of ACT Composite score as a predictor of FYGPA across high school and college settings. The regression models for RQ4 are fit using the total sample.

To examine RQ5 (*How much does the HSAR index vary across sociodemographic subgroups?*), we calculated the mean and standard deviation of the HSAR index by selected sociodemographic subgroups defined by gender, race/ethnicity, family income, parent education level, and school type. We present standardized mean differences (*d*) relative to selected reference groups. This analysis is repeated for ACT Composite score to put the results for the HSAR index into context. The analyses for RQ5 are conducted for the total sample.

2.5 Study limitations

A major limitation of the study is its reliance on students' self-reported coursework and grades. Prior research examined how often student-reported coursework agreed with actual coursework obtained from high school transcript data (Sanchez & Buddin, 2016). The accuracy of self-reported coursework varied considerably across the 30 high school courses, with a median percent correct of 94%, percent over-report of 5%, and percent under-report of 1%. Agreement of self-reported grades and actual grades also varied across the 30 courses, with median exact agreement rate of 67%, percent within one letter grade of 96%, percent under-reporting of 20%, and percent over-reporting of 14%. While the HSGPA measure is based on student self-report, it is highly correlated with HSGPA obtained from high school transcripts (r=0.84, Sanchez & Buddin, 2016). The performance of the HSAR index may have been adversely affected by student misreporting of coursework and grades.

Another study limitation is that the results are only directly applicable to the high school coursework and grades data as collected by the ACT test registration system. This system does not accommodate all possible high school courses, only distinguishes whole letter grades (A vs. B instead of A- vs. B+), and does not reveal other important aspects of each high school courses such as honors, AP, or dual credit designation. Additional research is needed to examine this approach using official high school transcript data and other systems for collecting student-reported coursework and grades.

The study used FYGPA as the basis of the HSAR index and for testing the predictive performance of the HSAR index. FYGPA is an important outcome because of its proximity to high school, its relevance to all first-year students, and its strong relationship with later college outcomes such as degree attainment. However, it's possible that an index designed to predict other outcomes, such as degree attainment or GPA at graduation, would improve prediction of those outcomes. FYGPA measures performance across a varying mix of college courses, so is not a standardized measure. Some students have easier first-year course loads, and this may confound the relationship of college readiness measures and FYGPA.

A large sample was used to estimate the HSAR index and examine its usefulness. The sample was not representative of the population of U.S. high school students, with Hispanic students underrepresented. In addition, very few students or colleges from the Northeast and Western regions of the United States were included in the sample.

3. RESULTS

3.1. Which high school courses relate most to FYGPA, either by simply taking the course or by earning higher grades in the course?

Participation rates varied extensively across the 30 courses, with virtually all students (>99%) taking English 9, 10, and 11; Algebra 1, Geometry, and U.S. History (Table 2). Courses with very low participation rates (< 10%) included German and other foreign language. Upper-level mathematics courses (Trigonometry, Other Advanced Math, and Calculus) had moderate participation rates (41-65%). Courses with the lowest average grades included Algebra 2 (3.29), Chemistry (3.30), Geometry (3.32), and Trigonometry (3.36). Courses with lower average grades had the greatest variation in grades earned. Generally, courses with higher participation rates had lower average grades, perhaps due to having higher concentrations of lower-achieving students. Participation in advanced, AP, or honors courses varied across subject areas from 26% in foreign languages to 67% in English (Table 2).

The multiple linear regression model used to estimate the HSAR index was estimated for the total sample, sample 1 (student cohorts of 2006-2009), and sample 2 (student cohorts of 2010-2013). The regression coefficients (and standard errors) for predicting FYGPA for the total sample are presented in Appendix Table A1. The results for the total sample are used to address the first research question and to identify counterintuitive results.

The courses with the largest contributions to FYGPA include English 11 (SD = 0.084), English 10 (SD = 0.057), Algebra 2 (SD = 0.051), Chemistry (SD = 0.051), English 12 (SD = 0.037), Other Advanced Math (SD = 0.037), and Trigonometry (SD = 0.035). Courses can have large contributions due to effects of simply taking the course, effects of earning higher grades in the course, or both reasons. The courses with the largest contributions to FYGPA, with the exception of Other Advanced Math and Trigonometry, were taken by the overwhelming majority of students in the sample. Therefore, the large contributions to FYGPA are due to the effects of earning higher grades in the course.

The difference between earning an A and C in English 11 was 0.234 points (FYGPA units), the largest among all courses. Some courses had very small course grade differentials. In American Government, for example, the difference between earning an A in C was only 0.034 FYGPA units. The difference between earning an A in Algebra 2 and not taking Algebra 2 was 0.144 FYGPA units, and the difference between earning an A in Chemistry and not taking Chemistry was 0.108 FYGPA units. Surprisingly, there was a small negative difference (-0.016) between earning an A in Physics and not taking Physics. There was some evidence of nonlinear relationships between course grades and FYGPA. In English 9 for example, the difference between earning a D and F was 0.191 FYGPA units, but the difference between earning an A and B was only 0.037 units.

Appendix Table A1 also reveals some counterintuitive results where a higher course grade is related to lower FYGPA. For example, predicted FYGPA is lower for students earning a C in Calculus (beta = -0.073) relative to students earning a D (beta = 0.003). Predicted FYGPA is lower for students earning an A in Computer Science (beta = -0.034) relative to students earning a B (beta = -0.001) or C (beta = 0.051). Predicted FYGPA is higher for students earning a C in both Economics and Geography, relative to students earning an A or B in those courses. Most of the counterintuitive weights are small.

Figure 1 shows the distribution of the HSAR index in the total sample. The distribution is skewed to the left, but not as severely as the distribution of unweighted HSGPA (Figure 2). The sample mean was 2.08, standard deviation 0.42, minimum -0.74 and maximum 2.84. Recall that the HSAR index is calculated as the predicted FYGPA, less the contribution of the 8th grade test scores. Therefore, the HSAR index scores are not inherently interpretable.

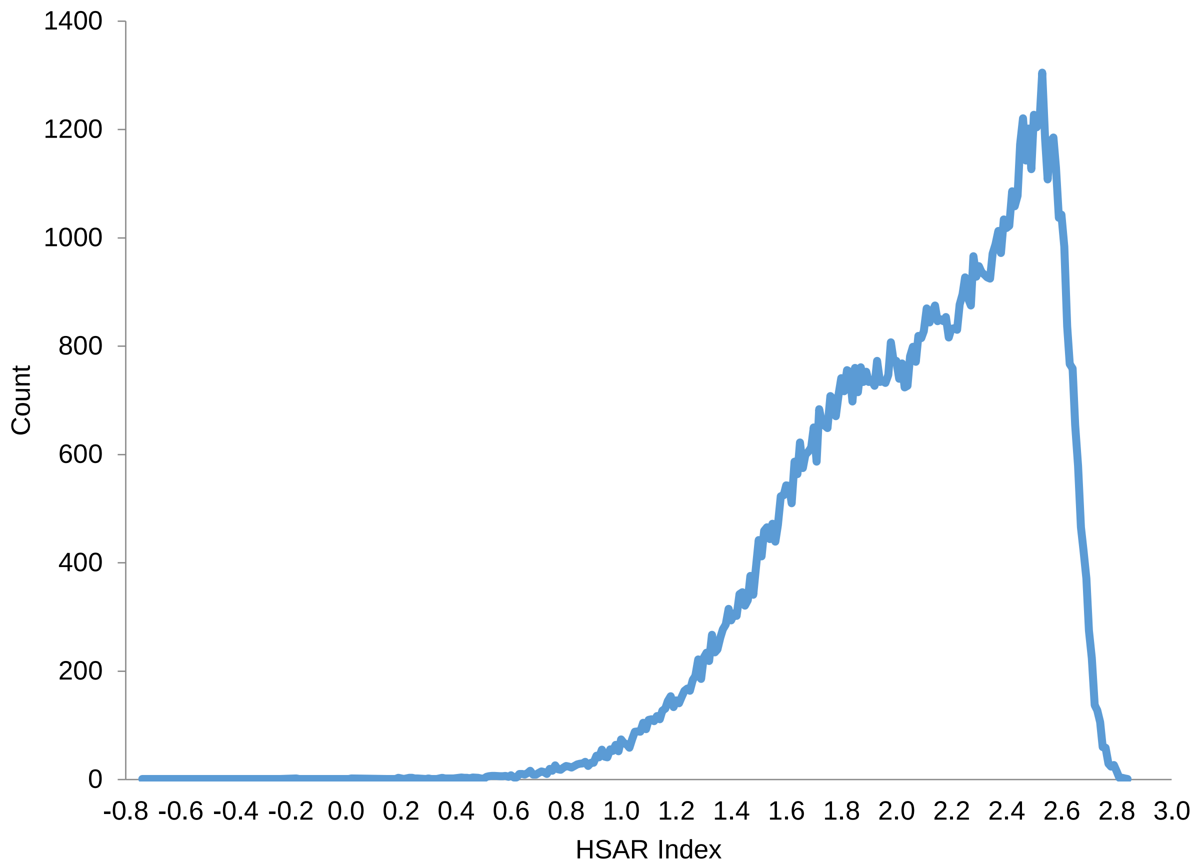

Figure 1. Frequency distribution of the HSAR index in the total sample

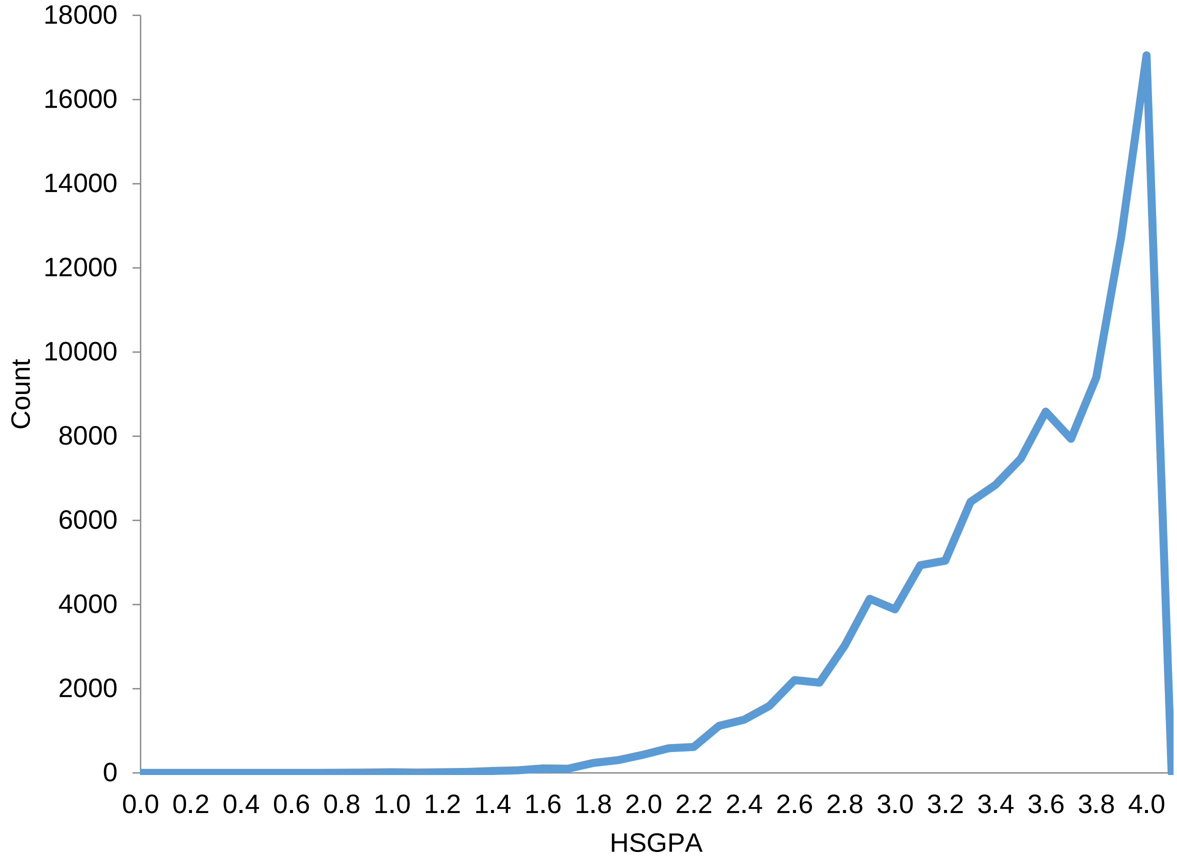

Figure 2. Frequency distribution of HSGPA in the total sample

3.2. Does the HSAR index provide incremental prediction of ACT Composite score?

Because both the HSAR index and ACT Composite score are impacted by high school coursework and the rigor of the high school experience, we expect the two measures to be correlated. The total group correlation of the HSAR index and ACT Composite score was 0.553 (Appendix Table A2); by comparison, the correlation of HSGPA and ACT Composite score was 0.515. Regression models for ACT Composite score were fit that included 8th grade Explore Composite score, HSGPA, and the HSAR index as predictors (Table 4). ACT Explore Composite score was the strongest predictor (beta = 0.665 for both sample 1 and sample 2), followed by the HSAR index (beta = 0.181 for sample 1, beta = 0.199 for sample 2). HSGPA had a small negative beta weight with ACT Composite score in both sample 1 and sample 2 (beta = -0.030). The HSAR index helps explain ACT Composite score, beyond what is explained by 8th grade Explore score and HSGPA.

Table 4. Regression Models for ACT Composite Score

Predictor	*r*	*β*	*SE*
Sample 1 (n = 58,599)			
8th grade ACT Explore Composite score	0.798	0.665	0.003
HSAR index	0.519	0.181	0.007
HSGPA	0.493	-0.030	0.007
Model *R*	0.815		
Sample 2 (n = 49,782)			
8th grade ACT Explore Composite score	0.813	0.665	0.003
HSAR index	0.557	0.199	0.007
HSGPA	0.524	-0.030	0.007
Model *R*	0.829		

Notes. *r* = correlation coefficient; *β* = Beta weight; *SE* = Standard Error.

3.3. Does the HSAR index provide incremental prediction of FYGPA?

By design, we would expect the HSAR index to be correlated with FYGPA, and that was the case (*r* =.499 for sample 1, *r* =.491 for sample 2) (Table 5). The HSAR index was a stronger predictor of FYGPA than HSGPA, which had correlations of 0.481 for sample 1 and 0.474 for sample 2. In the regression models for FYGPA, the HSAR index was the strongest predictor (beta = 0.284 for sample 1, beta = 0.309 for sample 2), followed by ACT Composite score (beta = 0.189 for sample 1, beta = 0.195 for sample 2). HSGPA was also a significant predictor of FYGPA (beta = 0.114 for sample 1, beta = 0.104 for sample 2).

Table 5. Regression Models for First-year College GPA

Predictor	Model 1		Model 2	
	β	*SE*	β	*SE*
Sample 1 (n = 58,599)				
HSGPA	0.385	0.004	0.114	0.010
ACT Composite score	0.209	0.004	0.189	0.004
HSAR index			0.284	0.010
Model R	0.515		0.526	
Sample 2 (n = 49,782)				
HSGPA	0.369	0.005	0.104	0.011
ACT Composite score	0.221	0.005	0.195	0.005
HSAR index			0.309	0.012
Model R	0.508		0.516	

Notes. β = Beta weight; *SE* = Standard Error.

Using only ACT Composite score and HSGPA as predictors, the model's multiple correlation (*R*) was 0.515 in sample 1 and 0.508 in sample 2. Adding the HSAR index to the model resulted in multiple *R* of 0.526 for sample 1 and 0.516 for sample 2. The HSAR index was the strongest predictor of FYGPA, and provided a small improvement over ACT Composite score and HSGPA for predicting FYGPA.

3.4. Does the relationship of the HSAR index and FYGPA vary by high school or college type?

Because the HSAR index does not attempt to capture quality of instruction or other within-school effects, it is possible that the strength of relationship between the HSAR index and FYGPA varies systematically by type of high school. To assess whether the strength of relationships varies by high school type, we tested the interaction of high school type and the HSAR index. The overall interaction test was significant ($p < 0.001$), indicating that the relationship of the HSAR index and FYGPA varies by high school type (Table 6). The interaction model provides separate beta weights by high school type, which can be used to interpret the size and meaning of the interaction.

Table 6. Beta Weights of College Readiness Measures, by High School and College Type

Predictor	HSAR Index		ACT Composite score	
	β	*SE*	*β*	*SE*
Model 1: High school type				
Catholic	0.466	0.010	0.342	0.012
Private	0.460	0.015	0.358	0.016
Other	0.548	0.035	0.426	0.032
Public, <20% FRL	0.450	0.006	0.310	0.006
Public, 20-40% FRL	0.513	0.005	0.384	0.005
Public, 40-60% FRL	0.532	0.005	0.406	0.006
Public, 60-80% FRL	0.483	0.008	0.422	0.010
Public, >80% FRL	0.401	0.015	0.358	0.019
Model 2: College type				
2-year	0.458	0.008	0.381	0.010
4-year, less selective	0.513	0.004	0.410	0.004
4-year, more selective	0.470	0.006	0.340	0.006

Notes. *β* = Beta weight; *SE* = Standard Error.

Across all types of high schools, the beta weights were positive. The beta weights were generally higher for public schools with moderate poverty levels (0.513, 0.532, and 0.483), relative to public schools with low poverty (0.450), public schools with very high poverty (0.401), private (0.460), and Catholic schools (0.466).

Interactions between high school type and ACT Composite score were also tested to provide a basis of comparison with the HSAR interactions. Similar to the HSAR index, the relationship of ACT Composite score and FYGPA varied across high school type. The beta weights for ACT Composite score tended to be higher for moderate-poverty public high schools. Overall, the interaction analyses suggest that the predictive strength of both the HSAR index and ACT Composite score generalize across different types of high schools.

We also examined whether the HSAR index functions the same across college settings. We tested the interaction of the HSAR index and college type (2-year, 4-year less selective, 4-year more selective). While the interaction test was significant, the beta weights were similar across college settings, ranging from 0.458 for 2-year colleges to 0.513 for less selective 4-year colleges.

3.5. How much does the HSAR index vary across sociodemographic subgroups?

Differences across student subgroups were observed for the HSAR index, and the differences were generally smaller than those observed for ACT Composite score (Table 7). While females scored lower on ACT Composite score than males (d = -0.18), they scored higher on the HSAR index (d = 0.28). African American (d = -0.75), Hispanic (d = -0.33), and students of other race/ethnicity (d = -0.18) scored lower on the HSAR index than White students. Mean HSAR index scores also varied by family income level, with higher-income students scoring higher. Mean HSAR index scores increased with parent education level. HSAR index scores were much lower for students whose parent(s) earned only a high school diploma or less (d = -0.47), compared to those whose parent(s) had earned higher than a bachelor's degree. Relative to students from low-poverty public schools, HSAR index scores were higher for students at private (d = 0.24) and Catholic (d = 0.10) high schools. HSAR index scores were lowest for students from high-poverty public high schools (d = -0.44).

HSAR index differences across race, family income level, parent education level, and high school type were smaller than those observed for ACT Composite score. Differences in HSAR help explain differences in ACT Composite scores across racial/ethnic groups. For example, the mean difference in ACT Composite score for White and African American students was 4.66 points. After adjusting for the HSAR index using multiple regression, the difference is 3.00 points. After adjusting for the HSAR index, 8th grade Explore Composite score, family income, parent education level, and school type, the difference is only 0.79 points.

Table 7. Subgroup Differences in College Readiness Measures

Group	ACT Composite			HSAR index		
	M	*SD*	*d*	*M*	*SD*	*d*
Female	23.22	4.28	-0.18	2.12	0.40	0.28
# Male	24.00	4.53		2.01	0.44	
Race/ethnicity						
African American	19.50	3.55	-1.06	1.80	0.42	-0.75
Asian	24.39	4.28	0.05	2.16	0.39	0.10
Hispanic/Latino	22.13	4.05	-0.46	1.98	0.43	-0.33
Other	23.63	4.53	-0.12	2.04	0.43	-0.18
# White	24.16	4.20		2.12	0.40	
Family income						
<$36,000	21.29	4.17	-0.92	1.93	0.44	-0.60
$36,000-60,000	23.03	4.20	-0.52	2.05	0.42	-0.31
$60,000-100,000	24.23	4.13	-0.25	2.13	0.40	-0.12
# >100,000	25.34	4.05		2.18	0.38	
Parent education level						
High school or less	22.03	3.94	-0.84	2.04	0.41	-0.47
Some college, less than Bachelor's	23.09	4.02	-0.60	2.10	0.39	-0.33
Bachelor's	24.74	4.04	-0.23	2.19	0.37	-0.12
# Graduate study or more	25.73	3.98		2.24	0.35	
High school type						
Catholic	24.81	4.08	0.01	2.12	0.40	0.10
Private	25.18	4.25	0.09	2.18	0.39	0.24
Other	25.33	5.03	0.13	2.08	0.41	0.01
# Public, <20% FRL	24.78	4.14		2.08	0.42	
Public, 20-40% FRL	23.86	4.24	-0.21	2.10	0.42	0.04
Public, 40-60% FRL	22.68	4.26	-0.48	2.07	0.42	-0.02
Public, 60-80% FRL	21.25	4.14	-0.80	1.99	0.44	-0.22
Public, >80% FRL	19.39	3.74	-1.23	1.90	0.43	-0.44

Notes. *M* = Mean; *SD* = Standard Deviation; *d* = Cohen's *d;* # Indicates reference group

4. DISCUSSION

4.1 Effects of high school courses on FYGPA

One of the surprising results of the study was that high school English courses, particularly English 10 and 11, had the strongest contributions to FYGPA. First-year college courses demand strong writing skills, and the vast majority of students take at least one English course during the first year (Radunzel, Westrick, Bassiri, and Li, 2017). Course grades in high school English courses carry a strong signal of readiness for the writing demands of first-year college courses. They might also signal higher motivation and stronger worth ethic, as prior research has shown that motivation has a stronger effect on success in first-year English courses, relative to math courses (Robbins et al., 2006). Students may have extra motivation for performing well in English 10 because, in many states, an end-of-course exam must be passed. College-bound students may be highly motivated in grade 11 college-prep courses due to their stakes for college admissions.

The study results affirm the importance of performing well in upper-level mathematics courses. Prior research showed the importance of taking upper-level mathematics courses for earning a bachelor's degree (Adelman, 1999, 2006). Gartner et al. (2014) found a positive effect of taking Algebra 2 on FYGPA and cumulative college GPA, but that Algebra 2 completion did not affect college and career outcomes to the same degree. Over 98% of the students in our sample took Algebra 2, so our analysis was concerned more with the effects of earning higher grades in Algebra 2 rather than the effect of taking Algebra 2. Taking higher-level math courses (Trigonometry, Other Advanced Math, and Calculus) and earning good grades (A or B) was associated with higher FYGPA.

Performance in high school science courses – particularly Chemistry and Biology – also helped predict FYGPA. Across the core subject areas, the social studies courses contributed least to FYGPA. Performance in the two social studies courses with the highest participation rates (U.S. History and World History) was most important.

Taking accelerated, AP, or honors courses in social studies, English, and foreign language was also related to higher FYGPA. Surprisingly, the effect of taking advanced mathematics and science courses was very small. These results could be due to advanced mathematics and science students taking more difficult first-year college courses, leading to lower FYGPA. The relationship between taking advanced courses and FYGPA may be confounded by the difficulty of the first year courses. This might also explain why we did not observe positive effects of taking high school Physics. Additional research is needed to examine this issue by accounting for the difficulty of the mix of first-year courses.

4.2 Performance of the HSAR index

Because it was designed to optimize the prediction of FYGPA, it was not surprising that the HSAR index outperformed both HSGPA and ACT Composite score as a predictor of FYGPA, with a correlation of about 0.50. It performed similarly across two cross-validation samples, and also helped explain variation in ACT Composite score.

While the HSAR index was the strongest predictor of FYGPA, it only resulted in a modest increase over ACT Composite score and HSGPA in a multiple regression model, with multiple R increasing by about 0.01 in both samples. Consistent with prior studies, the prediction of FYGPA mostly plateaus once HSGPA and college admissions test scores are included in the model. While additional significant predictor variables (e.g., student motivation, HSAR) can be identified, the model's overall predictive strength does not increase much.

Some variation in the performance of the HSAR index was observed across high school and college settings. The predictive strength was strongest at public schools with moderate poverty levels (20-80% of students eligible for free or reduced lunch), and slightly weaker at public schools with low poverty levels, public schools with the highest poverty levels, and private and Catholic schools. This general pattern was also observed for ACT Composite score.

The HSAR index was most predictive at less selective 4-year colleges. Generally, the variation in predictive strength across school settings was small.

Subgroup differences were observed for the HSAR index. The subgroup differences were generally in the same direction as ACT Composite score differences, but smaller in magnitude. One exception was for gender, where females outscored males on the HSAR index (d=0.28) but scored lower than males on the ACT test (d = -0.18). Prior studies have shown underprediction of first-year college grades for female students based on ACT Composite score and HSGPA (Mattern, Sanchez, and Ndum, 2017), and this result suggests that female underprediction would be reduced with inclusion of the HSAR index in the model. While ACT test scores are a cumulative measure of knowledge and skills attained through grade 11 or 12, the HSAR index attempts to only measure the high school academic experience. Therefore, the larger subgroup differences on the ACT test could be due to having a longer time period for differences to accumulate. It is also possible that HSAR index subgroup differences are smaller because differences in grading standards across high schools are not controlled. ACT test scores, on the other hand, have the same meaning across high schools.

4.3 Possible uses of the HSAR index

The results of the study can be used to identify the high school courses with the strongest relationships with FYGPA, net of the effects of pre-high school academic achievement. This information, in turn, can serve two purposes: 1) academic advising to college-bound students on which courses to take and perhaps which to focus engagement on, 2) guidance to researchers, survey developers, and college admissions personnel on which high school courses are most important to study. The HSAR index itself can be used for evaluation of the high school experience, or within a multiple measures college readiness model that includes HSGPA and ACT or SAT test scores.

4.4 Directions for additional research

Additional research is needed to address the study's limitations, and to expand the scope of the measurement of rigor. As discussed earlier, limitations include reliance on students' self-reported coursework and grades; lack of detailed data on each course's designation as honors, AP, or dual credit; the use of FYGPA without accounting for college course difficulty; and underrepresentation across racial/ethnic groups and geographic regions. These limitations could be jointly addressed by collecting official high school and first-year college course transcript data for a large and diverse sample of students. By addressing these limitations, bias related to student self-report and course difficulty could be reduced.

We did not attend to all aspects of rigor, such as course content and high school and classroom effects. Additional research is needed to test the performance of measures that attend to additional aspects of rigor.

Appendix

Table A1. Total Sample Regression Model for Estimating HSAR Index

Predictor	Regression Coefficients and Standard Errors						SD
Intercept	1.071 (0.159)						
8th grade ACT Explore test scores							
English	0.014 (0.001)						0.054
Mathematics	0.009 (0.001)						0.028
Reading	0.008 (0.001)						0.029
Science	0.015 (0.001)						0.042
High school courses	A	B	C	D	F	Not taken	
English 9	0.363 (0.102)	0.325 (0.102)	0.269 (0.102)	0.191 (0.104)	REF		0.032
English 10	0.220 (0.083)	0.135 (0.082)	0.061 (0.082)	-0.043 (0.085)	REF		0.057
English 11	0.213 (0.070)	0.095 (0.070)	-0.021 (0.07)	-0.124 (0.073)	-0.237 (0.093)	REF	0.084
English 12	0.108 (0.020)	0.044 (0.020)	-0.01 (0.025)	-0.093 (0.052)	-	REF	0.037
Other English	0.013 (0.006)	0.005 (0.010)	0.015 (0.032)	-0.112 (0.091)	-	REF	0.006
Algebra 1	-0.011 (0.029)	-0.021 (0.029)	-0.035 (0.030)	-0.035 (0.034)	-0.062 (0.083)	REF	0.008
Algebra 2	0.144 (0.020)	0.067 (0.020)	0.025 (0.020)	0.009 (0.024)	-0.102 (0.049)	REF	0.051
Geometry	0.026 (0.027)	-0.010 (0.027)	-0.026 (0.028)	-0.004 (0.032)	-0.155 (0.067)	REF	0.022
Trigonometry	0.075 (0.007)	0.009 (0.007)	-0.019 (0.011)	-0.052 (0.027)	-0.128 (0.066)	REF	0.035
Calculus	0.044 (0.008)	-0.029 (0.008)	-0.073 (0.016)	0.003 (0.051)	-0.041 (0.147)	REF	0.026
Other math beyond Algebra 2	0.081 (0.007)	0.035 (0.007)	-0.007 (0.012)	-0.062 (0.035)	-0.403 (0.095)	REF	0.037
Computer Math/ Science	-0.035 (0.008)	-0.001 (0.014)	0.051 (0.041)	-0.100 (0.110)	-	REF	0.013
Physical, Earth, General Science	-0.056 (0.008)	-0.057 (0.009)	-0.048 (0.013)	-0.087 (0.031)	-0.368 (0.107)	REF	0.022
Biology	0.082 (0.020)	0.048 (0.020)	-0.014 (0.021)	-0.010 (0.030)	-0.048 (0.086)	REF	0.031
Chemistry	0.108 (0.011)	0.027 (0.011)	-0.017 (0.012)	-0.024 (0.020)	-0.161 (0.046)	REF	0.051
Physics	-0.016 (0.007)	-0.043 (0.007)	-0.047 (0.013)	-0.122 (0.036)	-0.202 (0.089)	REF	0.020
U.S., American History	0.098 (0.053)	0.049 (0.053)	-0.004 (0.053)	-0.042 (0.058)	-	REF	0.033

World History, Civilization	-0.038 (0.008)	-0.068 (0.009)	-0.087 (0.014)	-0.095 (0.029)	-	REF	0.023
Other History	-0.012 (0.006)	0.000 (0.009)	-0.033 (0.019)	-0.101 (0.053)	-	REF	0.008
Government, Civics, Citizenship	0.008 (0.009)	-0.013 (0.009)	-0.026 (0.015)	-0.073 (0.036)	-	REF	0.012
Economics, Consumer Econ.	0.001 (0.006)	-0.006 (0.008)	0.031 (0.017)	-0.035 (0.048)	-0.197 (0.154)	REF	0.007
Geography	-0.044 (0.006)	-0.049 (0.009)	-0.004 (0.018)	-0.007 (0.045)		REF	0.022
Psychology	0.035 (0.006)	0.025 (0.008)	-0.013 (0.02)	-0.049 (0.060)	-0.104 (0.136)	REF	0.017
Spanish	0.019 (0.008)	-0.052 (0.009)	-0.081 (0.012)	-0.098 (0.026)	-0.231 (0.058)	REF	0.036
French	0.006 (0.009)	-0.039 (0.012)	-0.069 (0.022)	-0.050 (0.053)	-0.136 (0.105)	REF	0.013
German	-0.009 (0.014)	-0.063 (0.019)	-0.043 (0.04)	-0.091 (0.088)	-	REF	0.009
Other Language	0.017 (0.011)	-0.012 (0.017)	-0.006 (0.038)	0.028 (0.091)	-	REF	0.004
Art	-0.011 (0.005)	-0.083 (0.011)	-0.102 (0.025)	-0.221 (0.062)	-0.31 (0.118)	REF	0.024
Music	-0.012 (0.005)	-0.075 (0.018)	-0.063 (0.045)	-0.223 (0.088)	-	REF	0.013
Drama/Theater	-0.013 (0.006)	-0.044 (0.019)	-0.144 (0.049)	-0.147 (0.103)	-	REF	0.011
High school advanced coursework	Yes	No					
English	0.047 (0.006)	REF					0.022
Mathematics	0.006 (0.006)	REF					0.003
Social Studies	0.066 (0.006)	REF					0.033
Natural Sciences	0.007 (0.006)	REF					0.003
Foreign Languages	0.046 (0.006)	REF					0.020

Notes. *SD* = Standard Deviation; REF = reference group; predictors in bold are not significant (p>0.05)

Table A2. Correlation Matrix and Descriptive Statistics

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. 8th grade ACT Explore Composite	1.000										
2. HSAR index - Total	0.501	1.000									
3. HSAR index - English	0.408	0.892	1.000								
4. HSAR index - Mathematics	0.430	0.832	0.623	1.000							
5. HSAR index - Social Studies	0.391	0.778	0.601	0.617	1.000						
6. HSAR index - Natural Science	0.338	0.683	0.576	0.465	0.480	1.000					
7. HSAR index - Foreign Language	0.297	0.604	0.476	0.456	0.431	0.357	1.000				
8. HSAR index - Arts	0.164	0.356	0.262	0.218	0.231	0.228	0.195	1.000			
9. HSGPA	0.475	0.944	0.864	0.799	0.751	0.668	0.539	0.302	1.000		
10. ACT Composite score	0.806	0.553	0.436	0.494	0.435	0.361	0.317	0.181	0.515	1.000	
11. FYGPA	0.356	0.507	0.448	0.422	0.394	0.351	0.304	0.175	0.483	0.408	1.000
Mean	17.710	2.076	0.746	0.141	0.049	0.018	-0.012	-0.023	3.450	23.553	2.837
SD (Standard Deviation)	2.781	0.420	0.171	0.127	0.078	0.059	0.040	0.031	0.500	4.400	0.954

References

ACT (2004). *Crisis at the core*. Iowa City, IA: ACT.

ACT (2013). *ACT Explore technical manual.* Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/Explore-TechManual.pdf.

ACT (2014). *The ACT technical manual*. Iowa City, IA: ACT.

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

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

Bassiri, D. & Schulz, M. (2003). Constructing a universal scale of high school course difficulty. *Journal of Educational Measurement, 40*(2), 147-161.

Beatty, A. S., Sackett, P. R., Kuncel, N. R., Kiger, T. B., Rigdon, J. L., Shen, W., & Walmsley, P. T. (2012). *A comparison of alternate approaches to creating indices of academic rigor.* (College Board Research Report 2012-11). New York, NY: The College Board.

Berglund, P. A. (2010). *An introduction to multiple imputation of complex sample data using SAS v9.2.* Ann Arbor, MI: Institute for Social Research-University of Michigan. Retrieved from http://www.mwsug.org/proceedings/2010/stats/MWSUG-2010-21.pdf.

Clinedinst, M., Koranteng, A. M., & Nicola, T. (2015). *2015 state of college admission.* Arlington, VA: National Association for College Admission Counseling. Retrieved from https://www.nacacnet.org/news--publications/publications/state-of-college-admission/.

Dougherty, C., Mellor, L., & Jian, S. (2006). *Orange juice or orange drink? Ensuring that "advanced courses" live up to their labels.* Austin, TX: National Center for Educational Accountability.

Gardner, D. P. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Department of Education, National Commission on Excellence in Education.

Glander, M. (2016). *Documentation to the 2014–15 Common Core of Data (CCD) universe files* (NCES 2016-077). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/ccd/pdf/2016077_Documentation_062916.pdf.

Hansen, J. D., Sadler, P. M., & Sonnert, G. (2016). *Watching our weights: Using a graded response model to estimate high school course rigor.* Retrieved from https://pdfs.semanticscholar.org/0d46/5b07e40b6a8038efab9e4a79178116dd528b.pdf.

Klopfenstein, K., & Lively, K. (2015). Do grade weights promote more advanced course taking? *Education Finance and Policy, 11*(3), 310-324.

Mangan, K. (2016, July 22). As dual enrollments swell, so do worries about academic rigor. *The Chronicle of Higher Education.* Retrieved from http://www.chronicle.com.

Mattern, K. D. & Wyatt, J. N. (2012). *The validity of the academic rigor index (ARI) for predicting FYGPA.* (College Board Research Report 2012-5). New York, NY: The College Board.

Mattern, K. D., Sanchez, E., Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice, 36*(1), 47-57.

Radunzel, J., Westrick, P., Bassiri, D., & Li, D. (2017). *Development and validation of a preliminary ELA Readiness Benchmark based on the ACT ELA score*. Iowa City, IA: ACT.

Robbins, S. B., Allen, J., Casillas, A., Peterson, C. H., & Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology, 98*(3), 598–616.

Sadler, P. M. & Tai, R. H. (2007). Accounting for advanced high school coursework in college admissions decisions. *College and University Journal, 82*(4), 7-14.

Sanchez, E. & Buddin, R. (2015). *How accurate are self-reported high school courses, course grades, and grade point averages*? Iowa City, IA: ACT.

Strong, E. K., Jr. (1926). An interest test for personnel managers. *Journal of Personnel Research, 5*, 194–203.

Wyatt, J. N., Wiley, A., Camara, W. J., & Proestler, N. (2011). *The development of an index of academic rigor for college readiness.* (College Board Research Report 2011-11). New York, NY: The College Board.