RESEARCH REPORT 2014-2

CollegeBoard

Does the Level of Rigor of a High School Science Course Matter?

An Investigation of the Relationship Between Science Courses and First-Year College Outcomes

By Pamela K. Kaliski and Kelly E. Godfrey

Pamela K. Kaliski is an associate psychometrician at the College Board.

Kelly E. Godfrey is a research scientist at the College Board.

Acknowledgment

The authors would like to acknowledge and thank Brian Patterson for his review of this manuscript and suggestions for revision.

Correspondence concerning this article should be addressed to Pamela Kaliski, College Board Research, 661 Penn Street, Suite B, Newtown, PA 18940.
Email: pkaliski@collegeboard.org.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools.

For further information, visit www.collegeboard.org.

RESEARCH

For more information on College Board research and data, visit research.collegeboard.org.

Contents

Executive Summary

The focus of this research is to evaluate the relationship between advanced high school science courses and college outcomes, with a focus on the benefit of Advanced Placement Program® (AP®) participation and performance in comparison to other high school options (e.g., dual enrollment, honors, and regular science high school courses). Although there is a plethora of previous research comparing achievement-related outcomes of AP participation with other levels of high school course participation (e.g., honors course participation, dual enrollment), there has yet to be a study that compares achievement-related outcomes of AP, honors, dual enrollment, and regular science courses when using a rigorous statistical model. Multilevel modeling was employed to account for the nested structure of the data (i.e., students nested within colleges), and several student-level and college-level covariates were included in the model to provide a stringent test of these relationships. Overall, results suggest that higher performance in advanced science course work (e.g., AP) in high school is related to higher first-year GPA and higher first-year science GPA; as such, teachers should encourage students to participate in advanced high school science course work when planning their curriculum for science learning.

Introduction

In 1991, the National Center for Improving Science Education released its report, *The High Stakes of High School Science*, which outlined a call for science education reform, including requiring all students to take four years of science. Claiming that science education empowers citizens, the report emphasized the need to prepare students for an increasingly technology-driven workforce. Over two decades later, the recent release of the Next Generation Science Standards (NGSS) (Next Generation Science Standards, 2013), which was driven by the need to prepare young Americans for success in the global economy, provides yet another example of how science education remains critical. U.S. policymakers are increasingly concerned over the competitiveness of Americans in science. High school students are not only taking more science courses before graduating, but they are also taking higher levels of science (Dalton, Ingels, Downing, & Bozick, 2007). High school science courses of varying levels of rigor are often available for students to take; specifically, many students might be able to choose between Advanced Placement®, honors, dual enrollment, International Baccalaureate (IB), and a regular non-advanced science course when planning their high school curriculum.

With the implementation of the Common Core State Standards (CCSS) (Common Core State Standards Initiative, 2012), a current national educational priority is to prepare students for college-level course work. The CCSS Initiative Mission Statement articulates this goal clearly:

> *The Common Core State Standards provide a consistent, clear understanding of what students are expected to learn, so teachers and parents know what they need to do to help them. The standards are designed to be robust and relevant to the real world, reflecting the knowledge and skills that our young people need for success in college and careers. With American students fully prepared for the future, our communities will be best positioned to compete successfully in the global economy.* (Common Core State Standards Initiative, 2012)

As such, research that informs science education about the benefits of the various science course-taking patterns as they relate to college outcomes is essential. Moreover, the current Secretary of Education, Arne Duncan, emphasized the importance of student participation in advanced courses such as AP in high school, in order to develop a world-class education system in the United States (U.S. Department of Education, 2009). This study fills a gap in the current science education research by using rigorous statistical methods for examining the relationship between various levels of science courses (e.g., AP, honors, dual enrollment, regular) and first-year college outcomes, and in doing so, it can assist high school teachers in advising their students about what level of high school science course work will yield the greatest benefits in regard to college outcomes. The results of this study can increase the understanding of advanced science course work in relation to various college outcomes.

Theoretical Framework

Previous research has examined the relationship between participating in high school science course work and various outcomes. For example, Sadler and Tai (2007a) investigated the relationship between science course taking and college science performance and found that taking courses in biology, chemistry, and physics in high school led to positive outcomes in corresponding college courses; namely, performance in introductory biology, chemistry, and physics courses. In addition, researchers have found that taking more science courses leads to higher mathematics scores on the SAT® (Brody & Benbow, 1990). Trusty (2002) investigated the effects of advanced science and math courses on students' likelihood of choosing a science or math major in college using the National Education Longitudinal Study of 1988-94. Results indicated that course taking in high school does influence choosing a math or science major in college. For women, taking one calculus course more than doubled their chances of majoring in math or science. For men, taking physics in high school was a significant predictor (Trusty, 2002). To summarize, results suggest that there is a positive relationship between high school science course work participation and achievement-related outcomes (e.g., college course grades) as well as interest-related outcomes (e.g., choice of major).

However, within a particular science subject, such as biology, a student typically is faced with several choices in regard to level of rigor of the course when selecting which biology course to take. That is, there are several levels of rigor in high school science courses. These include, but are not limited to, AP, International Baccalaureate (IB)[1], dual enrollment, honors, and regular, non-advanced courses[2]. AP allows high school students to enroll in courses designed to cover the same material as the respective entry-level college courses. At the completion of the course, students take an exam for the opportunity to earn college credit. IB's Diploma Program offers students an advanced curriculum designed to help them succeed in college and beyond (International Baccalaureate, 2014). Unlike AP, assessments are administered at the conclusion of the program, as opposed to the completion of the course. Dual enrollment is another opportunity for students to earn college credit while completing high school course requirements. These students are instructed by either high school teachers or faculty members from higher education institutions, and the course counts simultaneously toward high school graduation requirements and college credits to degree (O'Brien & Nelson, 2004). Honors courses are offered during high school and fulfill high school graduation requirements, but they do not count toward college degree requirements, and they are usually taught at a more advanced level than regular courses. When these options are available for students to consider while they are designing their high school curriculum, it begs this question: What

1. International Baccalaureate is not included in the current study, due to unavailable data and lack of ability to partition out science effects as opposed to overall IB program effects.

2. One important difference between AP and honors that can complicate the comparison between AP and honors for some AP subjects — for example, biology and chemistry — is that AP Biology and AP Chemistry are second-year courses. That is, a student would take either regular biology or honors biology before taking AP Biology. In these cases, the student is deciding if he or she should take AP after having already taken one year of introductory course materials (either honors or regular).

are the benefits of taking a regular course in biology? How do the benefits compare to those of either an honors course, an AP course, or a dual enrollment course? In other words, the critical decision is not typically whether or not the student should take biology, but rather *which* biology course the student should take, and if the student should also take AP when he or she has already taken a regular or honors course[3]. As such, the above research that investigated the overall impact of science participation is not quite complete without a more detailed investigation regarding the level of science courses. When we refer to "level" of science course for the remainder of the paper, we are referring to the courses of varying levels of rigor — AP, honors, dual enrollment, and regular courses — as opposed to specific science subjects (e.g., biology, chemistry).

Importance of Advanced Course Work

Educators have long been exploring the impact of more rigorous course work on various college academic outcomes. Adelman (1999) found that the more advanced courses students take in high school, the more likely they are to graduate college. Roderick, Nagaoka, Coca, and Moeller (2009) found that almost two thirds of the students from advanced programs graduate high school with access to selective or highly selective four-year colleges and universities. However, these researchers did not compare the different options of advanced course work to one another. A brief review of previous research that has focused on comparing different options of course work is described below.

Some researchers have explored the role of advanced course work, particularly AP, in course taking and achievement. Of these, some researchers have examined AP more holistically as a program rather than specific AP subjects. For example, Chajewski, Mattern, and Shaw (2011) recently found that the odds of enrolling in a four-year institution are at least 171% greater for students who took at least one AP Exam, compared to students who did not take AP Exams.

Other researchers have conducted empirical studies that disaggregate AP subjects (e.g., AP Biology, AP Chemistry). Morgan and Klaric (2007) compared non-AP students who took the introductory college course to AP students, who received a score of a 3, 4, or 5, regarding both groups' subsequent course grade in college (i.e., the course that immediately follows the introductory course). Morgan and Klaric analyzed data from 10 AP Exams, including AP Biology and AP Chemistry. When controlling for SAT scores, the pattern for the AP Biology Exam was as expected — students who received an AP score of 5 had the highest subsequent course grade, followed by students who received an AP score of 4, followed by students who received an AP score of 3, followed by students who did not take AP. The AP4 and AP5 students were significantly different from the non-AP students. For AP Chemistry, although the same pattern was observed, the differences between the AP3, AP4, and AP5 students were not statistically significant.

Keng and Dodd (2008) conducted a more extensive study, looking at not only the outcome of subsequent course grade but also at overall college credit hours, overall college GPA, first-year credit hours, first-year GPA, subject credit hours, and subject GPA. They compared four different groups on each outcome: AP Credit (scores of 3, 4, and 5), AP Non-Credit (scores of 1 and 2), Concurrent Enrollment (synonymous for dual enrollment), and Non-AP students.

3. We realize that there are situations in high schools where students do not have much control over which course they take. For example, if a teacher or counselor recommends a student take honors instead of AP, the student might take honors for this reason. In these contexts, we believe the results of this study can assist guidance counselors who are making recommendations and decisions for students regarding their high school science curriculum.

AP Biology and AP Chemistry were the two sciences included in this study. The AP Credit group consistently outperformed other groups on all outcomes, with the exception of subject credit hours, where the dual enrollment group outperformed other groups. However, this study was descriptive in nature and not based on rigorous statistical models.

Other researchers have conducted similar work comparing honors science students to AP and non-AP students but with more of a specific focus on science. For instance, Sadler and Tai (2007b) focused on the relationship between AP Exam scores and introductory science college course grades, comparing those who took AP to those who did not. Their outcome of interest was introductory science course grades, regardless of subject (i.e., biology, chemistry, physics). The researchers found that AP students who scored a 4 or 5 on the exam tended to outperform other students in introductory science courses (which is to be expected, since taking the introductory course in college after AP in high school is essentially retaking the course). Students who took the AP course and did not score a 3 or higher on the exam tended to do no better than students taking honors or regular science courses at the high school level (Sadler & Tai, 2007b). However, conclusions were based on aggregated science courses as opposed to individual subjects studied separately, and sample sizes were small and not necessarily representative of the general population. Sadler and Sonnert (2010) expanded this work by reanalyzing a subset of data from Sadler and Tai (2007b), comparing AP to other levels of course work (i.e., honors and regular science courses) and examining the effect for three different subject areas (i.e., biology, chemistry, and physics), rather than aggregating across subjects. Results showed that, as expected, students who "passed" the AP Biology, Chemistry, or Physics exam — earning a score of 3, 4, or 5 — earned significantly higher grades in their introductory college courses across all three science areas than students with other experiences besides AP. AP students who "failed" the AP Exam (i.e., earned a 1 or 2 on the exam) did not earn higher grades from a statistical significance standpoint; however, for two subjects (biology and chemistry), the effect sizes for students in the AP fail group were much higher than those for students in the regular course group.

Although most existing research on rigorous course work is focused on AP, educators have recently begun to explore the effects of dual enrollment programs on students' college outcomes in more detail. In 2011, Speroni used data from the state of Florida to determine the effects of participating in dual enrollment programs on students' access to college and bachelor's degree attainment. Speroni found that although dual enrolled students were more likely than AP students to enroll in college after high school, they were less likely to first attend a four-year institution, and she found no statistically significant differences in degree attainment. Dual enrollment students were not found to be significantly different from AP students if the dual enrollment course was taught on the high school campus. In 2012, researchers in Texas followed suit and explored the effects of dual enrollment on students in their state (Struhl & Vargas, 2012). The authors used propensity score matching to create matched groups for comparison purposes and found positive findings for dual enrollment students in terms of enrollment, persistence, and graduation from college. However, the lack of accounting for other advanced college-level rigorous programs, such as AP and IB, led to critical limitations in the study's results and implications. In addition, a report was recently released that describes a small study comparing AP and dual enrollment in Indiana (Smith, 2012). Smith describes results from analyses of data from the Advancing Academic Excellence grant study, which is composed of 36 high schools in Indiana whose students were tracked through college. There were statistically significant differences in first-year GPA for students who completed and passed AP and those who completed but did not pass an AP course. However, the differences in first-year GPA between those who completed dual enrollment and those who did not complete dual enrollment were not statistically significant. Most recently, the National Center for Education Statistics (NCES, 2013) compared characteristics of dual enrollment students to exam-based (i.e., AP and IB) students using descriptive statistics. The primary focus of this report was not college outcomes but rather enrollment rates and the prevalence of these courses in public high schools. Finally, recent research at the College Board suggests that, in general, AP students perform as well if not better than students in dual enrollment (Godfrey, Matos-Elefonte, Ewing, & Patel, 2013; Wyatt & Patterson, 2013).

The Purpose of Study

Clearly, there is a plethora of previous research comparing achievement-related outcomes between AP participation to other levels of high school course participation (e.g., honors course participation, dual enrollment). However, there has yet to be a study that compares AP, honors, dual enrollment, and regular science courses on achievement-related outcomes, using one rigorous statistical model. The focus of this research is to evaluate the advantage of participation and performance in AP science classes above and beyond that of other levels of classes in a similar subject (i.e., honors, dual enrollment, regular, or none). The outcome variables of interest in these analyses are first-year college GPA (FYGPA) and first-year science GPA. Specific research questions are as follows:

(1) Do AP students have a higher FYGPA than honors, dual enrollment, regular course, or no course students when controlling for relevant student-level and college-level covariates? If so, does this apply to all five AP score category groups and those who took the course but no exam, or only certain AP score category groups?

(2) Do AP students have a higher science GPA than honors, dual enrollment, regular course, or no course students when controlling for relevant student-level and college-level covariates? If so, does this apply to all five AP score category groups and those who took the course but no exam, or only certain AP score category groups?

Method

Data Sources

There were three primary sources of data used for this study. The first source of data comes from a database that was created by the College Board as part of an ongoing effort to build a national higher education outcomes database. Official college transcript data exist in this database and were used for the current study. Specifically, first-year college transcript data from two cohorts of students (2008 and 2009) were obtained, with original sample sizes of 246,652 for the 2008 cohort and 262,949 for the 2009 cohort[4]. The second source of data is self-reported data from SAT takers; specifically, it is the answers to the SAT Questionnaire (SAT-Q) that students complete when registering for the SAT. This is where student-level demographic variables were obtained, as well as records of students' high school course participation. The third source of data is from the *Annual Survey of Colleges* (College Board, 2012) and contains college-level information.

4. See Patterson and Mattern (2011) and Patterson and Mattern (2012) for more information on participating colleges.

Several criteria were applied to reduce the original cohort sample sizes and prepare the final dataset for analyses. First, listwise deletion of incomplete cases led to the deletion of (a) students who did not take the SAT and (b) students who had no first-year science course work. Also, because we wanted to isolate the relationship with each course type and the outcomes, students who had completed AP, honors, and dual enrollment in one particular science subject area were removed. Moreover, if a student was listed as participating in AP and honors in a given science subject, the student was treated as part of an AP group, and if a student was listed as participating in dual enrollment and honors, the student was treated as part of the dual enrollment group. The two cohorts were then combined into one large dataset. The final sample size contained 257,877 students from 132 institutions. Tables 1 and 2 show the descriptive statistics for student-level and college-level characteristics, respectively. For each of the four subjects, every student was classified (using a self-report survey question on the SAT Questionnaire; more details below) into one of the following 10 groups (see Table 3 for the sample sizes of each group, as well as the gender and racial/ethnic breakdowns for each group):

1. *No Class:* The student did not take a class in that subject area.
2. *AP1:* The student took the AP Exam and scored a 1.
3. *AP2:* The student took the AP Exam and scored a 2.
4. *AP3:* The student took the AP Exam and scored a 3.
5. *AP4:* The student took the AP Exam and scored a 4.
6. *AP5:* The student took the AP Exam and scored a 5.
7. *No AP Exam (NoExam):* The student took the AP course but not the exam.
8. *Regular (Reg):* The student took a regular course in that subject area.
9. *Honors:* The student took the honors course in that subject area.
10. *Dual Enrollment (DE):* The student completed dual enrollment in that subject area.

Measures

What follows is a description of the variables that were included in the statistical model for these analyses.

College outcome variables. There were two college outcome variables of interest in this study: first-year GPA (FYGPA) and science GPA. FYGPA represents the students' grade-point average for all of their first-year courses. Science GPA was computed by calculating the average course grade of any science courses that were taken during the first year. All analyses were conducted twice — once for each of these two outcome variables.

Level of science course. The group comparisons of interest in the current study are among the 10 groups specified above. They are reflected by including dummy-coded variables into the model for the level of science course. Specifically, because there were 10 groups of interest (shown in Table 3), nine dummy-coded variables were added into the statistical model (to be described below); the reference group was the "No Class" group.

Student-level covariates. Many student-level covariates were gathered for these analyses. Specifically, SAT Mathematics (SAT-M), SAT Critical Reading (SAT-CR), and SAT Writing (SAT-W) were pulled from official College Board records. All SAT scores were treated as continuous; SAT section scores range from 200–800 in 10-point intervals. In addition, the students' reported high school grade point average (HSGPA) was gathered from the SAT Questionnaire (SAT-Q), which students complete when they register for the SAT. HSGPA was treated as continuous, and it was based on survey responses to 12 response options where 1 = A+ (97–100), 2 = A (93–96), 3 = A- (90–92), 4 = B+ (87–89), 5 = B (83–86), 6 = B- (80–82), 7 = C+ (77–79), 8 = C (73–76), 9 = C- (70–72), 10 = D+ (67–69), 11 = D (65–66), and 12 = F (65 or below). Demographic variables including gender, racial/ethnic identity, mother's education, and father's education were also obtained from the SAT-Q. All of these demographics variables are categorical and were dummy-coded for the analyses in the current study. Specifically, males were coded as 0, and females were coded as 1; non-white racial/ethnic groups were coded as 0, and whites were coded as 1; for parental education, no college degree or associate degree was coded as 0, and college degree (at least a bachelor's degree) or higher was coded as 1.

Further, the self-reported course work questions on the SAT-Q were used to identify in which of the 10 course groups the students were classified. Specifically, on the SAT-Q, students are given the following instructions: "For each year of secondary school, go down the list of courses and bubble in the ones you took in that year. For every course that is designated as Honors, Advanced Placement Program (AP) or Dual Enrollment (a course that allows you to simultaneously earn high school and college credit), fill in the oval in that column as well. Then fill in courses you plan to take (for example, if you are in 11th grade and plan to take calculus in 12th grade, fill in the corresponding oval). Finally, go down the list and mark the oval in the "None" column for courses that you have not taken and do not plan to take in high school."

College-level covariates. Several college-level covariates representing characteristics of the colleges were also pulled from the *Annual Survey of Colleges* and merged to the student-level dataset. First, college location had three categories: (1) urban (2) suburban, and (3) rural; two dummy codes were created (urban and suburban), with rural as the reference group. Second, college type (public or private) was a categorical variable, with public coded as 0 and private coded as 1. Third, college selectivity, measured in terms of the percent of students admitted, had three categories: (1) high = under 50%, (2) medium = 50%–75%, and (3) low = greater than 75%. Two dummy codes were created (medium selectivity, low selectivity), and high selectivity was the reference group. Fourth, college size had four categories: (1) small = 750 to 1,999 undergraduates, (2) medium to large = 2,000 to 7,499 undergraduates, (3) large = 7,500 to 14,999 undergraduates, and (4) very large = 15,000 or more undergraduates. Three dummy codes were created, and very large was the reference group. Finally, college region had six categories: (1) southwest, (2) midwest, (3) midstate, (4) northeast, (5) southern, and (6) western. The southwest group was specified as the reference group.

Data Analyses

A total of eight analyses were conducted. All analyses were conducted for each of the two outcome variables (FYGPA and science GPA), once for each AP science subject (AP Biology, AP Chemistry, AP Physics, and AP Environmental Science). Given the nested data structure of these data (i.e., students nested within colleges), it is likely that the assumption of independence, which is assumed by many traditional data analytic approaches that answer similar research questions (e.g., ANCOVA), is violated. For example, when students are nested within colleges, the students within one college are likely to be more similar to one another than students within a different college. More specifically, any systematic differences in how grades are assigned to students at one college (e.g., leniency) opposed to another would lead to violations of the independence of the student-level residuals. The violation of this assumption can lead to consequences when interpreting the findings if the selected data analytic approach does not account for this. Multilevel modeling, also known as hierarchical linear modeling (HLM), is an appropriate technique to employ when data are nested because it accounts for the hierarchical structure of the data (e.g., Raudenbush & Bryk, 2002). Thus, multilevel modeling was used in the current study for all analyses because there is a possibility that the assumption of independent observations has been violated due to the nesting of students in colleges. All models were estimated in SAS 9.1, using restricted maximum likelihood estimation.

First, intercept-only multilevel models were estimated for each subject in order to estimate the intraclass correlation (ICC) for the two dependent variables: first-year GPA (FYGPA) and science GPA. The ICC represents the proportion of total variance that is due to between-college variance. Put another way, the ICC captures the extent to which random variation in the outcome is attributable to the college level, and a significance test of the ICC is a test of the assumption of independent student-level residuals. ICC values greater than 0.05 typically warrant the use of HLM. This is because such values indicate that the independence assumption is violated (e.g., Snijders & Bosker, 1999).

Second, a random intercept model was estimated, and all student-level covariates (e.g., SAT scores, HSGPA, gender, racial/ethnic identity, parental education) were entered into the model to control for background characteristics, along with the college-level covariates (e.g., college location, college type, college size, college selectivity, college region) to examine differences in FYGPA and science GPA between levels of these college characteristics. Also, the level of science course group dummy codes that represent in which of the 10 groups a student is classified were entered; this is the variable that answers the two substantive research questions of interest. All student-level covariates and college-level covariates were grand-mean centered in advance. An example model is the following, where FYGPA is modeled, with slope coefficients fixed and intercepts varying randomly across colleges:

$$FYGPA_{ij} = \beta_{0j} + \beta_{1j}\,(SAT\text{-}CR) + \beta_{2j}\,(SAT\text{-}M) + \beta_{3j}\,(SAT\text{-}W) + \beta_{4j}\,(HSGPA) + \beta_{5j}\,(gender) + \beta_{6j}\,(racial/ethnic\ identity) + \beta_{7j}\,(mother's\ education) + \beta_{8j}\,(father's\ education) + \beta_{9j}\,(AP1) + \beta_{10j}\,(AP2) + \beta_{11j}\,(AP3) + \beta_{12j}\,(AP4) + \beta_{13j}\,(AP5) + \beta_{14j}\,(APNone) + \beta_{15j}\,(dual) + \beta_{16j}\,(honors) + \beta_{17j}\,(regular) + rij$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(urban) + \gamma_{02}\,(suburban) + \gamma_{03}\,(college\ type) + \gamma_{04}\,(medium\ selectivity) + \gamma_{05}\,(high\ selectivity) + \gamma_{06}\,(large\ college) + \gamma_{07}\,(medium\ college) + \gamma_{08}\,(small\ college) + \gamma_{08}\,(midwest) + \gamma_{08}\,(midstate) + \gamma_{08}\,(northeast) + \gamma_{08}\,(south) + \gamma_{08}\,(west) + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{17j} = \gamma_{170}$$

where $FYGPA_{ij}$ is student i in college j's FYGPA, β_{0j} is the FYGPA for college j when each of the student-level covariates are equal to the average across all students in all colleges and adjusted for the differences among colleges in all student-level covariates; β_{pj} is the covariate p slope for FYGPA for college j; r_{ij} is the deviation of the individual student i in college j's actual FYGPA from what was expected under the model; γ_{01} is the grand mean intercept and γ_{p1} is the grand mean covariate p slope for FYGPA; and μ_{0j} is the random intercept effect for college j (i.e., its deviation from the grand mean FYGPA intercept). Statistical significance tests of parameters as well as effect sizes (which are defined in this study as the magnitude of the predicted group mean differences on the outcome variable) were examined. As part of the multilevel modeling analyses, statistical significance tests of pairwise comparisons between the groups of interest were also estimated. The difference between any two slope coefficients for the nine dummy-coded groups serves as an unstandardized effect size and indicates the difference in GPA between those two groups, and the slope coefficients themselves are the differences between the relevant group and those having taken no high school course in the relevant area.

Interpretation of the student-level covariates and college-level covariates.

The slope coefficients for the continuous covariates at level 1 (SAT-M, SAT-CR, SAT-W, and HSGPA) are interpreted as the change in either FYGPA or science GPA for every unit change in the predictor, after controlling for the other covariates. The categorical variables (gender, racial/ethnic identity, father's education, mother's education) were all dummy-coded. For gender, because males were specified as the reference group, the gender coefficient equals average female FYGPA or science GPA minus the average male FYGPA or science GPA when controlling for the remaining covariates. For racial/ethnic identity, non-whites were the reference group, so the slope coefficient compares non-whites to whites. For father and mother education, the reference groups were students of fathers and mothers who have less than a bachelor's degree. The slope coefficients for the dummy-coded, college-level covariates are interpreted similarly; that is, the value of the coefficient represents that adjusted average difference between that group and the reference group on the relevant outcome.

Results

The means on the two outcomes of interest — FYGPA and science GPA — are shown in Table 4 for each of the 10 groups. These means do not account for any of the student-level or college-level covariates. Still, it is worth noting the general trends shown by these means in order to establish a baseline, before examining the results of the rigorous models, which include student-level and college-level covariates. As expected, the five means for the five AP Exam score category groups (i.e., AP1, AP2, AP3, AP4, and AP5) increase as AP Exam score increases. Honors students have means similar to students in the AP2 and the AP3 groups. The dual enrollment students, as well as students who took an AP course but no exam, have means similar to students in the AP2 group. Students who took regular science courses have means similar to students in the AP1 and AP2 groups. These trends are the same across all four science subjects and for both FYGPA and science GPA. The takeaway message from these general trends is that students who are in the AP4 and AP5 groups have the highest FYPGAs and science GPAs across all four science exams. However, these means do not account for any of the possible student-level or college-level variables that might contribute to these group differences, which raises the following question: What happens when a rigorous statistical model is applied to the data? The results for research questions 1 and 2, which are based on rigorous statistical models, are presented on the following page.

Research Question 1: FYGPA

The ICC from the intercept-only model with FYGPA as the outcome was 0.143[5]; this indicates the proportion of total variance in FYGPA that is due to between-college variability. This value is considered sufficiently large (i.e., greater than 0.05) and was statistically significant ($p < 0.001$), indicating that there is between-college variation in FYGPA and that multilevel modeling is necessary. The results of the multilevel model that was estimated for each of the four science subject areas are shown in Table 5. Specifically, these are the slope coefficient parameter estimates[6], which represent the relationship between that variable and FYGPA, in units of FYGPA. For example, the slope coefficient for gender shows that females have an average FYGPA 0.183 points higher than males, controlling for all other student-level and college-level covariates included in the model. The values for each of the nine levels of science course groups show the difference between that group and the reference group of No Class. For example, students who receive a score of 5 on the AP Biology Exam have an average FYGPA 0.161 points higher than students who did not take any biology course.

However, our research questions are focused not only on comparing these groups to students in the reference group (i.e., students who do not take a course). It is important that our questions are also focused on how the AP students from all AP score groups compare to the other groups, particularly the honors, dual enrollment, and regular course students. We conducted pairwise comparisons to answer this question. The results of the pairwise comparisons for the four sciences on the outcome of FYGPA are shown in Table 6. The pairwise comparisons revealed positive findings in favor of advanced science course taking, particularly for AP. These pairwise comparisons are based on the model results that were shown in Table 5. As such, they are incredibly stringent pairwise comparisons, given that they are controlling for all student-level and college-level covariates that were incorporated into the multilevel model. Each value in Table 6 can be interpreted as the average difference in FYGPA between those two groups (i.e., the AP Exam group and the comparison group), controlling for all of the student-level and college-level covariates. For example, students who receive a score of 5 on AP Biology have an average FYGPA 0.172 points higher than students who completed a dual enrollment biology course. The criteria used to determine whether or not one group outperforms another group were twofold, and both had to be met: (a) statistical significance — at the $p < 0.05$ level and (b) practical significance, which is defined in the current study as a difference greater than 0.05 in absolute magnitude. Although we are not aware of any established benchmarks for what constitutes a meaningful difference in FYGPA, a difference of 0.05 seems reasonable for this research. We based this on our awareness of meaningful criteria that are set by some institutions to be inducted into various honors societies or honors programs (e.g., Appalachian State University, 2013; New York University, 2013). A summary statement that is based on the expectations of the predictive model follows for each AP Exam group, focusing on which other groups that particular AP Exam group outperformed (if both criteria were not met, the pairwise comparison is not considered to be a meaningful difference and is not summarized below):

1. The AP5 group outperformed all other groups, across each of the four science content areas.

5. Because the ICC is based on an empty model with no other predictors, the one ICC applies to all four science subject areas.

6. Standard errors are available upon request

2. There were 28 pairwise comparisons for the AP4 group across the four science content areas; the AP4 group outperformed the comparison group in 20 of these 28 pairwise comparisons. Interestingly, all physics pairwise comparisons had practically meaningful differences, whereas only two of the environmental science pairwise comparisons had meaningful differences.

3. The AP3 group outperformed the AP2 group for physics only, AP1 group across all four sciences, as well as the NoExam, DE, Honors, and Reg group for chemistry.

4. The AP2 group outperformed the AP1 groups for biology, chemistry, and physics.

Research Question 2: Science GPA

The ICCs from the intercept-only model with science GPA as the outcome was 0.087, which again is considered a sufficiently large value and necessitates the use of multilevel modeling. The results of the multilevel model that was estimated for each of the four science subject areas, with science GPA as the outcome, are shown in Table 7, and the pairwise comparisons are shown in Table 8. As with FYGPA, the pairwise comparisons revealed positive findings in favor of advanced science course taking, particularly for AP. A summary statement that is based on the expectations of the predictive model follows for each AP Exam group, focusing on which other groups that particular AP Exam group outperformed (if both criteria were not met, the pairwise comparison is not considered to be a meaningful difference and is not summarized below):

1. There were 32 pairwise comparisons for the AP5 group across the four science content areas; the AP5 group outperformed the comparison group in 29 of these 32 pairwise comparisons.

2. There were 28 pairwise comparisons for the AP4 group across the four science content areas; the AP4 group outperformed the comparison group in 20 of these 28 comparisons.

3. There were 24 pairwise comparisons for the AP3 group across the four science content areas; the AP3 group outperformed the comparison group in 14 of these 24 comparisons.

4. There were 20 pairwise comparisons for the AP2 group across the four science content areas; the AP2 group outperformed the comparison group in seven of these 20 comparisons. Interestingly, five of these practically meaningful differences were for chemistry.

In summary, for both outcome variables — FYGPA and science GPA — the pairwise comparisons reveal slightly different trends for each of the four sciences. However, one apparent underlying theme is that students in more advanced courses (e.g., AP, honors) tend to have higher FYGPAs and science GPAs compared to students in regular courses. In Table 9, we show the same information that was in Table 4 — group means for the outcomes variables — but now with all of the student-level and college-level covariates accounted for. That is, these are the predicted group means based on the multilevel models, as opposed to the actual group means that do not account for the covariates. Note that these estimated group means directly reflect the pairwise comparisons that are shown in Tables 6 and 8; the group differences between any two means shown in Table 9 are equal to the pairwise comparison differences shown in Tables 6 and 8, save rounding error. The general trends shown in Table 9 are very similar to those noted in Table 4, but as expected, the differences between groups are smaller because there are many student-level and college-level covariates in the model. The pairwise comparisons provide the detailed information within each of the four sciences when comparing the groups to the five AP score groups. Generally, the AP5 and AP4 groups were the strongest performing groups, and the AP3 group performed at least as well if not better than most groups. There were mixed results for the AP2s, AP1s, honors, and dual enrollment groups. Figures 1 and 2 visually depict these predicted group means for FYGPA and science GPA, respectively.

Discussion and Conclusions

The overall results of this study indicate that advanced and rigorous course work, such as AP, is related to better first-year college outcomes. Our two research questions, estimating a very rigorous model, show positive effects for AP participation, particularly for high scorers, above and beyond other levels of science course participation. More specifically, the results of this study reveal that *some* AP students do have higher FYGPA and science GPA than honors, dual enrollment, regular course, AP course but no exam, or no course students, when controlling for student-level and college-level covariates. However, the results vary by the specific science subject area. That is, not all AP Exam takers have higher FYGPAs and science GPAs than all other groups. Mostly, the AP4s and the AP5s have higher values on the outcome variables than the other groups, and the AP3s are performing at least comparably to the other groups. However, in some cases, the AP3s, the AP2s, and even the AP1s outperform other groups. For example, for each of the two outcomes (i.e., FYGPA and science GPA), the AP3 group outperforms the AP1 group for all four science subjects. For chemistry only, the AP3 group outperforms the honors, dual enrollment, regular, and NoExam group.

Importance of Effect Size and Practical Significance

The role of practical significance when interpreting these results is critical. In other words, one should be cautious not to put too much emphasis on the statistical significance tests of the slope coefficients from the multilevel models, or the pairwise comparison tests, given the large sample sizes for most groups of interest. There are no established guidelines regarding how large an increase in GPA is considered practically meaningful; for the current work, we used 0.05 as our benchmark. We based this on our awareness of meaningful criteria that are set by some institutions to be inducted into various honors societies or honors programs (e.g., Appalachian State University, 2013; New York University, 2013). For example, if a student had a GPA of 3.60 one semester, and 3.65 the next semester, the student is on track to graduate magna cum laude at Appalachian State University; in this scenario, an increase of 0.05 made a difference to the student. We encourage science educators, researchers, and stakeholders to consider what effect sizes are appropriate for their own purposes when interpreting the results of this study. An idea for future research to help inform this issue would be to survey students and educators and ask them what magnitude of a GPA is meaningful.

Describing and Comparing Results for Each Outcome Variable

The study focused on two first-year college outcome variables: FYGPA and science GPA. The FYGPA outcome is broader and can be composed of many different types of courses, whereas the science GPA outcome is more specific. On one hand, significant and practically meaningful results for FYGPA have implications for students' general success their first year in college. On the other hand, the science GPA outcome might be more of interest to those who are focused on understanding the correlates of success in *science* courses in college. As such, the FYGPA results can be useful when seeking to understand correlates of students' general first-year success in college, whereas the science GPA results can be useful when seeking to understand correlates of success in science courses. With the push to recruit more STEM majors in this day and age, the science GPA outcome has even more relevance. There were similarities and differences between the analyses conducted with FYGPA as the outcome and the analyses conducted with science GPA as the outcome. The science GPAs for all groups were lower than FYGPAs for all groups, which is not surprising inasmuch as science courses are considered more strictly graded than other general education courses.

For both outcome variables, the same general trends were observed; specifically, AP5s and AP4s tended to outperform other groups, the AP3s performed at least comparably to other groups, and the results were mixed for AP2s, AP1s, honors, dual enrollment, regular, and NoExam students. However, some specific differences between the two outcome variables are worth noting. For chemistry, what is most fascinating about students who scored a 3 on the AP Exam is that they outperform honors, dual enrollment, regular, and NoExam class students on science GPA more so than FYGPA. Similarly, in environmental science, the AP4s outperform the AP3s on science GPA more than FYGPA. On the contrary, with physics, the AP5s are outperforming the AP4s, AP3s, AP2s, and honors students more on FYGPA than on science GPA. Interpreting why certain magnitude differences are greater for one outcome variable over another is challenging. For example, why would students who receive an AP score of 3 on the AP Chemistry Exam outperform other groups on science GPA more than FYGPA, as opposed to students who receive a score of 5 on the AP Physics Exam and outperform other groups on FYGPA more than science GPA? Future researchers and educators should explore these patterns in more detail to determine if there are theoretically meaningful explanations for the differences in these patterns between the two outcomes of interest studied in this paper. For now, acknowledging the benefits of advanced course work, and encouraging students to participate in advanced course work, is the take-home message.

Methodology Observations for Consideration in Future Studies

Several previous research studies investigating AP have not disaggregated the five AP score categories in analyses. Rather, students who scored 3 or higher are often considered "Passers" or "AP-Credit," and students who scored 1 or 2 are often considered "Non-Passers" or "No Credit" (e.g., Keng & Dodd, 2008; Sadler & Sonnert, 2010). The results of this study highlight the importance of disaggregating AP scores in such research studies because the magnitude of the relationship between AP scores and outcome differs depending on which AP score a student receives. For example, whereas previous studies have grouped students who scored a 2 on an exam in with the Non-Credit group, the benefits of taking an AP course but only receiving a 2 on the exam are overlooked; in this study, students who received an AP2 on AP Chemistry or AP Environmental Science perform as well if not better than dual enrollment students and regular course students on FYGPA, and even more so on science GPA. Moreover, previous research has also aggregated all sciences together when examining certain outcomes, such as grade in introductory college course (e.g., Sadler & Tai, 2007b). This research also highlights the importance of disaggregating the analyses by subject area. When the five AP score categories and subject-specific outcomes are grouped together, important score-specific and subject-specific relationships can be muted, resulting in a misleading picture of the relationship between course participation and relevant outcome variables.

Limitations of Current Study

This study has several limitations that are important to consider when interpreting the results. *First,* the dual enrollment sample sizes are small compared to the other student groups, particularly for environmental science. This might be a result of the way the question is worded on the SAT Questionnaire; students might not understand how "dual enrollment" is conceptualized, which could in turn limit our sample sizes. However, the sample size is still over 1,000 for biology, chemistry, and physics, so even though this is small relative to the other group sample sizes, it is certainly still an adequate sample size in an absolute sense. Moreover, recent research has found trends similar to the ones in the current study, in respect to AP and dual enrollment performance comparisons (Godfrey et al., 2013; Wyatt & Patterson, 2013), which gives us confidence that our AP comparisons to the dual enrollment groups are accurate. We must simply use caution when interpreting any comparisons to the dual enrollment groups and recognize that additional research with better measures of dual enrollment course taking are necessary in order to contribute to the growing body of research comparing AP students to dual enrollment students. *Second*, only students who took the SAT are included in these analyses. Therefore, these results cannot be generalized to all students who take standardized tests for college admission. *Third*, only first-year outcome variables are examined in this study, limiting the interpretations to only first-year college success. Future research should examine other outcome variables, such as college retention and choice of major. *Fourth*, the AP science programs are currently undergoing a redesign project for the courses and exams, which launched with AP Biology in 2013 (College Board, 2013). As such, this research should be replicated when the new science exams have been administered in order to yield score interpretations that are valid and not outdated. *Fifth*, the sample is limited to students who took at least one science course in their first year. As such, the results cannot be generalized to all first-year college students, and future research should examine different populations of first-year college students. *Sixth*, there are several limitations that accompany the study of grades, or any sort of GPA, as outcome and/or predictor variables. Grades are indirect measures of student learning and consist of multiple components, such as student absenteeism and test scores. In turn, the meaning of the GPA is often confounded with variables that are not directly related to student learning. Moreover, teachers utilize different grading standards and grading equivalency issues (e.g., grade inflation). In summary, although GPAs are used as outcomes in this research, and in many previous research studies, a consumer of research must be aware of the limitations. *Seventh*, the AP course is a second-year course for some subjects, such as biology and chemistry. As such, all students in AP Biology and AP Chemistry either completed an honors or a regular course in that subject area first. When interpreting the effects of AP performance for these subjects, it is important to be mindful that these students have a year of experience in that subject area before entering an AP course. This will contribute to a larger magnitude of the AP effect sizes.

Implications and Conclusions

The results of this study can help students, teachers, and educators know what levels of science course work relate to positive first-year college outcomes, and in turn can assist in designing high school students' curriculum. Regardless of a student's interest or intended major, these results should encourage students who are academically prepared to add advanced science course work to their high school classes. The results are particularly favorable for AP, because these predictive models suggest that a student who receives a higher score such as a 4 or a 5 is likely to outperform students who take other levels of science course work on FYGPA or science GPA. Even if students receive a 2 or a 3 on the AP Exam, they might perform as well if not better than students who took the honors, regular, or dual enrollment counterpart course.

This study contributes to the understanding of teaching and learning of science by applying a rigorous statistical model to examine the differences between students classified in various high school science course-taking groups on their FYGPAs and first-year science GPAs. Results indicate positive relationships between advanced science course work performance and college outcomes. As such, teachers should encourage students who are academically prepared to participate in advanced high school science course work when planning their curriculum for science learning. This study is relevant to the teaching of science because it provides a thorough examination of the high school science course-taking options and relevant first-year college outcomes. Moreover, this study is relevant to the learning of science because the results provide evidence that students who acquire a depth of science knowledge and skills that enable them to earn qualifying scores on the AP Exams tend to perform better in college. These results account for the nested structure of educational data and also incorporate many covariates into the model, providing a very stringent test of these relationships.

References

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment* [Monograph]. Retrieved August 1, 2012, from U.S. Department of Education, Office of Educational Research and Improvement website: http://www2.ed.gov/pubs/Toolbox/index.html

Appalachian State University. (2013). Office of the Registrar. http://www.registrar.appstate.edu/graduation/generalinformation.html

Brody, L. E., & Benbow, C. P. (1990). Effects of high school coursework and time on SAT scores. *Journal of Educational Psychology, 82,* 866–875.

Chajewski, M., Mattern, K. D., & Shaw, E. J. (2011). Examining the role of Advanced Placement exam participation in 4-year college enrollment. *Educational Measurement: Issues and Practice 30*(4), 16–27.

College Board. (2012). 1998–2010 *Annual survey of colleges* [Data File]. New York, NY: The College Board.

College Board (2013). *Advances in AP*. Retrieved from http://advancesinap.collegeboard.org/

Common Core State Standards Initiative. (2012). Retrieved from http://www.corestandards.org/

Dalton, B., Ingels, S. J., Downing, J., & Bozick, R. (2007). *Advanced mathematics and science coursetaking in the Spring High School senior classes of 1982, 1992, and 2004* (NCES 2007-312). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Godfrey, K., Matos-Elefonte, H., Ewing, M., & Patel, P. (July 2013). *AP completion: Comparing AP, dual-enrolled, and regular students on four college outcomes.* Presented at the AP Annual Conference. Las Vegas, NV.

International Baccalaureate. *The IB Diploma Programme* (2014). Retrieved from http://www.ibo.org/diploma/

Keng. L., & Dodd, B. G. (2008). *A comparison of college performances of AP and non-AP student groups in 10 subject areas* (College Board Research Report No. 2008-7). New York: The College Board.

Morgan, R. & Klaric, J. (2007) *AP students in college: An analysis of five-year academic careers* (College Board Research Report No. 2007-4). New York: The College Board.

National Center for Education Statistics. (2013). Dual credit and exam-based courses in U.S. Public High Schools: 2010-11. Retrieved from http://nces.ed.gov/pubs2013/2013001.pdf

National Center for Improving Science Education. (1991). *The high stakes of high school science*. Washington, DC.

Next Generation Science Standards. (2013). Retrieved from http://www.nextgenscience.org/

New York University. (2013). Latin Honors. Retrieved from http://www.nyu.edu/registrar/graduation/latin-honors.html

O'Brien, D.M. & Nelson, T.D. (2004). Strengthening college preparation and access through concurrent enrollment in high school and community college. Retrieved from http://zxtm01-ti-1.utdallas.edu/research/tsp-erc/pdf/wp_obrien_2004_strengthening_college_preparation.pdf.pdf

Patterson, B. F. & Mattern, K. D. (2012). *Validity of the SAT for predicting first-year grades: 2009 SAT validity sample* (College Board Statistical Report No. 2012-2). New York: The College Board. Retrieved from http://research.collegeboard.org/sr2012-2.pdf

Patterson, B. F., & Mattern, K. D. (2011). *Validity of the SAT for predicting first-year grades: 2008 SAT validity sample* (College Board Statistical Report No. 2011-5). New York: The College Board. Retrieved from http://research.collegeboard.org/sr2011-5.pdf

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Roderick, M., Nagaoka, J., Coca, V., & Moeller, E. (2009). *From high school to the future: Making hard work pay off*. Consortium on Chicago School Research, Chicago, April.

Sadler, P. M., & Sonnert, G. (2010). High school advanced placement and success in college coursework in the sciences. In P. M. Sadler, G. Sonnert, R.H. Tai, and K. Klopfenstein (Eds.), *AP: A critical examination of the Advanced Placement Program*. Cambridge, MA: Harvard Education Press.

Sadler, P. M., & Tai, R. H. (2007a). The two high-school pillars supporting college science. *Science*, *317*, 457–458.

Sadler, P. M., & Tai, R. H. (2007b). Advanced Placement exam scores as a predictor of performance in introductory college biology, chemistry, and physics courses. *Science Educator*, *16*(2), 1–19.

Smith, J. S. (2012). *Advancing academic excellence project summary evaluation report 2002–2012*. Retrieved from http://asainstitute.org/aar/docs/AAE-Evaluation-Report-Joshua-Smith-FINAL.pdf

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Speroni, C. (2011). *High school dual enrollment programs: Are we fast-tracking students too fast?* (NCPR Working Paper). New York, NY: National Center for Postsecondary Research.

Struhl, B., & Vargas, J. (2012). *Taking college courses in high school: A strategy for college readiness*. Boston, MA: Jobs for the Future.

Trusty, J. (2002). Effects of high school course-taking and other variables on choice of science and mathematics college majors. *Journal of Counseling & Development*, *80,* 464–474.

U.S. Department of Education (2009). *American Recovery and Reinvestment Act of 2009: Using ARRA funds to drive school reform and improvement*. Retrieved from http://www.ed.gov/policy/gen/leg/recovery/guidance/uses.doc

Wyatt, J., & Patterson, B. (July 2013). *A comparison of the college outcomes of AP and dual enrollment students*. Presented at the AP Annual Conference. Las Vegas, NV.

Table 1.

Descriptive Statistics of Student Level Characteristics

Variable	*N*	*M*	*SD*	*Min*	*Max*
SAT Math	255,135	588.50	95.17	200.00	800.00
SAT Critical Reading	255,135	562.27	93.05	200.00	800.00
SAT Writing	255,119	557.06	93.71	200.00	800.00
HSGPA*	248,440	2.92	1.46	0.00	12.00
Variable	**Frequency**				
Gender — Males	121,276 (47%)				
Gender — Females	136,601 (53%)				
Racial/Ethnic Identity–Non-White	78,056 (31%)				
Racial/Ethnic Identity–White	172,670 (69%)				
Mother's Educ. Bach.+	127,292 (55%)				
Mother's Educ. Bach.-	102,190 (45%)				
Father's Educ. Bach.+	136,751 (60%)				
Father's Educ. Bach.-	89,447 (40%)				

*HSGPA was treated as continuous, and it was based on survey responses to 12 response options where 1 = A+ (97–100), 2 = A (93–96), 3 = A- (90–92), 4 = B+ (87–89), 5 = B (83–86), 6 = B- (80–82), 7 = C+ (77–79), 8 = C (73–76), 9 = C- (70–72), 10 = D+ (67–69), 11 = D (65–67), and 12 = F (65 or below).

Table 2.

Descriptive Statistics of College Level Characteristics

Variable	*N*	%
Location		
urban	107,004	41.53%
suburban	111,108	43.13%
rural	39,515	15.34%
School Type		
private	59,060	22.90%
public	198,817	77.10%
Selectivity		
under 50%	52,866	50.50%
50%–75%	170,950	66.29%
over 75%	34,061	13.21%
Size		
small	7,896	3.06%
medium	36,220	14.05%
large	48,296	18.73%
very large	165,465	64.16%
Region		
midwest	40,268	15.62%
midstate	59,357	23.02%
northeast	20,980	8.14%
south	41,262	16.00%
west	51,008	19.78%
southwest	45,002	14.45%

Table 3.

Sample Sizes of Classified Groups for Each Science Subject

	Biology	Chemistry	Physics	Environmental Science
No Class	45,402	50,247	107,392	121,489
% Female	48	49	58	51
% White	69	69	70	66
AP1	4,729	4,387	1,775	1,201
% Female	67	57	52	68
% White	53	56	51	57
AP2	4,116	3,274	2,073	1,103
% Female	63	53	42	68
% White	63	64	62	66
AP3	4,853	4,978	3,486	1,284
% Female	57	48	34	61
% White	65	64	67	67
AP4	5,255	4,274	2,342	1,829
% Female	54	40	28	52
% White	66	66	70	72
AP5	5,966	3,845	1,938	926
% Female	51	30	19	41
% White	65	67	69	77
No AP Exam	12,849	12,500	18,096	2,984
% Female	57	52	39	57
% White	66	66	67	67
Regular	118,228	114,722	80,434	114,484
% Female	53	55	53	54
% White	69	69	68	71
Honors	55,029	57,990	38,896	12,233
% Female	54	55	53	54
% White	72	71	70	79
Dual	1,450	1,660	1,445	344
% Female	58	54	51	55
% White	67	68	67	82

Table 4.

Means of Relevant Groups with No Control Variables

First-Year GPA	No Class	Regular	AP1	AP2	AP3	AP4	AP5	No AP Exam	Honors	Dual
Biology	3.000	2.883	2.822	3.041	3.154	3.270	3.501	3.000	3.113	3.050
Chemistry	2.970	2.883	2.922	3.137	3.255	3.371	3.576	3.002	3.113	3.064
Physics	2.947	2.924	2.847	3.026	3.168	3.336	3.527	3.097	3.120	3.068
Env. Sci.	2.994	2.956	2.865	3.033	3.136	3.216	3.437	3.029	3.119	3.024
Science GPA										
	No Class	**Regular**	**AP1**	**AP2**	**AP3**	**AP4**	**AP5**	**No AP Exam**	**Honors**	**Dual**
Biology	2.686	2.527	2.428	2.724	2.860	3.002	3.240	2.652	2.824	2.717
Chemistry	2.644	2.520	2.605	2.889	3.055	3.087	3.335	2.674	2.823	2.739
Physics	2.604	2.577	2.491	2.739	2.876	3.112	3.259	2.792	2.845	2.752
Env. Sci.	2.661	2.625	2.456	2.732	2.794	2.957	3.201	2.698	2.861	2.771

Table 5.

Slope Coefficient Estimates from Multilevel Model Analyses, Outcome Is FYGPA

Fixed Effect	Biology	Chemistry	Physics	Env. Sci.
Student-Level Variables				
Intercept	2.995**	2.989**	3.003**	3.007**
SAT-Verbal	0.000**	0.000**	0.000**	0.000**
SAT-Math	0.001**	0.001**	0.001**	0.001**
SAT-Writing	0.001**	0.001**	0.001**	0.001**
HSGPA	0.-135**	-0.135**	-0.136**	-0.136**
Gender	0.183**	0.183**	0.183**	0.183**
Racial/Ethnic Identity	0.086**	0.088**	0.085**	0.085**
Mother's Education	0.045**	0.045**	0.045**	0.045**
Father's Education	0.078**	0.078**	0.078**	0.078**
AP1	-0.098**	-0.059**	-0.116**	-0.036
AP2	-0.028*	-0.025*	-0.061**	-0.004
AP3	0.016	0.072**	0.002	0.027
AP4	0.053**	0.113**	0.078**	0.038*
AP5	0.161**	0.219**	0.190**	0.127**
APNone	-0.024**	-0.018*	-0.018*	-0.020
Dual	-0.011	-0.020	-0.047*	-0.029
Honors	0.025**	0.027**	0.012*	0.002
Regular	-0.002	0.001	0.009*	0.012**
College-Level Variables				
urban	-0.066	-0.067	-0.067	-0.068
suburban	-0.054	-0.053	-0.054	-0.055
college type	0.021	0.021	0.022	0.023
medium selectivity	0.155**	0.154**	0.152**	0.152**
high selectivity	0.204**	0.204**	0.202**	0.202**
large college	-0.006	-0.006	-0.007	-0.006
medium college	0.030	0.030	0.029	0.031
small college	-0.121	-0.121	-0.123	-0.123
midwest	0.064	0.063	0.066	0.070
midstate	0.081	0.081	0.083	0.088
northeast	0.121	0.122	0.124	0.129
south	0.096	0.095	0.099	0.103
west	0.018	0.018	0.017	0.020

*$p < 0.05$
**$p < 0.01$

Table 6.

Pairwise Comparisons for Biology, Chemistry, Physics, and Environmental Science for FYGPA

AP Exam Group	Comparison Group	Pairwise FYGPA Difference for:			
		Biology	Chemistry	Physics	Env. Sci.
AP5	AP4	0.108***	0.106***	0.111***	0.090***
	AP3	0.145***	0.147***	0.192***	0.099***
	AP2	0.189***	0.194***	0.251***	0.131***
	AP1	0.258***	0.279***	0.306***	0.162***
	NoExam	0.185***	0.237***	0.208***	0.146***
	DE	0.172***	0.240***	0.237***	0.156***
	Honors	0.136***	0.192***	0.178***	0.125***
	Reg	0.163***	0.218***	0.199***	0.138***
AP4	AP3	0.037**	0.041	0.080***	0.009
	AP2	0.081***	0.088***	0.140***	0.041
	AP1	0.151***	0.173***	0.194***	0.072**
	NoExam	0.077***	0.131***	0.097***	0.057**
	DE	0.064***	0.134***	0.126***	0.066
	Honors	0.028**	0.086***	0.067***	0.035*
	Reg	0.055***	0.112***	0.087***	0.048**
AP3	AP2	0.044**	0.047	0.059***	0.032
	AP1	0.114***	0.132***	0.114***	0.063*
	NoExam	0.040***	0.090***	0.016	0.047*
	DE	0.027	0.092***	0.045*	0.057
	Honors	-0.009	0.045***	0.134	0.026
	Reg	0.018	0.071***	0.007	0.039*
AP2	AP1	0.069***	0.085***	0.055**	0.031
	NoExam	-0.004	0.043***	-0.043**	0.015
	DE	-0.017	0.046*	-0.014	0.025
	Honors	-0.053***	-0.002	-0.073***	-0.006
	Reg	-0.026*	0.024*	-0.052***	0.007
AP1	NoExam	-0.073***	-0.042***	-0.098***	-0.016
	DE	-0.086***	-0.039*	-0.069**	-0.006
	Honors	-0.123 ***	-0.086***	-0.128***	-0.037
	Reg	-0.096***	-0.061***	-0.107***	-0.024

Note. **p* < 0.05; ***p* < 0.01; ****p* < 0.001. Boldface differences indicate practical significance (i.e., greater than 0.05 in absolute value). Differences are estimated as the AP Exam group minus the comparison group; positive values favor the group in the left-hand column.

Table 7.

Parameter Estimates from Model, Outcome Is Science GPA

Fixed Effect	Biology	Chemistry	Physics	Env. Sci.
Student-Level Variables				
Intercept	2.664**	2.652**	2.670**	2.677**
SAT-Verbal	0.000**	0.000**	0.000**	0.000**
SAT-Math	0.002**	0.002**	0.002**	0.002**
SAT-Writing	0.001**	0.001**	0.001**	0.001**
HSGPA	-0.155**	-0.154**	-0.155**	-0.155**
Gender	0.116**	0.116**	0.117**	0.117**
Racial/Ethnic Identity	0.140**	0.143**	0.139**	0.139**
Mother's Education	0.054**	0.054**	0.054**	0.054**
Father's Education	0.090**	0.091**	0.091**	0.091**
AP1	-0.121**	-0.032*	-0.118**	-0.093*
AP2	-0.009	0.093**	-0.051*	0.018
AP3	0.039*	0.160**	-0.024	-0.013
AP4	0.080**	0.089**	0.104**	0.060*
AP5	0.153**	0.200**	0.131**	0.129**
NoExam	-0.023*	0.000	-0.025*	-0.025*
Dual	-0.019	-0.028	-0.056*	-0.056*
Honors	0.031**	0.042**	0.033**	0.033**
Regular	-0.005	0.000	-0.009	-0.009
College-Level Variables				
urban	0.016	0.015	0.015	0.014
suburban	0.025	0.025	0.025	0.024
college type	-0.089*	-0.089*	-0.088*	-0.088*
medium selectivity	0.085*	0.085*	0.082*	0.082*
high selectivity	0.013	0.013	0.010	0.010
large college	0.021	0.022	0.020	0.021
medium college	0.077	0.078	0.077	0.077
small college	0.120	0.121	0.119	0.119
midwest	0.171*	0.170*	0.173*	0.178*
midstate	0.148*	0.149*	0.150*	0.156*
northeast	0.196*	0.198*	0.199**	0.205**
south	0.178*	0.177*	0.181*	0.186*
west	0.169*	0.169*	0.170*	0.172*

*$p < 0.05$
**$p < 0.01$

Table 8.

Pairwise Comparisons for Biology, Chemistry, Physics, and Environmental Science for Science GPA

AP Exam Group	Comparison Group	Pairwise Science GPA Difference for:			
		Biology	Chemistry	Physics	Env. Sci.
AP5	AP4	0.073***	0.111***	0.027	0.069
	AP3	0.114***	0.040	0.155***	0.143***
	AP2	0.162***	0.108***	0.181***	0.112*
	AP1	0.274***	0.234***	0.249***	0.228***
	NoExam	0.178***	0.200***	0.156***	0.159***
	DE	0.172***	0.229***	0.186***	0.081***
	Honors	0.122***	0.158***	0.098**	0.117***
	Reg	0.158***	0.200***	0.140***	0.145***
AP4	AP3	0.041*	-0.071***	0.128***	0.074*
	AP2	0.088***	-0.003	0.155***	0.043
	AP1	0.201***	0.121***	0.222***	0.154*
	NoExam	0.105***	0.089***	0.129***	0.090**
	DE	0.098**	0.118***	0.160***	0.012
	Honors	0.049***	0.047**	0.071***	0.048*
	Reg	0.085***	0.089***	0.113***	0.076**
AP3	AP2	0.048*	0.068**	0.026	0.031
	AP1	0.160***	0.193***	0.094*	0.080*
	NoExam	0.064***	0.160***	0.001	0.017*
	DE	0.058	0.189***	0.031	-0.062
	Honors	0.008	0.119***	-0.057**	-0.026
	Reg	0.044**	0.160***	-0.015	0.002*
AP2	AP1	0.112***	0.125***	0.067*	0.111
	NoExam	0.017	0.092***	-0.026	0.048
	DE	0.010	0.121***	0.005	0.030
	Honors	-0.039*	0.051*	-0.083***	0.005
	Reg	-0.004	0.092***	-0.041	0.033
AP1	NoExam	-0.095***	-0.033	-0.093***	-0.064
	DE	-0.102***	-0.004	-0.062	-0.142
	Honors	-0.151***	-0.074***	-0.150***	-0.106
	Reg	-0.116***	-0.032*	-0.108***	-0.078

Note. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Boldface differences indicate practical significance (i.e., greater than 0.05 in absolute value). Differences are estimated as the AP Exam group minus the comparison group; positive values favor the group in the left-hand column.

Table 9.

Predicted Group Means for Biology, Chemistry, Physics, and Environmental Science for FYGPA

AP Group	Biology	Chemistry	Physics	Env. Sci.
FYGPA				
AP1	2.897	2.930	2.887	2.971
AP2	2.967	3.014	2.942	3.002
AP3	3.011	3.061	3.001	3.034
AP4	3.048	3.102	3.082	3.043
AP5	3.156	3.208	3.193	3.133
No AP Exam	2.971	2.971	2.985	2.987
Honors	3.020	3.016	3.015	3.008
Dual Enrollment	2.984	2.967	2.956	2.977
Regular	2.993	2.990	2.994	2.995
None	2.995	2.989	3.003	3.007
Science GPA				
AP1	2.543	2.620	2.553	2.584
AP2	2.655	2.744	2.620	2.685
AP3	2.703	2.812	2.646	2.664
AP4	2.743	2.741	2.774	2.737
AP5	2.817	2.852	2.801	2.806
No AP Exam	2.638	2.652	2.645	2.647
Honors	2.694	2.694	2.615	2.690
Dual Enrollment	2.645	2.624	2.703	2.727
Regular	2.659	2.652	2.661	2.662
None	2.664	2.652	2.670	2.677

Figure 1.

Graphical depiction of predicted group means for the four science subjects for FYGPA.

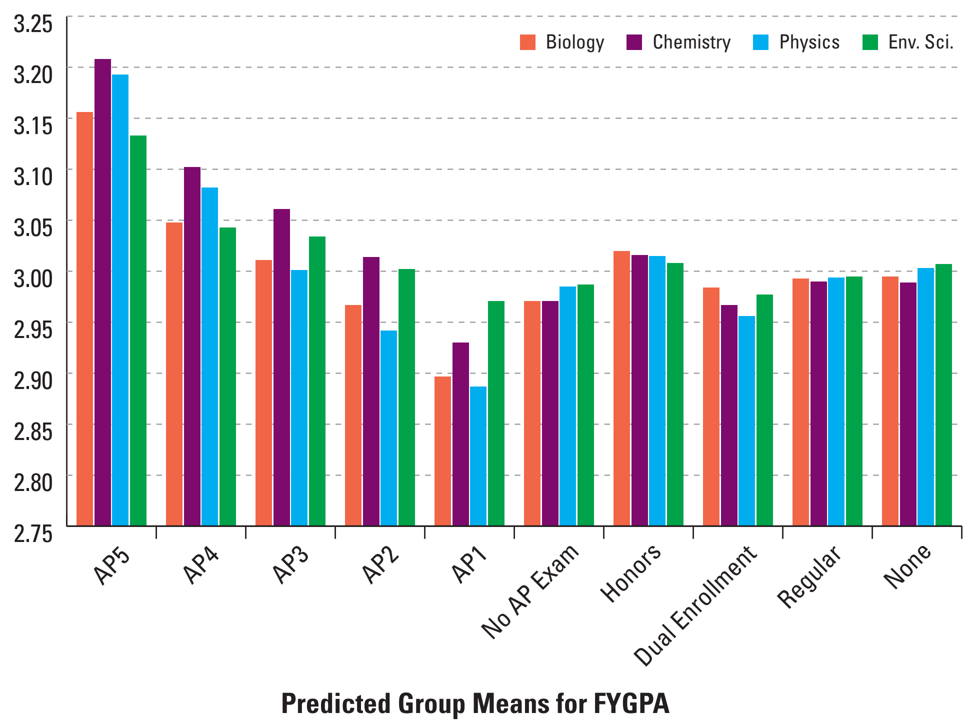
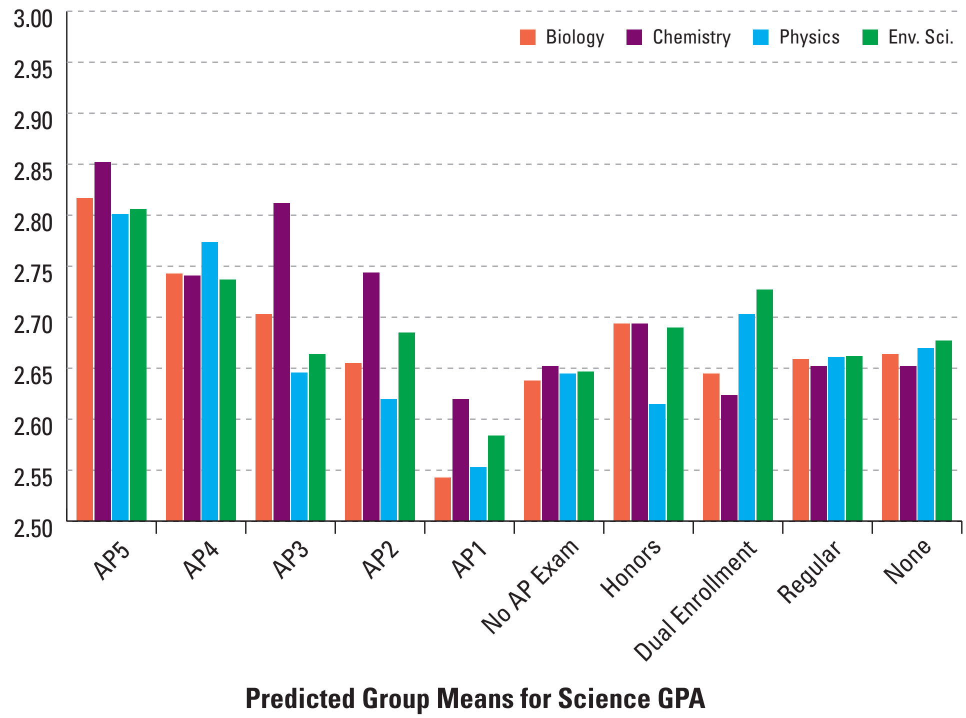


Figure 2.

Graphical depiction of predicted group means for the four science subjects for science GPA.



The Research department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: research.collegeboard.org

CollegeBoard

CollegeBoard

13b 8691