The Journal of Educational Research

ISSN: 0022-0671 (Print) 1940-0675 (Online) Journal homepage: https://www.tandfonline.com/loi/vjer20

A Closer Examination of the Academic Benefits of AP

Mary E. M. McKillip & Anita Rawls

To cite this article: Mary E. M. McKillip & Anita Rawls (2013) A Closer Examination of the Academic Benefits of AP, The Journal of Educational Research, 106:4, 305-318, DOI: 10.1080/00220671.2012.692732

To link to this article: https://doi.org/10.1080/00220671.2012.692732

Published online: 21 Mar 2013.

The Journal of Educational Research, 106:305–318, 2013

ISSN: 0022-0671 print / 1940-0675 online
DOI:10.1080/00220671.2012.692732

Routledge
Taylor & Francis Group

A Closer Examination of the Academic Benefits of AP

MARY E. M. McKILLIP
ANITA RAWLS
College Board

ABSTRACT. The authors sought to better understand the relationship between students participating in the Advanced Placement (AP) program and subsequent performance on the Scholastic Aptitude Test (SAT). Focusing on students graduating from U.S. public high schools in 2010, the authors used propensity scores to match junior year AP examinees in 3 subjects to similar students who did not take any AP exams in high school. Multilevel regression models with these matched samples demonstrate a mostly positive relationship between AP exam participation and senior year SAT performance, particularly for students who score a 3 or higher. Students who enter into the AP year with relatively lower initial achievement are predicted to perform slightly better on later SAT tests than students with similar initial achievement who do not participate in AP.

Keywords: Advanced Placement programs, assessments, high school students, multilevel modeling, propensity score matching

Researchers have found that students who have access to rigorous coursework during high school tend to have increased chances of later academic success in terms of high school assessments (Alexander & Pallas, 1984; Attewell & Dominia, 2008; Bottoms & Feagin, 2003; Gamoran & Hannigan, 2000; V. E. Lee, Croninger, & Smith, 1997), and increases in likelihood of college acceptance (Clinedinst & Hawkins, 2009), college enrollment (Clinedinst & Hawkins, 2009; Engberg & Wolniak, 2010; Long, Conger, & Iatarola, 2012), and college success (Adelman, 2006; Attewell & Dominia, 2008; Trusty, 2004). Furthermore, researchers find that students have benefitted academically from these rigorous curriculum offerings regardless of whether they were college bound (Alexander & Pallas, 1984), and regardless of race and socioeconomic backgrounds (Bottoms & Feagin, 2003). Unfortunately, students from traditionally underserved groups are less likely than others to have access to a rigorous curriculum in school (Adelman, 2006; Lee et al., 1997).

One example of a recognized rigorous secondary school program of study, as noted by the U.S. Department of Education (2010), is student participation in at least two Advanced Placement (AP) courses with a score of 3 or higher on the subsequent examinations. High school teachers design their own curriculum for AP courses, within a set of expectations established by college and university faculty for college-level courses. Only those courses with syllabi that meet or exceed these expectations are authorized to use the AP designation (College Board, 2011b). This course audit structure, combined with the fact that the AP course culminates in a national, standardized exam, produces some evidence of the rigor of AP courses. On the other hand, course rigor may vary according to teacher expectations and expertise, or the course subject matter (given that there are 37 AP courses and exams across 22 subject areas).

Past researchers have found evidence of positive benefits for AP courses and exams on later academic outcomes in high school and college (Mo, Yang, Hu, Calaway, & Nickey, 2011). However, much of this research also suggests that the positive benefits of AP are concentrated among students scoring a 3 or higher on an AP exam (Dougherty, Mellor, & Jian, 2006; Geiser & Santelices, 2004; Mattern, Shaw, & Xiong, 2009; Murphy & Dodd, 2009; Scott, Tolson, & Lee, 2010). Less positive findings are found for students who take AP courses but not AP exams, or who score a 1 or 2 on AP exams. Prior research on rigorous coursework, in general, has pointed to similar conclusions: that the benefits of these courses are heavily concentrated among students who perform well in the courses in terms of high grade point averages (Alexander & Pallas, 1984).

Additionally, some researchers have suggested that the benefits of AP are due to the fact that AP examinees are more likely than other high school students to have characteristics that make them likely candidates for later success. In other words, students who self-select into AP would perform well on later assessments and enroll and persist in college even if they did not take the AP courses and exams because they are better students: more motivated, more prepared at the start of high school, with more family advantages (Dougherty et al., 2006; Klopfenstein & Thomas, 2010). This is a particular concern, as researchers do not always carefully control

Address correspondence to Mary E. M. McKillip, Research & Development, College Board, 661 Penn St, Suite B, Newtown, PA 18940, USA. (E-mail: mmckillip@collegeboard.org)

for the selection bias whereby students who already have higher academic abilities and other advantaging characteristics are more likely to be enrolled in more challenging coursework than students with lower academic performance records.

Self-selection of higher performing students into AP is a concern in research on the impact of AP on later student performance, but current trends suggest that this selection bias may be lessening as access to AP courses and examinations are opening to more students in more high schools throughout the country now than ever before. College Board's *Annual AP Report to the Nation* for the graduating class of 2010 reports that the number of high school seniors taking an AP exam by the time of their graduation has nearly doubled since 2001, representing 28% of students from 12,705 U.S. public high schools (College Board, 2011a). With this growth, AP now reaches students who have not traditionally been found in these courses, with notable increases in the past decade in Black, Hispanic, and low-income AP examinees (College Board, 2011a).

An important policy question that is emerging as AP access expands is whether students who are lower performing at the start of high school should be allowed to participate in AP courses, or if the classes should be restricted to the highest achieving students only. While some argue that all students should be able to access these courses, others express concern about the rigor of the AP curriculum being diluted when students who are unprepared are admitted into these courses and the teacher must adjust to these lower-performing students (Matus, 2010).

Our study uses national-level data and advanced matching techniques to control for selection bias to better understand the relationship between a public high school student's participation in one of three popular AP subjects during junior year of high school and his or her Scholastic Aptitude Test (SAT) performance in the senior year. SAT assessment outcomes demonstrate a student's achievement levels at the end of high school and have been shown to be a solid predictor of first-year college grade point average (GPA; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008). Nearly 1.65 million college-bound seniors take the SAT (College Board, 2011c). We match AP examinees to similar students (i.e., similar in terms of characteristics such as parental education, race, Preliminary SAT/National Merit Scholarship Qualifying Test [PN] scores in fall of junior year) who took no AP exams during high school using propensity score matching techniques.

Our main research question was the following: What is the relationship between students who take AP during their junior year and SAT performance in their senior year when comparing similar AP to non-AP students?

Past research findings and policy questions reviewed previously suggest the need to examine this question more closely. Thus, we consider the AP and SAT relationship with four subquestions:

1. Does the AP relationship with SAT outcomes vary according to the student's AP exam score results?
2. Does the AP relationship with SAT outcomes vary according to AP subject area?
3. Does the AP relationship with SAT outcomes change once we account for other junior year AP exams?
4. Does the AP relationship with SAT outcomes vary according to the student's prior achievement level?

Method

Students who take College Board assessments are assigned a code that serves as a unique identifier and can be linked across exams to study a student's test-taking history throughout high school. These assessment data are connected to student demographic information that is collected during PN and SAT registration. For this study, we used a matched cohort file with College Board assessment history for the high school graduating class of 2010.

In order to better understand the relationship between AP participation and subsequent SAT performance, we focused on public school students who took the SAT during their senior year of high school from 2009–2010 and who had previously taken the PN in October 2008, at the start of their junior year. We excluded any students who were missing data on the variables of interest in our study. This resulted in a sample of 425,318 students. Within this sample, we focused on three groups of students, each group having taken one of three AP exams during their junior year in 2008–2009: Calculus AB, Chemistry, or English Language and Composition. These AP exams were selected for their strong correlations with the PN in Mathematics (for Calculus AB and Chemistry) and in Critical Reading and Writing (for English Language and Composition) and their popularity as junior year courses (Ewing, Camera, & Millsap, 2006; Ewing, Camera, Millsap, & Milewski, 2007). Though correlated with mathematics and popular among juniors, the AP Chemistry exam is a science course, not a mathematics course, and the AP Calculus AB is more commonly taken during a student's senior year rather than junior year.

Our research focuses on students who have taken an AP exam, under the assumption that most students who take the exam are taking it at the culmination of a year of AP coursework in the subject. Recent research on students taking exams in ten of the more popular AP subjects in 2008 found that 93% of AP examinees learned the material for the exam in an associated AP course (Kaira & Sireci, 2010). However, our data set is limited in its ability to report how many students take AP courses but do not take the associated exams. Thus, there is the possibility that we included students in our analysis who took an AP course but no exam in our non-AP comparison category. This limitation, however, would lead us to underpredict benefits of AP, if we are potentially comparing some AP course-takers to similar students who took AP courses and exams.

Given the selective sample of students who take a particular AP exam in their junior year, we created comparable groups of students using propensity score matching methods. In the absence of an experiment where researchers can carefully decide which student gets the treatment (i.e., AP) or is in the control group, propensity score matching methods are an alternative to account for selection bias in real-world situations (Rosenbaum & Rubin, 1983, 1984), and are increasingly being recommended for (Graham & Kurlaender, 2011) and used in educational research (Furstenberg & Neumark, 2007; J. C. Lee & Staff, 2007; Reardon, Cheadle, & Robinson, 2009). Each student who took a particular AP exam was matched to a comparable student selected from the 191,227 students who did not take any AP exams while in high school (but did take the PN junior year and the SAT senior year) on the following characteristics: reported GPA at the time of the PN exam; race/ethnicity; first language and best language (they could chose English only, another language only, or both English and another language); mother and father education; gender; and PN Critical Reading scores, Writing scores, and Mathematics scores. The data sets were well suited for propensity score matching, as we had a very large control group ($n = 191{,}227$) while the three treatment groups were comparatively small (Calculus AB, $n = 13{,}321$; Chemistry, $n = 15{,}503$; English Language and Composition, $n = 80{,}911$).

In our initial analysis to identify a matched sample, we entered the values for the previous characteristics into a logistic regression model and the outcome variable being whether a student took a particular AP exam. The measures for race/ethnicity, first and best language, mother and father education, and gender were treated as categorical variables, while the measures of GPA and PN subject scores were treated as continuous. From the results of this analysis, a propensity score was calculated for each student. In other words, based on the given characteristics of a student, the calculated score represents a student's likelihood to take a specific AP exam versus not take any AP exams throughout high school. This analysis was conducted separately for all three AP subjects using the following equations to estimate a student's propensity or likelihood to participate in the specified AP exam (π):

$$\begin{aligned} Logit(\pi) = \alpha &+ ß_1(Race)_1 + ß_2(Gender)_2 \\ &+ ß_3(FstLang)_3 + ß_4(BstLang)_4 \\ &+ ß_5(Father)_5 + ß_6(Mother)_6 \\ &+ ß_7(PNJRCR)_7 + ß_8(PNJRM)_8 \\ &+ ß_9(PNJRW)_9 + ß_{10}(GPA)_{10} \qquad (1) \end{aligned}$$

Students were then matched according to the estimated propensity score using a SAS (Version 9.1; SAS Institute, Cary, NC) greedy matching algorithm, with non-AP examinees matched one by one to students taking the given AP exam. With the greedy match, once a match is made, the match is not reconsidered. This method uses nearest available pair matching, starting with the best matches, then making the next best matches. We matched using a 5-to-1 digit match, meaning that students who had a propensity score in the treatment and control groups were matched if their scores differed less than 0.00001, then matches were sought with a score difference less than 0.0001 and so on, until the difference was less than 0.1 (Parsons, 2001). Among the available matching methods, this algorithm is the recommend technique for observational studies or when the outcome data are available (Stuart & Robin, 2011).

Propensity score matching techniques take into consideration the cautions of previous methodological flaws of studies of this type to reduce spurious effects (Attewell & Dominia, 2008). As can be seen by comparing characteristics of the sample of students before matching in Tables 1–3, junior year AP examinees in all subjects are notably different in characteristics as compared to students who did not take AP exams in high school. For example, on average, AP examinees have higher PN junior test scores and have parents with higher education backgrounds as compared to non-AP examinees. AP examinees are more likely to be Asian and less likely to be Black. With propensity score matching, each AP examinee was matched to one of the 191,227 non-AP students who had similar background characteristics to the AP examinees. In the event that an AP examinee was not adequately matched to a non-AP examinee or the examinee was missing data for any of the included characteristics, they were excluded from the final sample. Overall, the exclusion of examinees for the previously noted reasons resulted in a loss of 3,822 of AP Calculus AB examinees, 4,773 of AP Chemistry examinees, and 29,736 of AP English Language examinees (Tables 1–3). The AP examinees who could not be well matched to non-AP students tended to be the highest performing students prior to enrollment in AP. Despite the loss of data, our sample for analysis was still quite large and representative of the population, thus presenting no known limitations for generalizability.

Once all AP examinees were matched to a non-AP student where possible, a new data set was created with only the AP students and their comparable non-AP students. This resulted in three separate data sets, one for each AP exam of interest. Tables 1–3 show differences in characteristics of students in the samples before matching and after matching. As can be seen by comparing the means and standard deviations, or the percentage of the samples falling within a given category of a variable, the resulting samples of students were well matched. To test this further, a d statistic which represents the standardized differences between groups was calculated for the samples of students before and after propensity score matching. These differences were defined by Austin (2009) as

$$d = (x_{treatment} - x_{control})/\sqrt{((s^2_{treatment} + s^2_{control})/2)},$$

TABLE 1. Characteristics of AP Calculus AB Examinees and Nonexaminees Before and After Matching

	Unmatched sample							Matched sample						
	AP = 13,321			No AP = 191,227				AP = 9,499			No AP = 9,499			
	M	*SD*	%	*M*	*SD*	%	*d*	*M*	*SD*	%	*M*	*SD*	%	*d*
PN Critical Reading (20–80)	56.8	9.4		43.5	8.8		1.46	54.5	8.8		54.4	8.7		.01
PN Writing (20–80)	56.1	9.3		42.3	8.4		1.56	53.6	8.7		53.6	8.6		.00
PN Mathematics (20–80)	62.8	7.9		45.7	8.8		2.04	60.1	7.1		60.1	7.1		.00
GPA (0–4.33)	3.8	0.4		3.1	0.6		1.37	3.7	0.4		3.7	0.4		.00
Race														
American Indian			0.3			0.6	−.04			0.4			0.4	.00
Asian			27.0			5.8	.60			19.4			18.6	.02
Black			5.2			17.8	−.40			6.4			6.5	.00
Hispanic			12.0			13.6	−.05			13.7			14.2	−.01
White			52.1			59.5	−.15			57.1			57.2	.00
Other			3.3			2.6	.04			3.0			3.1	−.01
Gender														
Female			50.1			54.4	−.09			50.4			50.2	.00
Male			49.9			45.7	.08			49.6			49.8	.00
First language														
English only			65.6			82.2	−.38			71.2			70.9	.01
English and another			18.4			10.9	.21			15.7			15.7	.00
Another			16.0			6.9	.29			13.1			13.4	−.01
Best language														
English only			80.4			85.9	−.15			81.8			81.4	.01
English and another			17.1			12.8	.12			16.2			16.3	.00
Another			2.5			1.3	.09			2.1			2.3	−.01
Father education														
High school diploma or less			20.0			38.3	−.41			23.3			23.8	−.01
Some college			13.5			21.3	−.21			15.4			15.4	.00
Associate			5.0			7.3	−.10			5.6			5.5	.00
Bachelor's			27.0			20.1	.16			26.9			27.1	.00
Graduate degree			34.5			13.1	.52			28.8			28.3	.01
Mother education														
High school diploma or less			19.8			32.6	−.29			22.3			22.8	−.01
Some college			14.8			23.1	−.21			16.4			16.7	−.01
Associate			7.8			11.4	−.12			8.6			8.6	.00
Bachelor's			31.3			20.6	.25			29.7			29.3	.01
Graduate degree			26.4			12.3	.36			23.0			22.5	.01

Note. AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test; GPA = grade point average.

where x is the mean, and s is the standard deviation of continuous variables, or

$$d = (p_{treatment} - p_{control})/\sqrt{((p_t(1-p_t) + p_c(1-p_c))/2)},$$

where p is the percent of nominal variables.

In all cases, propensity score matching reduced the d statistics notably, and resulted in balanced samples, with AP examinees matching well on every characteristic in all three samples with d statistics under .10 (Austin, 2009).

After the match was completed, we estimated multilevel regression models using SAS PROC MIXED with maximum likelihood estimation for each AP subject with the related SAT assessment section as the outcome. There were four models estimated as a result: AP Calculus AB and SAT Mathematics, AP Chemistry and SAT Mathematics, AP English Language and SAT Writing, and AP English Language and SAT Critical Reading (Tables 4–7). In each case AP examinees in the subject were compared to similar students who took no AP exams in high school. We began with a null model for each matched AP sample with the SAT scores as outcomes to check whether to proceed with multilevel modeling (first column model labeled "Null" in

TABLE 2. Characteristics of AP Chemistry Examinees and Nonexaminees Before and After Matching

	Unmatched sample							Matched sample						
	AP = 15,503			No AP = 191,227				AP = 10,730			No AP = 10,730			
	M	*SD*	%	*M*	*SD*	%	*d*	*M*	*SD*	%	*M*	*SD*	%	*d*
PN Critical Reading (20–80)	56.9	9.4		43.5	8.8		1.47	54.1	8.5		54.1	8.8		.00
PN Writing (20–80)	56.0	9.5		42.3	8.4		1.53	52.9	8.4		52.9	8.8		.00
PN Mathematics (20–80)	62.0	8.8		45.7	8.8		1.85	58.7	7.9		58.7	7.7		.00
GPA (0–4.33)	3.8	0.4		3.1	0.6		1.37	3.7	0.4		3.7	0.4		.00
Race														
American Indian			0.4			0.6	−.03			0.5			0.5	.00
Asian			29.6			5.8	.66			19.1			18.8	.01
Black			5.7			17.8	−.38			7.2			7.4	−.01
Hispanic			9.5			13.6	−.13			11.3			11.7	−.01
White			51.3			59.5	−.17			58.7			58.3	.01
Other			3.6			2.6	.06			3.3			3.3	.00
Gender														
Female			48.6			54.4	−.12			50.1			50.0	.00
Male			51.4			45.7	.11			49.9			50.0	.00
First language														
English only			63.5			82.2	−.43			71.6			71.7	.00
English and another			20.1			10.9	.26			15.8			15.9	.00
Another			16.5			6.9	.30			12.6			12.4	.01
Best language														
English only			79.5			85.9	−.17			81.7			81.4	.01
English and another			18.1			12.8	.15			16.3			16.7	−.01
Another			2.4			1.3	.08			2.0			1.9	.01
Father education														
High school diploma or less			19.1			38.3	−.43			23.0			23.1	.00
Some college			13.3			21.3	−.21			15.9			15.9	.00
Associate			5.0			7.3	−.10			6.1			6.2	.00
Bachelor's			26.2			20.1	.15			26.7			27.0	−.01
Graduate degree			36.5			13.1	.56			28.5			27.9	.01
Mother education														
High school diploma or less			19.4			32.6	−.30			22.5			22.4	.00
Some college			14.1			23.1	−.23			16.0			16.3	−.01
Associate			7.5			11.4	−.13			8.7			8.7	.00
Bachelor's			31.4			20.6	.25			29.9			30.3	−.01
Graduate degree			27.6			12.3	.39			22.9			22.3	.01

Note. AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test; GPA = grade point average.

Tables 4–7), following Schreiber and Griffin's (2004) suggestions for presenting multilevel models. The null model is the following:

Level 1(Students): $SATScore_{ij} = \beta_{0j} + r_{ij}$,

where $SATScore_{ij}$ represents the SAT score of student i in school j, β_{0j} is the average SAT score for school j, and r_{ij} is the deviation from the mean for student i in school j.

Level 2(Schools): $\beta_{0j} = \gamma_{00} + u_{0j}$,

where γ_{00} represents the average intercept and u_{0j} is the random "school effect," school-level deviance in the mean SAT score.

For these first models, we examined the intraclass correlation coefficients (ICC), or the percentage of variability in the dependent variable (i.e., SAT score) accounted for by the presence and degree of nonindependence in the clustering of students within a school (Raudenbush & Bryk, 2002). The ICC was calculated as the quotient of the intercept variance (Level 2: Between schools) and the sum of the intercept variance and the residual variance (Level 1: Within schools) of the null model, where values above 0.01 suggest some of the variance can be explained by students within the same school (Raudenbush & Bryk, 2002). For example, in Table 4, the ICC = 1449.16 / (1449.16 + 4739.49). The ICC is 0.23 for AP Calculus AB and 0.22 for AP Chemistry with SAT Mathematics score as the outcome variable. The ICC for AP English Language is 0.22 with SAT Critical Reading as an outcome variable and also 0.22 with SAT Writing as an outcome variable. Thus, multilevel models

TABLE 3. Characteristics of AP English Language Examinees and Nonexaminees Before and After Matching

	Unmatched sample							Matched sample						
	AP = 80,911			No AP = 191,227				AP = 51,175			No AP = 51,175			
	M	*SD*	%	*M*	*SD*	%	*d*	*M*	*SD*	%	*M*	*SD*	%	*d*
PN Critical Reading (20–80)	53.9	9.2		43.5	8.8		1.16	50.1	7.9		50.1	7.9		.00
PN Writing (20–80)	53.1	9.3		42.3	8.4		1.22	49.2	7.8		49.3	7.9		−.01
PN Mathematics (20–80)	55.3	9.6		45.7	8.8		1.04	51.7	8.6		51.7	8.3		.00
GPA (0–4.33)	3.7	0.4		3.1	0.6		1.18	3.5	0.5		3.5	0.5		.00
Race														
American Indian			0.5			0.6	−.01			0.6			0.6	.00
Asian			14.7			5.8	.30			9.6			9.2	.01
Black			9.4			17.8	−.25			11.6			11.8	−.01
Hispanic			16.9			13.6	.09			17.2			17.7	−.01
White			55.1			59.5	−.09			58.0			57.6	.01
Other			3.5			2.6	.05			3.1			3.1	.00
Gender														
Female			63.7			54.4	.19			62.1			62.4	−.01
Male			36.3			45.7	−.19			37.9			37.6	.01
First language														
English only			72.8			82.2	−.23			77.1			77.0	.00
English and another			15.9			10.9	.15			13.6			13.6	.00
Another			11.3			6.9	.15			9.3			9.4	.00
Best language														
English only			82.3			85.9	−.10			83.2			83.2	.00
English and another			16.9			12.8	.12			15.9			15.9	.00
Another			0.7			1.3	−.06			0.8			0.9	−.01
Father education														
High school diploma or less			24.4			38.3	−.30			29.5			29.6	.00
Some college			16.4			21.3	−.13			18.8			19.0	−.01
Associate			5.8			7.3	−.06			6.7			6.8	.00
Bachelor's			25.7			20.1	.13			24.4			24.4	.00
Graduate degree			27.7			13.1	.37			20.6			20.3	.01
Mother education														
High school diploma or less			22.8			32.6	−.22			26.6			26.6	.00
Some college			17.5			23.1	−.14			20.0			20.3	−.01
Associate			8.6			11.4	−.09			9.8			9.8	.00
Bachelor's			28.7			20.6	.19			26.1			25.9	.00
Graduate degree			22.4			12.3	.27			17.5			17.4	.00

Note. AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test; GPA = grade point average.

were necessary to adjust for the clustering in the data. In a stepwise fashion, for each AP and SAT analysis, Step 1 included a control for PN score in the subject corresponding to the SAT outcome, Step 2 included the effect of the AP exam score, Step 3 included an additional measure for other AP exams taken junior year (the counts of AP examinees taking additional AP exams junior year are reported in Table 8) and Step 4 included the interaction between initial PN score and whether or not the student took an AP exam. The following equations represent Step 4:

Level 1(Students): $SATScore_{ij} = \beta_{0j} + \beta_{1j}(PNJRScore)_{1ij} + \beta_{2j}(AP1)_{2ij} + \beta_{3j}(AP2)_{3ij} + \beta_{4j}(AP3)_{4ij} + \beta_{5j}(AP4)_{5ij} + \beta_{6j}(AP5)_{6ij} + \beta_{7j}(APJRExtra)_{7ij} + \beta_{8j}(PNJRScore)(APTaker)_{8ij} + r_{ij}$,

where $SATScore_{ij}$ represents the SAT score of student i in school j; β_{0j} is the intercept, which is allowed to vary across schools; $\beta_{1j} - \beta_{8j}$ is the expected effect of the explanatory variable in school j; and r_{ij} is the deviation from the mean for student i in school j.

Level 2(Schools): $\beta_{0j} = \gamma_{00} + u_{0j}$,

where γ_{00} represents the average intercept of mean SAT score in school j and u_{0j} is the random school effect.

Results

Table 4, after the null model, presents the multilevel regression models for the outcome of Grade 12 SAT Mathematics scores. Step 1 includes a control for Grade 11 PN Mathematics score. In this model, a student scoring a 50 on the PN Mathematics section at the start of junior year

TABLE 4. Multilevel Regression Models of Grade 12 SAT Mathematics with Grade 11 AP Calculus AB and Grade 11 PN Mathematics (n = 18,998)

	Null		Step 1		Step 2		Step 3		Step 4	
	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*
Initial status (Intercept)	611.18***	0.80	147.26***	3.44	175.56***	3.38	176.55***	3.38	132.73***	4.39
PN Mathematics			7.75***	0.06	7.00***	0.06	6.98***	0.06	7.71***	0.07
AP score 1					8.64***	1.29	4.51**	1.45	104.38***	6.52
AP score 2					28.50***	1.40	24.09***	1.57	127.68***	6.78
AP score 3					39.25***	1.33	34.63***	1.53	140.14***	6.89
AP score 4					48.04***	1.31	43.15***	1.53	150.57***	7.01
AP score 5					65.57***	1.35	60.27***	1.60	170.50***	7.20
Other AP Jr exam							3.07***	0.49	3.48***	0.49
PN Mathematics*AP									−1.76***	0.11
−2 log likelihood	218,071.0		205,243.5		202,248.3		202,209.7		201,965.60	
Variance components										
Level 1: Within schools	4739.49***	56.12	2606.95***	30.46	2257.13***	26.15	2251.10***	26.09	2214.69***	25.74
Level 2: Between schools	1449.16***	64.22	361.91***	24.47	260.02***	19.23	261.77***	19.28	270.60***	19.50

Note. SAT = Scholastic Aptitude Test; AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test.
*p < .05. **p < .01. ***p < .001.

is predicted to score a 535 on the SAT senior year (SAT Mathematics score = 147.26 + 7.75 * 50), while a student scoring a 70 on the PN Mathematics is predicted to score a 690 (SAT Mathematics score = 147.26 + 7.75 * 70). To better understand the strength of the relationship between PN and SAT score, we measured the percent of explainable variance by the inclusion of PN Mathematics into this model through the variance component estimates (Singer, 1998). For Level 1 variance, (4739.49 − 2606.95) / 4739.49, or 45% of the within-school variance in SAT Mathematics scores is explained by the students' PN Mathematics scores, while (1449.16 − 361.91) / 1449.16 or 75% of the between school variance in SAT Mathematics score is explained by the students' PN scores. The strength of this relationship is not surprising, especially given the short time period between the PN (October of 2008) and the SAT (fall 2009 or spring 2010), and given the content similarities between the two assessment tests.

Step 2 in Table 4 considers a students' score on the AP Calculus AB exam (with the reference group of students

TABLE 5. Multilevel Regression Models of Grade 12 SAT Mathematics with Grade 11 AP Chemistry and Grade 11 PN Mathematics (n = 21,460)

	Null		Step 1		Step 2		Step 3		Step 4	
	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*
Initial status (Intercept)	599.77***	0.81	138.67***	2.89	159.92***	2.86	162.20***	2.89	126.27***	3.74
PN Mathematics			7.91***	0.05	7.27***	0.05	7.23***	0.05	7.84***	0.06
AP score 1					15.63***	0.97	9.62***	1.20	91.73***	5.55
AP score 2					35.85***	1.26	29.58***	1.45	115.87***	5.87
AP score 3					46.41***	1.16	40.02***	1.37	127.99***	5.96
AP score 4					56.33***	1.37	49.50***	1.56	139.78***	6.16
AP score 5					67.05***	1.85	59.83***	2.02	152.71***	6.45
Other AP Jr exam							4.82***	0.50	5.52***	0.50
PN Mathematics*AP									−1.48***	0.10
−2 log likelihood	249,029.8		232,100.5		229,016.8		228,923.0		228,694.9	
Variance components										
Level 1: Within schools	5439.08***	60.02	2684.66***	29.05	2357.92***	25.17	2345.47***	25.05	2315.88***	24.78
Level 2: Between schools	1558.23***	67.71	298.62***	21.15	208.25***	16.20	210.60***	16.22	215.59***	16.35

Note. SAT = Scholastic Aptitude Test; AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test.
*p < .05. **p < .01. ***p < .001.

TABLE 6. Multilevel Regression Models of Grade 12 SAT Critical Reading with Grade 11 AP English Language and Grade 11 PN Critical Reading (n = 102,350)

	Null		Step 1		Step 2		Step 3		Step 4	
	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*
Initial status (Intercept)	527.55***	0.56	118.47***	1.13	154.33***	1.21	155.50***	1.22	113.20***	1.48
PN Critical Reading			8.15***	0.02	7.25***	0.02	7.22***	0.02	8.07***	0.03
AP score 1					−14.41***	0.7	−17.06***	0.72	87.08***	2.27
AP score 2					11.36***	0.47	8.21***	0.52	121.43**	2.40
AP score 3					35.13***	0.50	31.11***	0.58	152.47***	2.57
AP score 4					52.78***	0.77	47.89***	0.84	176.62***	2.79
AP score 5					65.41***	1.43	59.51***	1.50	195.84***	3.18
Other AP Jr exam							3.96***	0.28	5.01***	0.28
PN Critical Reading*AP									−2.36***	0.05
−2 log likelihood	1,189,058.0		1,105,935.2		1,096,711.0		1,096,516.0		1,094,208.0	
Variance components										
Level 1: Within schools	5879.97***	27.21	2781.34***	12.71	2549.68***	11.68	2542.79***	11.66	2492.21***	11.41
Level 2: Between schools	1673.91***	41.60	158.48***	6.74	128.47***	6.11	132.26***	6.22	117.32***	5.73

Note. SAT = Scholastic Aptitude Test; AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test.
*$p < .05$. **$p < .01$. ***$p < .001$.

who took no AP exams in high school). Here, AP junior year Calculus AB exam taking is also a significant predictor of SAT senior year Mathematics test performance, with scores increasing for higher AP exam scores. The AP to SAT relationship varies by AP exam score. Students who take the AP exam and score a 1 are not predicted to have SAT Mathematics scores that vary notably from non-AP students—the score difference is about 9 points (8.64). However, there is a notable boost to SAT Mathematics scores for students with AP scores of 2 or higher as compared to SAT scores for similar non-AP students, ranging from 29 points (28.50) for students scoring a 2 on the AP Calculus AB up to a predicted score boost of 66 points (65.57) on the SAT Mathematics for students scoring a 5 on the AP. While the AP Calculus AB exam participation and score are not as strong at predicting differences in SAT Mathematics scores as is the PN Mathematics scores, they explain an additional 13% of the Level 1 variance and an additional 28% of the Level 2 variance in scores.

TABLE 7. Multilevel Regression Models of Grade 12 SAT Writing with Grade 11 AP English Language and Grade 11 PN Writing (n = 102,350)

	Null		Step 1		Step 2		Step 3		Step 4	
	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*
Initial status (Intercept)	514.04***	0.54	170.24***	1.24	202.26***	1.27	203.52***	1.27	164.13***	1.58
PN Writing			7.00***	0.03	6.11***	0.03	6.08***	0.03	6.89***	0.03
AP score 1					−10.54***	0.77	−13.98***	0.80	75.82***	2.37
AP score 2					17.65***	0.52	13.38***	0.58	109.73***	2.46
AP score 3					42.45***	0.56	36.83***	0.64	138.93***	2.61
AP score 4					62.06***	0.83	55.14***	0.91	162.49***	2.81
AP score 5					83.26***	1.54	74.82***	1.60	187.62***	3.22
Other AP Jr exam							5.62***	0.31	6.30***	0.31
PN Writing*AP									−2.03***	0.05
−2 log likelihood	1,183,362.0		1,123,629.0		1,112,872.0		1,112,543.0		1,110,930.0	
Variance components										
Level 1: Within schools	5556.50***	25.65	3211.99***	14.72	2893.72***	13.30	2880.80***	13.25	2843.55***	13.07
Level 2: In initial status	1609.73***	38.70	435.75***	12.68	385.36***	11.77	395.78***	11.98	364.76***	11.26

Note. SAT = Scholastic Aptitude Test; AP = Advanced Placement; PN = Preliminary SAT/National Merit Scholarship Qualifying Test.
*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 8. Frequency Distribution of Additional AP Junior Year Course-Taking Among Samples of AP Examinees

Number of additional AP Jr exam(s)	AP Calculus AB		AP Chemistry		AP English Language	
	n	%	*n*	%	*n*	%
None	1, 803	19.0	2, 538	23.7	16, 280	31.8
One	2, 839	29.9	3, 630	33.8	22, 874	44.7
Two	2, 985	31.4	3, 125	29.1	9, 337	18.3
Three	1, 451	15.3	1, 144	10.7	2, 234	4.4
Four or more	421	4.4	293	2.7	450	0.9

Note. Advanced Placement (AP) examinees in all three subjects were matched to a sample of students who took no AP exams in any subject during high school, so the comparison groups are not included in this table.

Step 3 considers whether taking additional AP exams in a students' junior year help students more on SAT outcomes than if they only take one AP exam. This variable can only explain differences within the AP group, as the non-AP group did not take any AP exams in high school. The addition of this variable, though significant, does not explain any additional proportion of the Level 1 or Level 2 variance in SAT Mathematics scores, and thus contributes very little to the model.

Step 4 includes an interaction term for PN Mathematics and AP exam taking along with all the other predictor variables included in the earlier steps. Though weak (explaining only 2% of the Level 1 variance and none of the Level 2 variance), this interaction is significant, and suggests a slight interaction such that, the higher a student's initial PN Mathematics score, the less they stand to gain on later SAT Mathematics scores from AP participation.

We present a figural representation of the model in Table 4 in Figure 1. This figure presents the predicted SAT Mathematics score outcome for three hypothetical groups of students in the AP Calculus AB matched sample: students starting junior year with a PN Mathematics score 1 standard deviation below the mean of the sample (53); students with a PN Mathematics score at the sample mean (60), and students with a PN Mathematics score 1 standard deviation above the mean (67). In each group, each bar represents the predicted SAT score for one of six cases: a student who took no AP exams in high school (the white bar), followed by five students, each of whom took the AP Calculus AB and scored a 1, 2, 3, 4, or 5.

Figure 1 represents the findings of the multilevel regression models in Table 4 as we presented them previously: The group with a PN Mathematics score of 53 will tend to have lower SAT Mathematics scores than those with a score of 60, who will tend to have lower SAT Mathematics scores than those with a PN score of 67. Within each group, the higher the AP exam score, the higher the predicted SAT Mathematic score. The interaction between PN Mathematics score and AP exam participation is shown by the slope of the line above each group: The slope is slightly steeper for the lowest PN Mathematics score group, and slightly flatter for the highest PN Mathematics score group.

The same modeling approach was used for the remaining analyses, and the results are presented in Table 5 for AP Chemistry and SAT Mathematics, in Table 6 for AP English Language and Composition and SAT Critical Reading, and in Table 7 for AP English Language and SAT Writing. In each case, Step 1 demonstrates that prior achievement is a powerful predictor of later achievement. Students with higher PN scores in the fall of junior year tend to have higher SAT scores during their senior year, and the inclusion of PN score in the model explains large proportions of the explainable variance when comparing the variance components in Step 1 to those from the null model.

In each of the remaining analyses, too, the inclusion of AP exam performance compared to the non-AP group in Step 2 explains additional variance at both Level 1 and Level 2, and higher AP exam scores predict higher SAT scores. As with the AP Calculus AB models in Table 4, however, AP exam scores are not as powerful as PN scores in predicting SAT scores in the remaining Tables 5, 6, and 7, as they explain 12%, 8%, and 10% of additional Level 1 within-schools variance, respectively. For AP Chemistry, in Table 5, there appears to be some benefit for any AP examinees compared to non-AP students, with a predicted score on the SAT Mathematics section of 16 points for students who scored a 1 on the AP exam compared to non-AP students, up to a 67 point increase for students who scored a 5 on the same AP exam. In terms of AP English Language in Tables 6 and 7, results are a little different. Students who take the AP English Language exam and score a 1 are predicted to gain less on their SAT Critical Reading (14 points less) and Writing (11 points less) than the comparable group of non-AP students. Meanwhile, students who score a 2 on the AP English Language exam are predicted to have slightly higher SAT Critical Reading (about 11 points more) and Writing scores (about 18 points more) than non-AP students, while there are substantial benefits for students scoring 3 or higher on the AP, ranging from 35 points on SAT Critical Reading for a score of 3 on the AP English Language exam, to 83 points on the SAT Writing section for a score of 5 on this same AP.

As with the AP Calculus AB sample, the consideration of additional AP junior year exams in Step 3 explains no additional variance at Level 1 or Level 2 in any of the three remaining samples. While additional AP exams were common among the AP examinees in every subject, as can be seen in Table 8, the benefit of this additional AP participation does not help to explain the increase in SAT scores.

The interaction terms between PN score and AP exam-taking added in Step 4 also does not contribute much to the models, although it does explain 11% of the Level 2 variance in SAT Critical Reading and 8% of the Level 2 variance

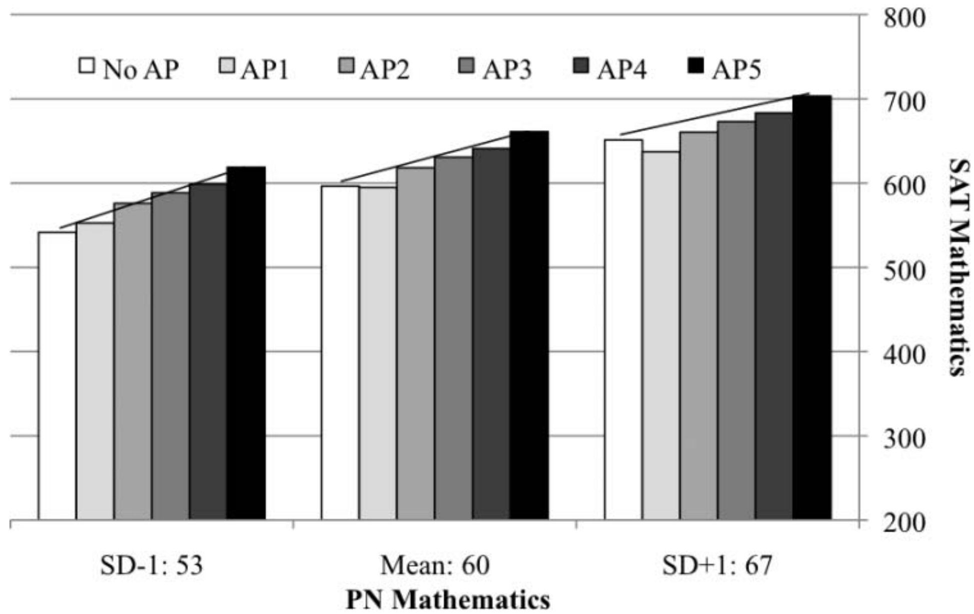

FIGURE 1. Representation of multilevel regression model estimates of AP Calculus AB and SAT Mathematics performance, based on PN Mathematics average and standard deviation of sample.

in SAT Writing scores for the AP English Language and Composition models in Table 6 and 7. The higher a student's initial PN score, the less he or she is predicted to gain from AP participation on the SAT.

Figures 2, 3, and 4 follow the same style as Figure 1, presenting the predicted SAT outcomes based on the estimates from Step 4 of Tables 5, 6, and 7, respectively, and based on the PN means and standard deviations of the matched sample. In these figures, the difference in slopes of lines is most pronounced in Figure 3, when considering the SAT outcome difference for an AP English Language student who scored a 5 on the exam and a comparable student with a similar starting PN score. In this example, a student starting with a PN Critical Reading score of 42 and scoring a 5 on the AP English Language exam (first group of bars, black bar) is predicted to score 97 points higher than a similar student who took no AP in high school (first group of bars, white bar). A student starting with a PN Critical Reading score of 58 who scores a 5 on the AP English Language exam (far right group of bars, black bar) is predicted to score 59 points higher than a similar student who took no AP in high school (far right group, white bar).

Discussion

Our findings demonstrate the importance of looking more closely at the background characteristics of students taking AP exams when studying later academic outcomes, and also considering differences by exam score and course subject matter. We do not find clear differences suggesting benefits to AP students when compared to non-AP students across subjects and AP exam scores, but rather the differences depend on the above distinctions.

In terms of our first research question, whether the AP relationship with SAT varies according to the student's AP exam score results, we find evidence that a positive benefit of AP on SAT performance increases notably with higher performance on the AP exam, particularly for students scoring 3, 4, or 5 on the AP exams. Students scoring the lowest on the exam (with a score of 1) tend to be the same or worse off than comparable non-AP students on the later SAT, most notably among students who take an AP English Language and Composition course who are predicted to score lower on the SAT assessment than their non-AP peers.

We are unable to answer why the students scoring lower on the AP exam are not benefitting from the AP coursework as their higher-scoring peers are, but we can propose two possibilities. First, students who score higher on AP exams may be more motivated than students who score a 1 or 2. This motivation may translate into students being more engaged and pushing themselves to achieve academically in their junior year of high school, preparing for the AP exam in a way that students scoring lower on AP exams have not, as past research has found links between motivation and achievement in school (Bandura, 1996; Zimmerman, Bandura, & Martinez-Pons, 1992). Second, despite our consideration for variation across schools, our results may demonstrate some additional variation happening across schools and classrooms—as some AP teachers may do a better job of preparing students for the culminating AP exam than other teachers, with many students with lower scores clustered in one classroom, and many students with higher scores

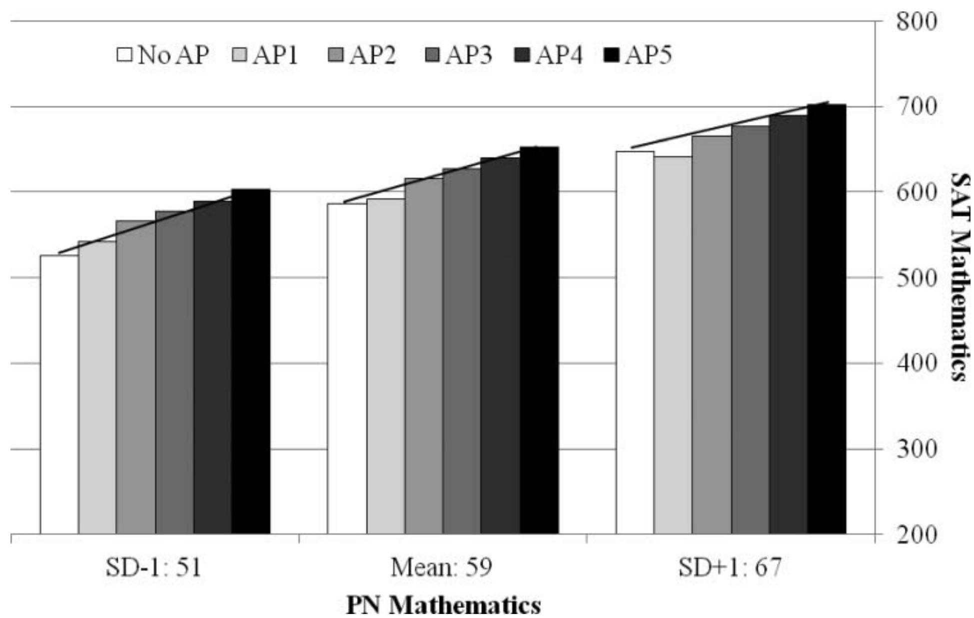

FIGURE 2. Representation of multilevel regression model estimates of AP Chemistry and SAT Mathematics performance, based on PN Mathematics average and standard deviation of sample.

clustered in another. One AP Calculus AB class may look very different from another AP Calculus AB class, and thus the outcomes in terms of exam scores are likely to be different. Though we used a multilevel model, we did not consider school-level characteristics, or allow the relationships between any of our predictor variables and the outcome to vary across schools. Future researchers would benefit from consideration of additional school-level characteristics and

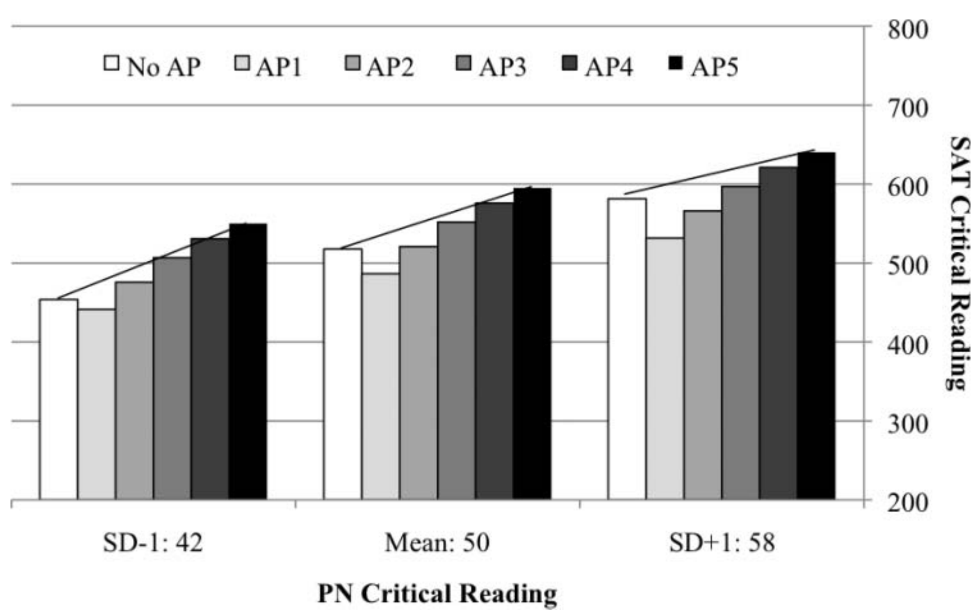

FIGURE 3. Representation of multilevel regression model estimates of AP English Language and SAT Critical Reading performance, based on PN Critical Reading average and standard deviation of sample.

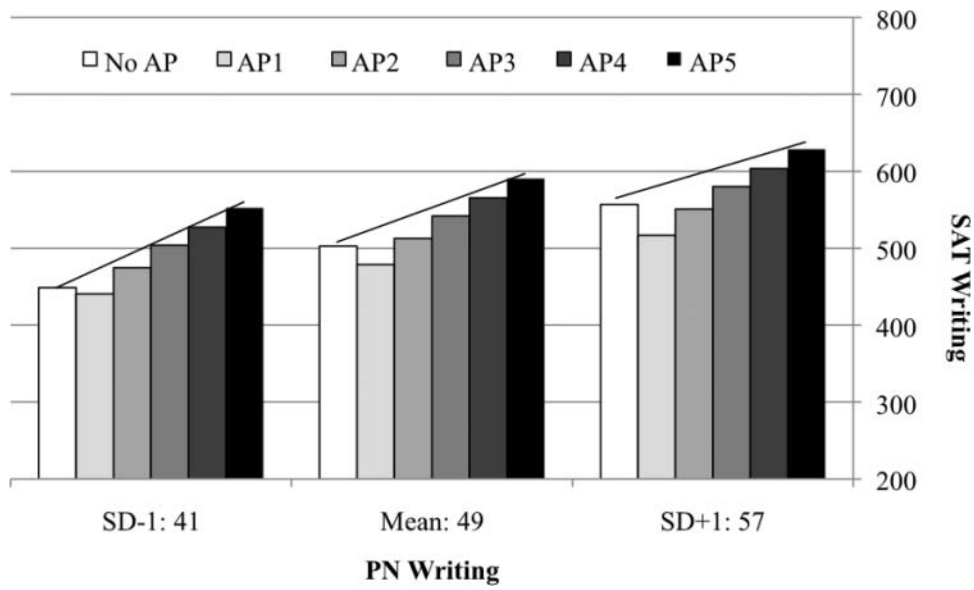

FIGURE 4. Representation of multilevel regression model estimates of AP English Language and SAT Writing performance, based on PN Writing average and standard deviation of sample.

also classroom-level characteristics that likely impact the AP exam performance of students.

In terms of the second research question as to differences in the relationship between AP and SAT depending on the subject of the course or exam, we do find that the relationships between AP participation and subsequent SAT performance tend to follow a similar pattern for all three AP subjects. Each AP course subject considered has a positive relationship with SAT outcomes, such that as AP exam scores increase SAT scores also increase. However there are some differences across the exams. For example, as noted previously, AP English Language and Composition examinees who scored a 1 tend to have lower subsequent SAT Critical Reading and Writing scores than their matched no-AP sample, a finding that is not found for AP Calculus and AP Chemistry students on their subsequent SAT Mathematics scores. This study did not examine the relationship across all 37 AP exams. Given the slight variation in results with AP English Language, it would be advantageous to examine the findings of other exams, particularly those with smaller volumes of test takers.

Our third research question considered other AP coursework and exams taken during junior year, to see whether the AP relationship with SAT was due to a general schedule of challenging AP coursework and exam taking in multiple subjects or to the particular course we were focusing on. The number of additional AP exams taken junior year added little to the model. While this finding is interesting, we cannot draw strong conclusions with it, as this measure is unable to account for other specific types of rigorous courses students took, nor can we consider the impact of AP courses among students who took no AP exams. This is a limitation of the study due to limitations in our data set.

Our final research question asked whether the relationship of AP with SAT outcomes varied according to the student's prior achievement level. The positive benefit of AP was more pronounced among students who began AP coursework with lower PN scores than the typical AP student. While the findings were not strong, they contradict arguments that only the highest performing students should be in rigorous courses such as AP.

This does not mean, however, that any low-scoring PN student should be enrolled in AP courses during his or her junior year. The results of this study may only be generalized to the population from which our sample was drawn: students taking AP exams in Calculus AB, Chemistry, or English Language and Composition and a group of students who have not taken AP exams but are comparable to the AP students. Students taking AP exams tend to have much higher starting achievement than the typical public high school junior, as can be seen in the AP and non-AP sample groups before matching in Tables 1–3. They are also more likely to be Asian or White, less likely to be Black, and more likely to have college-educated parents. Furthermore, as we noted previously, unmeasured factors such as student self-efficacy and motivation (as well as prior course history) may play a strong role in determining which lower performing students are later successful on the AP and SAT assessments. Further research is needed to help schools identify these other factors so that they can better target students for AP courses.

In addition, future research should also examine the relationship between the AP curriculum and gifted education programs. Previous research gives recommendations about how to improve the AP curriculum to meet the needs of the gifted learner (Hertberg-Davis, Callahan, & Kyburg, 2006). Some of the recommendations included increase depth of course content, increase student awareness of general learning as a benefit of AP, improve AP teacher training, increase equity, increase use of differentiated instruction, and offer other educational enrichment opportunities. The research urges AP curriculum to become more flexible to reach a diverse group of learners. We have been unable to contribute to this conversation, as our study analyses tended to exclude the highest performing pre-AP students since they could not be successfully matched to a comparable student who did not have any AP experiences in high school.

In conclusion, the results of this study suggest that students with higher AP exam scores seem to benefit the most from AP experiences, though this varies depending on the AP subject. Additionally, the findings urge schools and districts to consider other background characteristics of students beyond pre-AP test scores when selecting students for participation in AP. Future research should work to identify what additional characteristics that can be used, and how various classroom contexts may help different types of AP learners to be successful in the classroom and beyond.

ACKNOWLEDGMENT

The authors appreciate the thoughtful review of Dr. Robert Henson of an earlier version of this manuscript.

REFERENCES

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

Alexander, K. L., & Pallas, A. M. (1984). Curriculum reform and school performance: An evaluation of the "New Basics." *American Journal of Education*, *92*, 391–420.

Attewell, P., & Dominia, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, *30*, 51–71.

Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulations. *Biometrical Journal*, *51*, 171–184.

Bandura, A. (1996). Social cognitive theory of human development. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed., pp. 5513–5518). Oxford, UK: Pergamon Press.

Bottoms, G., & Feagin, C. (2003). *Improving achievement is about focus and completing the right courses*. Atlanta, GA: Southern Regional Education Board.

Clinedinst, M. E., & Hawkins, D. A. (2009). *2009 state of college admission*. Arlington, VA: National Association for College Admission Counseling.

College Board. (2011a). *7th annual AP report to the nation*. New York, NY: Author.

College Board. (2011b). *AP course audit*. New York, NY: Author. Retrieved from http://www.collegeboard.com/html/apcourseaudit/

College Board. (2011c). *College-bound seniors news release*. New York, NY: Author. Retrieved from http://media.collegeboard.com/pdf/cbs_2011_nat_release_091411.pdf

Dougherty, C., Mellor, L., & Jian, S. (2006). *The relationship between Advanced Placement and college graduation*. Austin, TX: National Center for Educational Achievement.

Engberg, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education*, *51*, 132–153.

Ewing, M., Camera, W. J., & Millsap, R. E. (2006). *The relationship between PSAT/NMSQT scores and AP examination grades: A follow-up study*. New York, NY: College Board.

Ewing, M., Camera, W. J., Millsap, R. E., & Milewski, G. (2007). *Updating AP potential expectancy tables involving PSAT/NMSQT Writing*. New York, NY: College Board.

Furstenberg, F. F., & Neumark, D. (2007). Encouraging education in an urban school district: Evidence from the Philadelphia Educational Longitudinal Study. *Education Economics*, *15*, 135–157.

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis*, *22*, 241–254.

Geiser, S., & Santelices, V. (2004). *The role of Advanced Placement and honors courses in college admissions*. Berkeley, CA: Center for Studies in Higher Education.

Graham, S. E., & Kurlaender, M. (2011). Using propensity scores in educational research: General principles and practical applications. *The Journal of Educational Research*, *104*, 340–353.

Hertberg-Davis, H., Callahan, C. M., & Kyburg, R. M. (2006). *Advanced Placement and International Baccalaureate programs: A fit for gifted students* (RM06222). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Kaira, L., & Sireci, S. (2010). *Validating Advanced Placement exams using student survey data*. Unpublished report.

Klopfenstein, K., & Thomas, M. K. (2010). Advanced Placement participation: Evaluating the policies of states and colleges. In P. M. Sadler, G. Sonnert, R. H. Tai, & K. Klopfenstein (Eds.), *AP: A critical examination of the Advanced Placement program* (pp. 167–188). Cambridge, MA: Harvard Education Press.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. (2008). *Validity of the SAT for predicting first-year college grade point average*. New York, NY: College Board.

Lee, J. C., & Staff, J. (2007). When work matters: The varying impact of work intensity on high school dropout. *Sociology of Education*, *80*, 158–178.

Lee, V. E., Croninger, R. G., & Smith, J. B. (1997). Course-taking, equity and mathematics learning: Testing the constrained curriculum hypothesis in U.S. secondary schools. *Educational Evaluation and Policy Analysis*, *19*, 99–121.

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

Mattern, K. D., Shaw, E. J., & Xiong, X. (2009). *The relationship between AP exam performance and college outcomes*. New York, NY: College Board.

Matus, R. (2010, January 7). State schools chief agrees too many in AP. *St. Petersburg Times*. Retrieved from www.tampabay.com/news/education/k12/florida-education-chief-agrees-states-ap-program-overenrolled/1063907

Mo, L., Yang, F., Hu, X., Calaway, F., & Nickey, J. (2011). ACT test performance by Advanced Placement students in Memphis city schools. *The Journal of Educational Research*, *104*, 354–359.

Murphy, D., & Dodd, B. (2009). *A comparison of college performance on matched AP and Non-AP student groups*. New York NY: College Board.

Parsons, L. S. (2011, April). *Reducing bias in a propensity score matched-pair sample using greedy matching techniques*. Paper 214–226 presented at the SAS Conference, Boston, MA.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Reardon, S., Cheadle, J., & Robinson, J. (2009). The effect of Catholic schooling on math and reading development in kindergarten through fifth grade. *Journal of Research on Educational Effectiveness*, *2*, 45–87.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*(387), 516–524.

Schreiber, J. B., & Griffin, B. W. (2004). Review of multilevel modeling and multilevel studies in The Journal of Educational Research (1992–2002). *The Journal of Educational Research*, 98, 24–33.

Scott, T. P., Tolson, H., & Lee, Y.-H. (2010). Assessment of Advanced Placement participation and University academic success in the first semester: Controlling for selected high school academic abilities. *Journal of College Admission*, *208*, 26–30. (ERIC Document Reproduction Service No. EJ893892)

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*, 323–355.

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Why matched subjects designs are superior to ANCOVA. In J. Osbourne (Ed.), *Best practices in quantitative methods* (pp. 155–176). Thousand Oaks, CA: Sage.

Trusty, J. (2004). *Effects of students' middle-school and high-school experiences on completion of the Bachelor's degree* (S. o. E. Center for School Counseling Outcome Research, Trans.). Amherst: University of Massachusetts.

U.S. Department of Education. (2010). *Academic competiveness grant*. Washington, DC: Author. Retrieved on from http://studentaid.ed.gov/PORTALSWebApp/students/english/AcademicGrants.jsp

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, *29*, 663–676.

AUTHORS NOTE

Mary E. M. McKillip is an Assistant Research Scientist at the College Board. She is a sociologist whose research focuses on identifying K–12 school characteristics that best help students, particularly those from traditionally underrepresented minority students and first-generation students, to succeed beyond high school in college and the workforce.

Anita Rawls is an Assistant Psychometrician at the College Board. Her research interests include test validity and computer-based testing.