CollegeBoard
connect to college success™

Research Report
No. 2007-5

The Relationship of AP® Teacher Practices and Student AP Exam Performance

Pamela L. Paek, Henry Braun, Catherine Trapani, Eva Ponte, and Don Powers

www.collegeboard.com

College Board Research Report No. 2007-5
ETS RR-05-10

The Relationship of AP® Teacher Practices and Student AP Exam Performance

Pamela L. Paek, Henry Braun, Catherine Trapani, Eva Ponte, and Don Powers

The College Board, New York, 2007

Pamela L. Paek is a research associate at the Charles A. Dana Center at the University of Texas at Austin.

Henry Braun is the Boisi professor of education and public policy at Boston College.

Catherine Trapani is a research project manager at Educational Testing Service (ETS).

Eva Ponte is an assistant professor at the University of Hawaii at Manoa.

Don Powers is a principal research scientist at Educational Testing Service (ETS).

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,400 schools, colleges, universities, and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com.

Additional copies of this report (item #070482345) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is $15. Please include $4 for postage and handling.

Printed in the United States of America.

Acknowledgments

The authors wish to extend thanks to Phil Handwerk for his statistical analyses, as well as his review of this report; Brent Bridgeman, Nancy Burton, Wayne Camara, Dan Eignor, and Kristen Huff for their insightful comments and reviews; Andrew Wiley, Mike Johanek, and Walt Jimenez for their interest, help, and patience in conducting this research; and the many AP teachers and AP Coordinators who agreed to participate in this study.

Contents

Tables

Figure

Abstract

This report analyzes the relationship of Advanced Placement Program® (AP®) teacher practices and student performance on AP Biology and AP U.S. History Exams. Using a national survey of AP teachers, the study developed four models for each subject with public school teachers only and both public and nonpublic school teachers, using two standards of success (scoring 3 or better and scoring 4 or better on the exams). Professional development and school and class context were statistically significant across all models; however, types of professional development differed. Resources were important for AP U.S. History teachers, while class size and schedule impacted AP Biology teachers. This indicates additional resources might enhance learning in AP U.S. History, while AP Biology teachers might be more effective with smaller, daily classes.

Introduction

The purpose of this report is to analyze the relationship of Advanced Placement Program (AP) teacher practices to student performance on AP Exams in both Biology and U.S. History and identify those practices that are associated (statistically) with higher-than-anticipated performance on the AP Exam. The reason for the AP Program is that our study focuses on a distinctive, specific type of teaching milieu: (a) the course content is unique to AP programs, in that it is more comprehensive and detailed than the usual high school course in the same subject area; and (b) students are assessed through a national high-stakes exam. These factors create a unique environment that has not usually been the focus of research on teaching.

While research has established that teachers have a measurable effect on student achievement, it has proven much more difficult to pinpoint the specific characteristics of teachers or aspects of their pedagogy that can be linked to higher student achievement (Olson, 2003). Nonetheless, this issue is now more relevant than ever; current legislation such as the No Child Left Behind Act (NCLB) requires more accountability for the achievement of our nation's students. One focus of the legislation is the preparation of a quality teaching force that will provide students with the best education possible. It carries an expectation that improvements in the professional development of teachers will promote positive changes in teaching practices, which will, in turn, enhance student achievement.

Data regarding teacher practices were obtained from national surveys of AP teachers of Biology and U.S. History. These surveys have enabled us to document the range of teaching practices used by AP teachers (Paek, Ponte, Sigel, Braun, and Powers, 2005). In this report, we look to link teacher practices and an appropriate measure of student achievement (to be described later). Because students were not randomly assigned to the teachers, the study can best be described as observational, not experimental. Therefore, no causal inferences are warranted. Rather, we are able to identify a number of associations between teacher practices and student performance, suggesting areas that may warrant further (perhaps experimental) study.

In a study such as this, choice of the performance criterion is critical. Because a student's AP score is likely related to his/her prior academic achievement, we attempted to construct a criterion from which a substantial proportion of the contribution of prior student achievement had been statistically eliminated. The particular approach we used is described in the Methodology section. The resulting statistic, a type of residual performance (or value-added measure), is an appropriate criterion because it is more likely to be sensitive to differences in teacher effectiveness than are raw scores. Consequently, the outcomes of this study can highlight areas that may warrant further study in understanding AP teacher practices. It will also add to the growing literature on effective teacher practices.

Measuring Teaching Practices

The first step in examining relationships between teacher practices and student achievement is to measure both teaching practices and the contexts in which teaching takes place. As Clare and Aschbacher (2001, p. 40) indicate, "Despite the fact that the effectiveness of school reform efforts for raising student achievement ultimately depends on the quality of instruction, we do not yet have effective and efficient ways to describe classroom practice and directly monitor the influence of reform efforts on students' learning environments."

Because of the complexity of classroom dynamics and the role of exogenous factors, researchers have developed several ways to document teaching practices and teacher-related factors that may influence those practices. These methods can be described along a continuum of more distal from the classroom to more proximal to the classroom, including surveys of teacher beliefs and practices, examination of teacher assignments, analysis of student work, student evaluation of teachers, classroom observations and videotapes of classroom observations, as well as combinations of the previous methods.

Our study employed surveys of teacher practices.

Such surveys are frequently used, as they are a cost-effective and practical means of obtaining information from a broad range of participants. However, the validity of the survey data is more problematic. Leaving aside the representativeness of the sample respondents, there are often discrepancies between what is learned from indirect methods, like surveys, and direct methods, like classroom observations. Researchers have found inconsistent relationships between surveys and other indicators of classroom performance such as logs of daily activities and classroom observations. Some researchers have found high correlations among methods (see Leighton, Mullens, Turnbull, Weiner, and Williams, 1995; Smithson and Porter, 1994); others have noted more mixed results (see Burstein et al., 1995; Chaney, 1994; Mullens et al., 1999).

However, even a high correlation between a survey and other indicators of classroom performance is not always sufficient to assure that survey results accurately capture the essential elements of teacher practice (Mayer, 1999). To address this problem, researchers have offered suggestions for improving survey data quality. Mayer argues that teachers should be asked about specific and recent periods of time in their teaching that represent typical practices (e.g., the last two weeks of instruction). Moreover, teachers should not only indicate how often they use a teaching strategy but also for how long (i.e., number of minutes). To these proposals, Mullens and his colleagues (1999) add the suggestion that survey items should have a clear and understandable definition and unmistakably defined limits, and respondents and analysts should have a uniform understanding of items. A summary of survey development is discussed later in this paper. For a detailed discussion of how we constructed the surveys employed in our study, the sampling techniques, and results of descriptive analysis, please see the first report of this project (Paek et al., 2005).

Analyzing the Effects of Practice on Student Achievement

If measuring teacher practices accurately and reliably presents a number of difficulties, then linking those practices to student outcomes can prove to be an even greater challenge. Even though many programs of professional development, induction, mentoring, and educational reform claim that their interventions benefit students, only a very small percentage have documented the impact those programs have on student learning. In her research on schoolwide reform programs, Herman (1999) found that only 3 of 24 programs had strong evidence of effects on student achievement. Similarly, in a review of research on professional development, Kennedy (1998) found only seven studies in mathematics and five in science that examined the impact of professional development on student achievement.

All parties with an interest in education have been calling for evaluation of program effectiveness. It is in this context that a new impetus for linking teaching and learning emerged in the 1990s, as evidenced by two major studies that took place in that decade. One was the Third International Mathematics and Science Study (TIMSS), which carried out a cross-country analysis of teaching practices and linked differences in practices to differences in student achievement. A second was the Tennessee Student–Teacher Achievement Ratio study (Project STAR; Bain, Lintz, and Word, 1989; Word et al., 1990), which examined the effects of class size reduction on student learning. Other relevant studies, albeit smaller in scope and influence, are the California Learning Assessment System (CLAS), Merck Institute for Science Education (MISE), and RAND Corporation studies. More recent studies were conducted by Rowan, Correnti, and Miller (2002) and Wenglinsky (2002). Next we briefly describe the findings of each study.

TIMSS aimed to "learn more about mathematics and science curricula and teaching practices associated with high levels of student achievement" (Robitaille and Garden, 1996). Using traditional correlational methods to analyze student achievement data, these researchers found that for every association between a variable (say, amount of homework) and achievement, there were counterexamples (Beatty, 1997). While TIMSS has worked to link teacher practices and student achievement, their findings show that teacher practices have a very weak relationship with classroom achievement (Williams et al., 2000).

The STAR study analyzed the effects of class size on student achievement by using traditional analysis of variance techniques. The hypothesis tested was that smaller class sizes lead to higher student achievement. The study found that students in smaller classes posted significant test-score gains across all subject areas (Finn, 2002; Krueger and Whitmore, 1999; Word et al., 1990). The study also found that these effects were significantly larger in kindergarten and first grade and that the effect of using teacher aides was not as powerful as reducing the size of the class.

A few researchers from the STAR project conducted a study to determine what effective teachers do to promote learning in reading and mathematics (Bain et al., 1989). They found that effective teachers shared the following traits: They had high expectations for student learning, provided clear and focused instruction, closely monitored student learning progress, retaught using alternative strategies when children failed to learn, used incentives and rewards to promote learning, were highly efficient in their classroom routines, set and enforced high standards for classroom behavior, and maintained excellent personal interactions with their students. This information is useful for portraying what effective teachers do in small classrooms; it is unclear, however, if these kinds of traits are facilitated by smaller classes or by particular teachers. Another drawback to this study is that teacher effectiveness was not determined by specific links between teaching practices and student achievement or statistical comparisons by teacher quality.

The CLAS was implemented in California in the early 1990s. Cohen and Hill (1998) conducted a study to examine the influence of assessment, curriculum, and professional development on teacher practice and student achievement. To that end, they carried out an analysis employing information from a survey of California elementary school teachers, student achievement data, as well as nonexperimental evidence. They found a modest relationship between teaching and learning: "Schools in which teachers report classroom practice that is more oriented to the math framework have higher average student scores on the fourth-grade 1994 CLAS, controlling for the demographic characteristics of schools" (Cohen and Hill, p. 28).

In a series of evaluations of the MISE conducted by the Consortium for Policy Research in Education (CPRE), researchers analyzed the impact of the program on student performance by applying hierarchical linear models (HLM; Consortium for Policy Research in Education, 1999, 2000, 2001, 2002). Unfortunately, they found little evidence of program impact on student achievement, and the effects that were found were rather modest. Researchers concluded that the main problem in studying changes in student performance was that available measures of science achievement were not sensitive enough to gauge the effect of inquiry-based teaching (Corcoran, 2003).

In a study conducted by RAND (Klein et al., 2000), the authors included multiple measures of student achievement in their model and utilized linear regression techniques to analyze the data. They observed small but consistently positive relationships between teachers' use of standards-based instruction and student achievement but were unable to detect any differences among achievement measures.

Rowan et al. (2002), in an analysis of the methods and results presented by Scheerens and Bosker (1997), pointed to the following problems: (a) attention is paid only to students' achievement status at a single point in time and (b) the models do not control for potentially confounding effects of other variables (e.g., socioeconomic status, prior achievement). The solutions presented by Rowan et al. are to (a) measure the amount of change per year in one classroom with the same teacher and (b) utilize measures of change adjusted for differences across classrooms in students' prior achievement, home and social background, and the social composition of the schools they attend.

Thus, these researchers decided to utilize HLM methods. Bryk and Raudenbush (1992) indicate that HLM addresses the shortcomings of traditional analysis tools (especially the lack of analytical means to address the hierarchical nature of the data) in three ways: improving the estimation of individual effects (both continuous and categorical variables can be specified to have random effects), expanding the modeling of cross-level effects (independent variables or covariates can be included in the model at different levels), and fine-tuning the partitioning of variance–covariance components (variability can be partitioned at each level). In addition, the data can be unbalanced at any level, and there is no theoretical limit to the number of levels one can include in the model. In practice, however, running a three-level model is already a complicated task.

Rowan and his colleagues found that, "After controlling for student background and prior achievement, the classrooms to which students are assigned account for somewhere between 4–18 percent of the variance in students' cumulative achievement status in a given year, which translates into a *d*-type effect of 0.21 to 0.42" (Rowan et al., 2002, p. 9). In an attempt to determine what might account for classroom-to-classroom differences in achievement, these researchers analyzed the influence of four types of variables: presage, context, product, and process.[1] Their work revealed that instructional effectiveness could be explained by differences in those variables. The effects sizes ranged from 0.10 to 0.20 for variables such as teacher experience, use of whole-class instruction, and curriculum coverage. These authors warn us that no single instructional variable should be expected to explain differences between classrooms, but rather a set of instructional effects should be combined to produce large effects.

Like Rowan et al. (2002), Wenglinsky (2002) also reviewed the literature on teacher effects and indicated three problems of traditional methods in linking teacher practices and student learning: (a) The methods are not sensitive to the multilevel nature of the data, (b) they fail to take measurement error into account, and (c) they fall short in measuring interrelationships among independent variables. Wenglinsky asserts that the best way to proceed is by using multilevel structural

[1] Presage variables are defined as teacher properties that have an influence on the interactive phase of teaching. Context variables have direct effects on instructional outcomes. Product variables are the possible outcomes of teaching, and process variables are properties of the interactive phase of instruction.

equation modeling (MSEM), which addresses each of these problems. First, MSEM includes separate models for each level of analysis, and these models are also related to one another. Second, these models address measurement error by explicitly measuring the amount of variance in the latent variables unexplained by the manifest variables and by generating latent variables from multiple manifest variables. Third, the path models estimate interrelationships among independent variables, thus making possible the estimation of indirect effects. In addition, Wenglinsky compared HLM and MSEM techniques and concluded that HLM has the advantage of being able to treat as a dependent variable not only a student outcome but also the relationship between that outcome and a student's background characteristics. The advantage of MSEM is that it makes it possible to explicitly model measurement error and more fully test relationships among independent variables.

Wenglinsky (2002) found that several aspects of teacher quality are related to student achievement when class size and socioeconomic status (SES) are taken into account. In particular, teachers' college major, professional development in using higher-order thinking skills, professional development in diversity and hands-on learning, and higher-order thinking skills are positively associated with achievement.

Another approach to dealing with the "single point in time" measurement problem indicated above was presented by Sanders and Rivers (1996). Instead of focusing on a one-year change, as Rowan et al. (2000) did, these authors analyzed the cumulative and residual effects of teachers on student academic achievement. Using data from the Tennessee Value-Added Assessment System (TVAAS), these authors analyzed the cumulative teacher effects in mathematics from grades 3 to 5. Using regression techniques, they carried out an analysis in several phases: one phase focused on shrinkage estimation for the teacher effects and grouping of teachers into "effectiveness quintiles" and encoded individual student records with the teacher effectiveness quintiles for each grade. The analysis indicated that while an effective teacher who receives students from a relatively ineffective teacher can facilitate excellent academic gain for students during the school year, the impact of relatively ineffective teachers from prior years has a significant effect on subsequent student achievement. In a later phase, the authors focused on student achievement gains based on the initial achievement level of students, and they found that lower-achieving students profited more than higher achieving students when assigned to "average teachers." The validity of this methodology has been questioned by researchers who think the methodology has a bias toward finding only supporting evidence (Haertel, 1999; Kuppermintz, 2003).

In sum, these studies indicate some of the difficulties in linking teacher practices to student achievement gains: the difficulty of representing the complexity of classroom practices and dynamics, the challenge of accurately measuring student learning, problems in collecting the data at the teacher level, and the temptation to make causal inferences from observational studies. We will address these issues later.

Survey Development

A survey method was selected as the method of choice to get a relatively large sample as needed for adequate statistical analysis. Most items in the survey were five-point rating scale items, with a few four-point scales, and one open-ended question. The survey went through three main phases before the final version was administered to the large sample in spring 2003.

The three phases include (1) an initial draft, (2) a focus group, and (3) a pilot study. In the initial draft phase, we reviewed pre-existing surveys, such as those used to gather information about the National Assessment of Educational Progress (NAEP) and about AP Summer Institutes, for possible item types. In the second phase, focus groups of AP teachers were asked to respond to each item from the initial draft and indicate which items they deemed most or least relevant. They were then asked for suggestions about other items to be included in the survey to identify key issues that may have been overlooked. The pilot study phase included a revised draft based on the feedback and analyses from the focus groups. As a result of the pilot study feedback (127 AP Biology and 97 AP U.S. History teachers returning the survey), four types of changes were introduced: (1) the scale of some items was modified to better obtain distributions of responses across all respective categories regarding the frequency of use; (2) some items were combined, (3) some items were eliminated, and (4) a few additional items were added. Specifics about these phases can be found in the report on AP teacher practices (Paek et al., 2005).

In addition, based on the concerns of Mayer (1999) and Mullens et al. (1999) in survey development, we developed items that had defined time limits, such as a specific number of hours estimated in doing a certain activity. In addition, we also designed items to be related to one another, to better understand how much emphasis was placed on a certain topic. Instead of asking how important a teacher thought a certain topic was, we asked how much emphasis was placed on a topic in comparison to other topics. The reason we designed items this way was to dissuade inflation of ratings, as a socially desirable response may result in teachers rating all topics as very important.

Final Survey Construction

In the AP Teacher Practices surveys, we aimed to create items that included several practices representative of what little is known about AP teachers' practices; while also including specific practices that we hypothesized, based on the literature review, to be more effective. Next, we provide a general discussion of the relationship we thought existed between each factor and effective teaching practices.

First, we expect that teachers' training and expertise have an effect on the quality of teachers' practices. For instance, we expect that as the teachers' participation in professional development activities increase so do the passing rates of the students of those teachers. Second, we predict that the school context shapes teachers' practices in ways that may be conducive to effective or ineffective AP practices—that is, practices that lead to high rates of passage on the AP Exam. For example, we hypothesize that schools that provide more preparation time for teachers to plan their AP classes afford teachers the opportunity to implement better-quality teaching practices. Third, we pay attention to the classroom context and how this context influences teachers' practices. In this case we postulate that the preparation of the students in the teachers' classes interacts with the instructional practices, which may be related to the final passing rates of students in the AP Exam. Fourth, and related to the school context, we conjecture that as the quantity and quality of the materials and resources provided by the school increases so do teachers' capacity to implement successful teaching practices. As an illustration, let's consider the scenario in which one teacher has access to several computers that are working properly versus a teacher who has no access to any computer. We stipulate that the teacher with access to computers is working in an environment that is more conducive to effective teaching practices.

Measuring these constructs will provide us with information about teachers' readiness to implement successful teaching practices. Our next objective is to describe how those practices take place in the classroom. To that end, we have created three main constructs embedded in a dimension we call "the factors that affect teacher practice." First, we measure the instructional and assessment practices of teachers. In this construct our intent is to depict a picture of teachers' practices. This construct will take different forms for different subject matter. In the case of U.S. History, we foresee that teachers who make extended use of in-depth essays combined with practice with multiple-choice tasks are more effective than teachers who focus only on the latter instructional form. In the case of Biology, we envision that teachers whose students independently design and conduct their own science projects will be more effective than those whose students do not have such opportunities. Second, we determine teachers' content coverage of the themes and topics of AP curriculum. We propose that the higher the alignment between the teachers' curriculum and the AP curriculum, the more effective the teachers' practices will be. Third, we aim to ascertain teachers' use of test-specific instructional activities and practices. Our view is that the stronger these test-specific practices are, the more effective the teachers' practices will be. For instance, we presuppose that teachers who conduct after-school sessions will have more students who would pass the AP Exam.

In the following sections we present each dimension, its factors, the items corresponding to each factor, and the analyses conducted for each factor (Please refer to the 2005 report by Paek et al. for full copies of the surveys).

Dimension 1: Factors Affecting Teachers' Practices

Substantive Expertise and Training

Substantive expertise and training refers to the teacher's experience with the content of the given course. This is a product of numerous factors, such as the educational background of the teacher (including educational level, major, and teaching certification), previous experience teaching courses in this subject area (AP and otherwise), and the teacher's ongoing professional development through workshops, institutes, university classes, and seminars. Presented in this light, professional development refers both to further exposure to course content as well as to experiences that bolster one's pedagogy.

School Context

School context refers to the nature of the learning environment. It measures a variety of matters related to how the school context provides or does not provide a positive setting for teaching and learning. For instance, this factor provides data for scheduling, the amount of classes and prep time that the teacher has during the day, and the amount of influence that the teacher has in organizing her/his AP class.

Classroom Context

Classroom context describes the factors that affect the composition and organization of the classroom, such as the class size.

Dimension 2: Analysis of Teachers' Practices

Instructional and Assessment Practice

Instructional and assessment practice (teachers' pedagogical practice) may be manifested through the nature of assignments (e.g., how students are configured for in-class and out-of-class activities and assignments). It also concerns the relative role of various styles of instructional delivery that teachers use in their courses. In addition to measuring the teachers' decisions about how to deliver instruction, this factor also reflects the emphasis that the teachers place on various "types" of knowledge realized through their courses. For instance, it measures the relative focus that the teachers place on different kinds of knowledge and ways to depict such knowledge, such as reciting facts and terminology, understanding key concepts from the course, and developing particular types of reasoning skills, etc. This factor also deals with the ways in which teachers assess students' understanding and provide feedback to students based on those assessments. For example, it gauges information about the type of tests used by teachers and the frequency with which they use those types of tests. Additionally, this factor addresses teachers' use of technology in the classroom. Finally, it covers issues directly related to instructional practice that do not usually take place during instruction, such as teachers' preparation time and students' homework load.

Content Coverage

Content coverage addresses the manner in which teachers cover the materials included in the AP course. The first issue addressed is how depth of course concepts is negotiated relative to breadth of course content. Second, teachers report the specific topics and themes that they find more relevant and thus tend to place more emphasis on, and rank those topics/themes regarding the degree of difficulty students have learning them. Last, content coverage refers to the extent to which the content of the AP class under examination is aligned with the content of the AP Exam. To examine such alignment, we used the information teachers gave about the emphasis they place on topics and themes and about topic difficulty, and compared that information with the number of questions per topic/theme included in the exam. We consider that the number of items corresponding to each topic is an indicator of the emphasis the curriculum places on them.

Test-Specific Instructional Activities and Practices

Test-specific instructional activities and practices refer to the instructional activities and pedagogical practices that the teacher uses specifically because he/she is teaching an AP class. This factor addresses teachers' instructional decisions, both inside and outside of class time, related to getting students ready to take and pass the AP Exam, and accounts both for activities, such as after-school review sessions, as well as pedagogical decisions, such as using AP practice tests to familiarize students with the AP Exam. It also considers the extent to which the teachers encourage or require students to participate in extracurricular activities, such as districtwide competitions, inasmuch as these activities relate to gaining knowledge about course content and preparing for the AP Exam.

Additional Data

This is a source of information not covered by the factors listed above but used in different ways to gain insight into teacher backgrounds, like teachers' age, race/ethnicity, and sex.

Sample

The initial sample, provided by ETS, included a comprehensive list of AP Coordinators at schools offering AP U.S. History or AP Biology. This list constituted the sampling frame for the first phase of the project. From this list a sample of schools was selected to achieve representative samples of U.S. schools offering AP U.S. History and AP Biology. Because AP Biology and AP U.S. History schools were selected independently, some overlap existed, meaning a small proportion of AP Coordinators contacted were asked to provide teacher contact information for both AP Biology and AP U.S. History teachers. AP Coordinators were asked to return a form, providing contact information for all teachers of the subject at the school. The teacher information gathered in this process formed the population for the second phase of the project wherein a sample of teachers was asked to complete a questionnaire. This sample was selected from schools with AP Coordinators that responded to the first phase of the project. This two-step sampling process—school selection followed by teacher selection within these schools—offered a representative sample of teachers in both AP Biology and AP U.S. History by developing a list of AP teachers in the United States where none was available.

A nationally representative random sample of 3,484 AP Coordinators was surveyed to obtain contact information for all AP teachers in both U.S. History and Biology at their respective schools. The samples were designed to give full representation to public and nonpublic[2] high schools in various regions of the nation, and to high schools in different size categories (number of students). Biology and U.S. History were the two chosen subjects because both have relatively large volumes and can be seen as representative of the science and humanities offerings of the AP Program. AP

[2] Schools self-designate themselves as either parochial or private, seemingly then differentiating religious and secular schools; however, some religious schools call themselves private and not parochial, which leads to confusion when trying to compare parochial and private school teachers. Therefore, the category nonpublic was introduced.

Biology has a very broad and dynamic curriculum that poses many pedagogical challenges while U.S. History has a well-established curriculum and an innovative assessment structure.

The purpose of the AP Coordinator survey was to gather contact information for the target respondents of the survey: all teachers who had taught AP Biology or U.S. History during the current school year or the two years prior. The survey consisted of an advance e-mail, a mailed packet (cover letter, request for contact information, and a return envelope), a reminder postcard, an e-mail reminder (supplemented with a telephone call for AP Coordinators missing e-mail), and a second mailed packet for nonrespondents. This phase of the research was conducted in fall 2001, and the sample was then refreshed in fall 2002 to include schools that offered the AP course for the first time in the 2002-03 academic year.

Based on feedback from teachers, survey experts, teacher-education researchers, and information from analyses of the pilot study data, a final instrument was created and delivered to 1,874 teachers of AP Biology and 2,336 teachers of AP U.S. History in spring 2003.

The study population was defined as teachers of AP Biology and U.S. History at high schools in the United States. Teachers were eligible if they had taught either subject at any time from spring 1999 to spring 2003. Of those, 1,171 AP Biology and 1,219 AP U.S. History teachers responded (a 62 percent return rate for AP Biology and 52 percent return rate for AP U.S. History). Most respondents were public school teachers: 912 or 78 percent for Biology and 932 or 76 percent for U.S. History. These numbers are representative of the surveys mailed out, as 76.6 percent of surveys were mailed to public school teachers, while about 21 percent of mailed surveys were targeted at nonpublic school teachers.[3] A similar percent of the nonpublic-targeted surveys were returned: 21.2 percent for Biology and 20.3 percent for U.S. History. The remaining teachers were classified as "missing" in that their school-type was not yet on file with the College Board (for schools administering AP for the first time). Almost 2 percent of schools were administering AP courses for the first time in the mailed sample; more AP U.S. History teachers from these new schools responded than AP Biology teachers (3 percent and 0.8 percent, respectively). In addition, the sample was representative by both region and school size.

Not all of the 1,219 questionnaires returned by AP U.S. History teachers and 1,171 questionnaires returned by AP Biology teachers were used in the analyses linking survey results with student test performance. Because there were no unique identifiers for teachers in schools with more than one AP teacher per subject, there were fewer questionnaires available for this study. Obviously, for schools in which there was only one AP U.S. History or one AP Biology teacher, the link was easily established. To include larger schools in this study, we asked for help from AP Coordinators at the schools. After numerous attempts to contact these AP Coordinators, we recovered an additional 76 U.S. History teachers and 12 Biology teachers. The data from the other schools were not used in the analyses in this report. In all, 920 U.S. History surveys and 909 Biology surveys could be included in the analyses linking teacher practices with student achievement data. The resulting school sample was not representative of all the schools surveyed, because schools with one teacher per AP subject (usually smaller schools) were overrepresented. Finally, after some exploratory analyses, we decided to eliminate classes with fewer than eight students with usable data. (This was done mainly to assure reasonable stability of class averages.) Dropped classes included those that originally had fewer than eight students taking the AP Exam, as well as those classes that resulted with eight or fewer students without a PSAT/NMSQT® score. Student AP Exam scores for both subjects were pulled from the spring 2003 administration. PSAT/NMSQT scores for these students were matched from previous years, from fall 2002 and previous administrations; as such, all students in this study took the PSAT/NMSQT prior to the AP Exams. Our final sample included 566 public school teachers and 172 nonpublic school teachers for AP U.S. History and 473 public school teachers and 194 nonpublic school teachers for AP Biology.

Methodology

As alluded to earlier, since higher AP performance is to be expected among those students with well-developed academic ability, it did not seem reasonable to use raw AP scores (reported on a 1–5 scale) as a criterion. We decided to construct a new criterion from which prior student achievement (as indicated by PSAT/NMSQT scores) was removed, at least approximately. Camara and Millsap (1998) found a fairly strong relationship between PSAT/NMSQT scores and AP performance (correlations ranging from 0.5 to 0.6 for most subjects), which provided the rationale for using PSAT/NMSQT as a workable proxy for prior student achievement in this context. In addition, the College Board refers to the PSAT/NMSQT as a reasoning test designed to measure critical thinking and reasoning skills, which validates our categorization of PSAT/NMSQT as a measure of prior achievement.

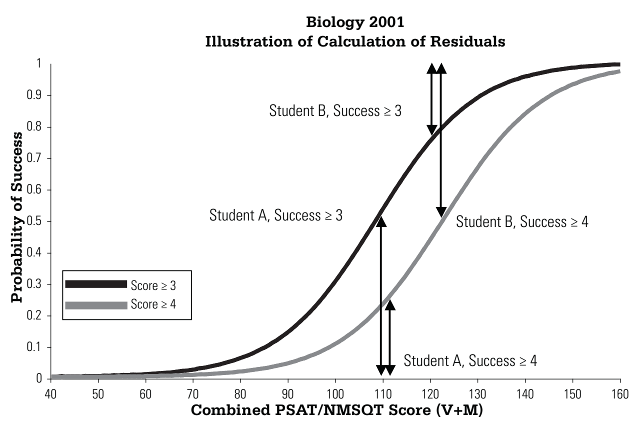
Rather than working with the exact AP score, we decided to dichotomize the score scale. In one case, we

[3] We originally planned to analyze public, private, and parochial schools. This was under the assumption that private schools were secular and parochial schools were religious. However, parochial schools can be classified as private, so we decided to combine private and parochial schools into nonpublic schools.

coded scores of 3 and above as "1" and 2 and below as "0." In the second case, we coded scores of 4 and above as "1" and 3 and below as "0." We treat a "1" as a success and "0" as a failure. The rationale for employing two criteria is that, while traditionally scores of 3 or better were accepted for college credit, many colleges are now requiring scores of 4 or better. In what follows, we ran two parallel analyses, using these different definitions of success.

First, we ran a logistic regression at the individual student level, predicting success (as defined above) from that individual's PSAT/NMSQT scores. For AP Biology, the sum of the PSAT/NMSQT Verbal and PSAT/NMSQT Math scores were used; for AP U.S. History, only PSAT/NMSQT Verbal was used as a predictor. With the fitted logistic regression, we were able to compute an estimated probability of success for each student. Next we created an individual residual by subtracting the probability of success from the (coded) score of 1 or 0 (See Figure 1). This residual is simply the difference between the actual (coded) score and its expected value based on the fitted logistic regression.

This residual is analogous to (though clearly not equivalent to) a gain score for the student. That is, we treat the PSAT/NMSQT score as a proxy for prior achievement as well as unidentified confounding factors (Camara and Schmidt, 1999; Zwick, 2001). Although performance on the PSAT/NMSQT is admittedly a "noisy" indicator of student achievement, it should be adequate for this study; that is, we expect that the residual, in comparison to the raw score, will be a more sensitive criterion for ferreting out possible contributions of teachers to student achievement. Since our analyses are conducted at the teacher level, we constructed a mean class residual (MCR) by taking the average of the residuals for all the students in the class. Recall that only classes with eight or more students with usable data were included in this study. In what follows, we will denote the criterion as MCR3 (AP Exam scores ≥ 3) or MCR4 (AP Exam scores ≥ 4) when we want to highlight which definition of "success" is being employed in its creation. Use of the MCR as a criterion places a burden on subsequent analyses inasmuch as much of the systematic variation has been removed through the preliminary logistic regression. The MCR contains a lot of noise even with averaging over students in a class; this is primarily due to the fact that the dichotomous outcome variable does not have that much information to begin with, but also because we are working with a large number of correlated variables, and what may emerge as significant may be due to the relationships among the predictors.



Note: Student A had combined PSAT/NMSQT scores of 110 and received a 2 on the AP Exam, a "failure." The residual is 0 – 0.2259 = –0.2259 for success ≥ 4, and 0 – 0.5245 = –0.5245 for success ≥ 3. Student B had combined PSAT/NMSQT scores of 120 and received a 4 on the AP Exam, a "success." Student B's residual for success ≥ 3 is 1 – 0.7431 = 0.2569, and for success ≥ 4 is 1 – 0.4278 = 0.5722.

Figure 1. Probability of success on the AP Biology Exam as a function of PSAT/NMSQT score.

In our literature review of possible analysis techniques of measuring the relationship of teachers with students, we discussed the outcomes of different studies in addition to advice from researchers on ways to analyze data. As such, we employed as much advice as we could in our analyses, one of which was to control for other confounding variables, like prior achievement. We also utilized variables from previous studies that have been shown to relate to student achievement, like teacher background variables (professional development, college major). Researchers in this field also recommended looking at multiple years of data. Since this study is the first in a possible series researching this relationship, our analyses are limited to the data we have, which is a single year of AP Exam data mapped to teachers' accounts of their practices. Further research in the cumulative and residual effects of teacher practices could be analyzed when additional years of testing for a single teacher taking this survey have been compiled.

Out of the possible analyses discussed, HLM seemed to be more powerful than the traditional analyses used by some of the other studies. However, we decided not to study the hierarchical nature of the data in this initial set of analyses. Since we are interested in school effects in addition to teacher effects on student achievement, we know the complications and difficulties in successfully running a three-level HLM. The other issue in the three-level HLM model was the sparseness of data, since we had very few schools with several teachers nested in a school. Although we could have run a two-level HLM, we felt more comfortable trying to build a value-added model with a more simple and straightforward set of analyses, as we felt this study was a preliminary one in better understanding AP teacher practices and the possible relationship with one year of student AP Exam data.

Analyses were run for all teachers and then repeated for public school teachers only. Originally, we intended to conduct separate analyses for private school teachers and for parochial school teachers. However, we discovered that the school type variable was based on school self-reports and that some religious schools chose to call themselves parochial while others chose private. Since we were unable to clearly distinguish private schools from parochial schools, we chose to collapse these categories into one (nonpublic schools). In addition, due to the small number of nonpublic school teachers, we combined our analyses of public and nonpublic school teachers, with a test for difference in school type.

Creation of Regression Models

We carried out a systematic series of exploratory unweighted regression analyses using the general linear models (GLM) methodology. The independent variables were organized into groups of AP teacher practices that were the basis for the AP Teacher Study Survey. Because of the presumed importance of the overall quality of a school, all regressions included the five-year mean PSAT/NMSQT score for the school in which the class was located.[4] This variable included all students who took a PSAT/NMSQT at a given school, regardless of whether they participated in the AP program.[5]

In grouping variables that addressed similar aspects of teaching practice, we relied on the theoretical framework we had created to guide the construction and analysis of the survey (Paek et al., 2005). Every regression was run with the mean school PSAT/NMSQT entered first. Data from each group of substantively related questions was then entered on its own. Individual questions were included in the analysis as separate variables and were treated as categorical, thus creating implicit dummy variables to represent the different response levels.

For example, the first question on the U.S. History Teacher Survey related to teaching objectives. This was a multipart question that asked how much emphasis was placed on each of six possible objectives. The responses were on a five-point Likert scale ranging from "Less than Average Emphasis" to "More than Average Emphasis." These six objectives were entered into the regression as a group of variables, after the mean school-level PSAT/NMSQT score. Using this procedure, the preceding class levels and analysis of variance (ANOVA) table was created using MCR4 as the dependent variable. See Tables 1 and 2.

At each stage, the significant effects were identified using the partial sums of squares (Type III[6]), generally using as a criterion a *p*-value of 0.15 for inclusion in subsequent models. This moderately "liberal" *p*-value was chosen because of possible collinearity among the variables. We did not want to eliminate anything too rashly. In this case, we retained variables on learning facts, skills for stating and supporting claims, and developing historical research skills for the next stage of analysis. Least squares means were used to compare effects associated with different response levels for a particular question. For example, the variable for stating and supporting claims had an MCR of −0.103 for the response "slightly less than average emphasis" compared to an MCR of −0.017 for the response "more than average emphasis." This comparison was significant with a *p*-value of 0.0112. The variables at our disposal fell naturally into two general categories: context for teaching

Table 1

AP U.S. History Learning Objectives: Levels and Values

Learning Objectives	*Levels*	*Values*
Emphasis on learning facts, dates, events, and terminology	5	1 2 3 4 5
Emphasis on understanding themes	5	1 2 3 4 5
Emphasis on viewing history as multiple perspectives/stories	5	1 2 3 4 5
Emphasis on developing skills for stating and supporting claims	4	2 3 4 5
Emphasis on developing historical research skills and techniques	5	1 2 3 4 5
Emphasis on developing interest in U.S. History	5	1 2 3 4 5

Note: Although the survey had five values for "emphasis on developing skills for stating and supporting claims," no teacher chose a "1," which is why there are only four levels for that learning objective.

[4] Initial descriptive analyses revealed a strong association between MCR and school mean PSAT/NMSQT score.

[5] One drawback to using the mean PSAT/NMSQT at the school level is that it is not uniformly administered across states. For instance, some places require all students to take the PSAT/NMSQT, not just college-bound students. These scores will be lower overall than in a district where the test is optional.

[6] This method calculates the sums of squares of an effect in the design as the sums of squares adjusted for any other effects that do not contain the design and orthogonal to any effects (if any) that contain it. Because Type III sums of squares are invariant with respect to the cell frequencies, as long as the general estimate remains constant, they are often considered useful for an unbalanced model with no missing cells, as is the case in this study.

Table 2

Type III: AP U.S. History Public School Teachers, MCR4

Source	*DF*	*Type III SS*	F	p
Mean PSAT/NMSQT score	1	1.4557	47.61	<0.001
Emphasis on learning facts, dates, events, and terminology	4	0.3773	3.08	0.0156
Emphasis on understanding themes	4	0.1972	1.61	0.1694
Emphasis on viewing history as multiple perspectives/stories	4	0.0352	0.29	0.8860
Emphasis on developing skills for stating and supporting claims	3	0.3568	3.89	0.0090
Emphasis on developing historical research skills and techniques	4	0.2438	1.99	0.0939
Emphasis on developing interest in U.S. History	4	0.1466	1.20	0.3102

and teacher practice. The context for teaching variables comprised five groups:

1. Teacher variables related to professional experience (five individual variables, including years of teaching experience, years of AP teaching experience, educational level, major, and teaching certificate);
2. Professional development (17 variables dealing with participation in specific AP professional development activities [PAPPD]);
3. School support (six variables, including school policy for assigning teachers to AP classes, number of classes taught, number of students per class, type of teaching schedule, adequacy of different school resources, and influence of AP resources);
4. Classroom control (seven variables, including hours dedicated to prepare for AP class, teaching autonomy, school criteria for AP enrollment, school policies determining which students take the AP Exam, percentage of students who take the AP Exam, teaching freedom, and class size);
5. Hours of preparation and number of classes taught (two variables and their interaction).

The significant variables ($p < 0.15$) from each group were then entered together in a regression model that included mean school PSAT/NMSQT score as a covariate. Those variables that again achieved significance ($p < 0.15$) were retained to form a parsimonious context for teaching model.

A similar sequence was carried out for the teacher practice variables. These variables comprised six groups:

1. Learning goals (six variables);
2. Instructional methods (eight variables);
3. Assessments (five variables);
4. Feedback (six variables);
5. Student activities/tasks (eight variables); and
6. AP Exam preparation techniques (seven variables, including focus on multiple-choice, free-response, or both portions of the test; percent of class time dedicated to prepare for the AP Exam during the year; percent of class time dedicated to prepare for the AP Exam during the month prior to the exam; and type and frequency of review activities implemented to prepare for AP Exam).

All the significant teacher practice variables were entered in a new regression with mean school PSAT/NMSQT score as a covariate. Those that attained significance were retained in a parsimonious teacher practices model.

To construct the final model,[7] the mean school PSAT/NMSQT score was entered first since it has a strong statistical relationship to the criterion that needs to be taken into account. Next, the parsimonious teacher context variables were entered into the model to account for contextual differences among teachers. Lastly, the parsimonious teacher practices variables were entered. After accounting for individual prior student achievement (by means of the preliminary logistic regression), mean school PSAT/NMSQT score, and other context variables, we are able to estimate the incremental contribution of teacher practices to AP Exam results, either MCR3 or MCR4. This is a very stringent standard for identifying teacher practices that may be related to improved student achievement.

To compare teachers with different responses to an item, we computed the corresponding least squares means (LSM) of the MCRs, which control for all other terms in the model by accounting for the effects of correlated variables. We compared responses within an item (e.g., excellent, good, fair, and poor are response choices within one item) using a significance level of $p < 0.01$ to reduce capitalization on chance and

[7] Variables were included in the final model if their p-value was ≤ 0.10.

thus the possibility of finding a false positive result. This increased the chances that significant differences represent real differences.

In addition to identifying statistically significant variables, we also computed effect sizes for each variable in the final eight models using the partial-η squared (η_p^2). Because of our concern in potentially overestimating the effect size, we used Kennedy's (1970) method for computing partial η-squared, which is the sum of squares of the variable over the total sum of squares,

$$\eta_p^2 = \frac{SS_{effect}}{SS_{total}}$$

We used Kennedy's formula rather than Cohen's (1969) and Friedman's (1968) suggestion of using the sum of squares of the variable divided by the sum of the sum of squares of the variable plus the sum of squares of the error,

$$\eta_p^2 = \frac{SS_{effect}}{(SS_{effect} + SS_{error})}$$

because the latter could inflate the effect sizes because it was not accounting for all other sources of variance (Halderson and Glasnapp, 1972; Wuensch, 2006). The partial η-squared values can be found in the appendixes next to the *p*-value for each variable in each model. Effect sizes for each variable were generally small, around 0.02, which individually does not seem to have a large effect, but additive-wise, the η-squared for the models tend to hold moderate-sized effects (averaging around 0.30), since the η-squared values are similar to the *R*-squared values.

Summary of Results

The results are listed in eight separate analyses. The first four analyses are for AP Biology: (a) for public school teachers with the criterion MCR3, (b) for public school teachers with the criterion MCR4, (c) for public and nonpublic school teachers combined with the criterion MCR3, and (d) public and nonpublic school teachers combined with the criterion MCR4. A parallel set of four analyses follows for AP U.S. History.

Table 3 summarizes the *R*-square and root mean square error (RMSE) for the mean PSAT/NMSQT score and then the mean PSAT/NMSQT score + Context + Teacher Practices.[8] We did not try to maximize the adjusted *R*-square by combining response categories for individual questions to reduce the number of dummy variables—and degrees of freedom—needed. Rather, the idea was to carry out a systematic but exploratory analysis by looking specifically for variables that showed differences in student performance. We employed the baseline *R*-square, using only the mean PSAT/NMSQT score to get an idea of how much the context variables and the teacher practice profile contribute.

In addition to the *R*-square, which is not adjusted for the number of predictors, we also employed the RMSE of the fitted model as a diagnostic. The RMSE is a kind of generalized standard deviation and an excellent scalar measure of predictive efficacy of the model.

We were interested in how much more of the criterion variance we could account for with context and teacher practice variables, as well as how much we could reduce the RMSE error. For AP Biology public school teachers, the initial regression using only mean school PSAT/NMSQT score as a predictor of the MCR3 criterion

Table 3

R-Square and RMSE of Models

Model	*R^2 (RMSE) of mean PSAT/NMSQT score only*	*R^2 (RMSE) of mean PSAT/NMSQT score and context and teacher practices*
Biology public schools only (MCR3)	0.06 (0.182)	0.27 (0.167)
Biology public schools only (MCR4)	0.06 (0.179)	0.33 (0.158)
Biology combined public and nonpublic schools (MCR3)	0.09 (0.182)	0.20 (0.175)
Biology combined public and nonpublic schools (MCR4)	0.05 (0.187)	0.40 (0.154)
U.S. History public schools only (MCR3)	0.09 (0.189)	0.32 (0.169)
U.S. History public schools only (MCR4)	0.06 (0.171)	0.26 (0.159)
U.S. History combined public and nonpublic schools (MCR3)	0.10 (0.188)	0.16 (0.185)
U.S. History combined public and nonpublic schools (MCR4)	0.08 (0.178)	0.34 (0.160)

[8] Although we ran models with the mean PSAT/NMSQT score and context for teaching variables, we did not include the *R*-square or RMSE in Table 3 because we do not discuss any of these models in detail in this report.

yielded an R-square of 0.06 and an RMSE of 0.182. Our final model combines selected context variables and selected teacher practices variables, along with the mean school PSAT/NMSQT score. This model has an R-square of 0.27 and an RMSE of 0.167. Thus, the selected survey variables do predict a modest amount of (residual) student performance, accounting for an additional 21 percent of the variance over mean school PSAT/NMSQT score alone. There is also a nearly 8 percent reduction in the RMSE.

Similarly, for AP Biology public school teachers (MCR4), the initial regression that used mean school PSAT/NMSQT score as the only predictor yielded an R-square of 0.06 and an RMSE of 0.179. Our final model has an R-square of 0.33 and an RMSE of 0.158. The context and teacher practices variables account for an additional 27 percent of the variance over the mean PSAT/NMSQT score alone. Moreover, we have found both teaching-context and teacher-practice variables that help account for differences in student performance. In contrast to the fitted model for MCR3, this model includes more teacher-practice variables, possibly indicating that teacher practices are more important at higher performance levels.

Combining AP Biology public and nonpublic school teachers (MCR3), our final model has an R-square of 0.20 with a reduction of the RMSE to 0.175. Thus, an additional 11 percent of the total variance is accounted for by the five variables relating to context for teaching and teacher practices. Combining AP Biology public and nonpublic school teachers (MCR4), our model accounts for an additional 35 percent of the variance. Like the public school model with MCR4, this model includes more teacher-practice variables than the model for MCR3, possibly indicating that, again, teacher practices are more important at the higher levels.

For AP U.S. History public school teachers, our final model accounts for 32 percent of the total variance of the criterion MCR3, 23 percent more than mean PSAT/NMSQT score alone. For the model with MCR4, an additional 20 percent of the variance was accounted for by teacher contexts and practices. From the original model, we can see that the RMSE has decreased slightly, from 0.171 to .0159.

For the model combining AP U.S. History public and nonpublic school teachers (with the MCR3 criterion), the initial regression that used mean school PSAT/NMSQT score alone resulted in an R-square of 0.10 and an RMSE of 0.188. No teacher-practice variables were significant, so the context for teaching model is the final model, and accounts for only an additional 6 percent of the variance. For the model with MCR4, the model accounts for 34 percent of the variation in the student residuals, which is 26 percent more than the original model and has reduced the RMSE from 0.178 to 0.156.

AP® Biology Public School Teachers with Criteria of MCR3 and MCR4

For the model of AP Biology public school teachers with criterion of MCR3, 442 teachers were included in the analyses; for the criterion of MCR4, there were 435 teachers. All variables from the context for teaching and teacher practice profile models were subjected to the analyses as described earlier to create a model for each criterion (see Table 4). The most parsimonious models were both significant ($p < 0.01$); the overall test of models show $F(31, 408) = 4.95$ for MCR3 and $F(35, 402) = 5.70$ for MCR4. Please see Tables A1 and A2 in Appendix A for detailed statistics.

The variables from the survey listed in Table 4 are the ones that differentiated more successful AP Biology public school teachers from less successful teachers. The first variables listed are the ones that were significant in both models, followed by the significant variables in MCR3 or MCR4, but not both. Note that this is based on comparison of LSMs.[9] The dependent variable was MCR3 or MCR4 (i.e., the actual number of a teacher's students who met the criterion AP score minus the number the mean PSAT/NMSQT score predicted would meet the criterion divided by the total number of students with AP scores). It should be noted that all omitted variables were not significant. This could be due to one or more of the following reasons: (a) There is no association between the variable and the criterion, (b) the variable was highly correlated with other variables included in the model, (c) there is little heterogeneity among teachers on the variable, or (d) there is insufficient power in the analysis, since too many degrees of freedom were required to code the response categories. For all models, the exclusion of a variable does not mean it is unimportant but, rather, that it does not account for differences in mean student outcomes among teachers, given the other variables in the model.

There were three variables that differentiated AP Biology public school teachers in both models (MCR3 and MCR4). The partial η-squared for these variables were usually the largest in each model outside of the PSAT/NMSQT score. The variables were frequency of class meetings (partial η-squared values were 0.023 and 0.032 for MCR3 and MCR4, respectively), percentage of students who take the AP Exam (0.032 and 0.047), and the use of AP Exam topics and/or rubrics (0.017 and 0.031). In sum, students whose AP Biology classes met every day, whether for 30–60 minutes or 61–110 minutes per day, throughout the school year, performed better than those students whose classes met less frequently. The LSM

[9] Significance here means that after including the other variables in the model, each makes a significant reduction in the mean-square error.

Table 4

AP Biology Public School Models: Statistically Significant ($p < 0.10$) Variables

Variables	*MCR3*	*MCR4*
Percentage of students who take the exam	+	+
Influence of resources: Frequency of class meetings	+	+
Influence of resources: AP Exam topics and/or scoring rubrics	+	+
Number of years teaching	+	
Participation in AP professional development activities (PAPPD): Attending an AP Institute	–	
PAPPD: Reviewing AP Biology Teacher's Guide	–	
Class size	–	
Objective: Learn scientific methods	+	
Assessment: Tests requiring lengthy written responses	+	
Influence of resources: Frequency of using exemplary syllabi from other AP Biology classes		–
Type and frequency of review activities: Teacher estimate of student time dedicated to study course material on their own		+
Teaching test-taking strategies		X
Computer use: Teacher researching information on the Internet		–
Focus of attention to prepare students for AP Exam		X
Percentage of class time dedicated to prepare for the AP Exam during the month prior to the AP Exam		X

Note: Response choices to the variables are listed in increasing order unless the variable is listed as nominal. Explanations of the symbols used in this table are: "+" represents a positive relationship with the variable, "–" represents an inverse relationship with the variable, and "X" represents a nonlinear relationship with the variable (e.g., the variable either is not ordinal or the relationship is neither positive nor negative).

showed a significant difference between block scheduling and a daily schedule of 61–110 minutes per day, with a residual difference of 0.08. The proportion of students who take the AP Exam in the class also has an effect: classes in which fewer than 50 percent of students take the AP Exam tend to perform significantly worse than classes with 75–100 percent of students taking the exam. In short, the higher the percent of students in the class taking the exam, the better classes tend to do. Last, most teachers say that they find the use of AP Exam topics and/or rubrics to be between "somewhat" and "extremely" influential. For MCR3, the more useful the teacher found this resource, the better his or her classes performed. For MCR4, there was no pattern, except that teachers who said these resources were "somewhat" influential had classes scoring significantly lower than those claiming "very" and "extremely" influential. In general, the more often teachers found these resources to be useful, the better their classes performed.

The variables that were significant only for the model with MCR3 were mainly teaching related (e.g., professional development and teaching experience): overall teaching experience, the use of essay topics and/or rubrics as a resource, attending an AP Institute, use of the AP Biology Teacher's Guide, and class size. As expected, classes of beginning teachers (teaching less than three years) perform worse (lower MCRs) than classes of more experienced teachers. There are significant differences between teachers within their first three years of teaching and those who have been teaching four to six years. We see the residual is 0.14 smaller for classes of beginning teachers than for teachers of four to six years. This finding is not surprising, as studies on beginning teachers indicate that teachers need at least three years to become adept in their teaching (Darling-Hammond, 1997; Feiman-Nemser, 2001). It may be that experienced teachers, who are more familiar with the course content and the exam, can fine-tune their teaching, which results in better class performance. An interesting aside is that biology teachers are much more experienced in teaching biology than teaching AP Biology. In fact, 75 percent of teachers have 10 or more years experience teaching biology, which may imply that teachers do not teach AP Biology until they have substantial teaching experience.

Two variables dealt with AP professional development, the first of which was attending an AP Institute. AP Institutes are usually weeklong gatherings where teachers work together on different lessons. Fifty percent of teachers claimed they had never attended an institute, while 40 percent of teachers had only attended once. These teachers' classes tend to perform similarly on the AP Exam. The remaining 10 percent had attended an AP Institute more than once, but their classes are performing worse than the classes of the other two groups. Perhaps teachers are sent by their supervisors to attend an additional institute because their classes did not perform adequately on the AP Exam the first time. These teachers may not be as effective in the classroom, regardless of the number of times they attend an institute.

The AP Biology Teacher's Guide provides sample course outlines designed by successful AP Biology teachers at seven diverse schools. Each of the detailed outlines includes a description of the school environment, the objectives of the course, prerequisites, student selection, a timeline for topics, and assignments. Most teachers use this guide to inform their teaching. A small number of teachers who had taught more than 10 years who indicated they did not use the AP Biology Teacher's Guide had classes that performed much better than classes of less experienced teachers. It may be that these teachers have developed their own guide for teaching AP Biology or no longer needed to use the AP guide due to their extensive experience teaching AP Biology. Finally, larger classes are more likely to underperform on the AP Exam; classes of 15 or fewer students have significantly better performances than all other larger classes.

The two remaining variables for this model are related to teacher practices, which are learning the scientific method and tests requiring lengthy responses. Most teachers (68 percent) emphasize learning the scientific method with average or slightly above average emphasis. In general, the more a teacher emphasizes this learning objective, the better his or her classes tend to perform. Teachers who employ at least average emphasis have classes performing 0.02 to 0.06 better than those who stress learning the scientific method with less than average emphasis. In terms of assessment methods, teachers who used tests requiring lengthy written responses at least once or twice a month had classes perform significantly better on the AP Exam than those who did this only several times per year or rarely.

The variables in the model for MCR4 only are mostly related to teacher practices: the teacher's estimate of the time students dedicated to studying course material on their own, the teaching of test-taking strategies, teachers researching information on the Internet, focus of attention to prepare students for the AP Exam, and percent of time directly related to helping students pass the AP Exam the month before the exam. The use of exemplary syllabi as a resource was the one remaining variable included in this model for measuring teaching context. The more often teachers used this resource, the lower the predicted student success. It appears that teachers who rely less on this resource have classes that perform better than those who rely on it more heavily. This may be related to experience and degree of comfort with the course content (e.g., teachers who are more familiar with the course probably need to consult the syllabi less often).

Regarding teachers' estimates of the time students dedicate to studying AP material, about 67 percent of teachers believe their students study at least 4 to 9 hours per week on their own for the AP Exam. A small number of teachers claim their students study over 20 hours per week, and these students performed significantly better than students who, according to the teacher's estimate, studied less often. With regard to test-taking strategies, although LSM showed significant differences among groups for teaching test-taking strategies, there was no consistent pattern across groups. For instance, teachers who taught test-taking strategies the least had the highest performing classes, whereas those teaching them once or twice a month had better performances than those teaching them several times per year, once or twice per week, or almost every class period. The analysis of teachers' use of the Internet indicates that most teachers (92 percent) seek information on the Internet, and teachers who do not use the Internet had classes that performed better than those teachers who did. It would be interesting to see where on the Internet these teachers obtained their information. In terms of preparing students for the exam, most teachers prepare their students equally for both parts. It appears, however, that teachers who focused on the multiple-choice section of the exam produced significantly higher-performing classes (a difference in residual of 0.09). The last variable included in the model is the percent of time dedicated to preparing students a month before the AP Exam takes place. The distribution of this variable is roughly uniform across the five choices: less than 20 percent, 21–40 percent, 41–60 percent, 61–80 percent, and more than 80 percent. Most teachers spending less than 20 percent of their time the month before the exam in preparing their students obtained the best results, followed by those who spend more than 80 percent. Those that spend less than 20 percent of their time may not need that extra preparation time because they feel confident in how they have prepared their students over the course of the year. And those that spend extra time the month before (80 percent or more) may not have spent any time earlier in the school year specifically preparing students for the AP Exam. They may be having students take practice tests and work specifically with retired exams, which would give the students familiarity with the exam format.

Combined Analysis of AP Biology Public and Nonpublic School Teachers for MCR3 and MCR4

Recall that our original plan was to conduct separate analyses of public, private, and parochial school teachers. However, even when combining private and parochial school teachers into a category we defined as "nonpublic," the sample size was quite small. Therefore, our next set of analyses combines public and nonpublic school teachers, with a test for difference in school type.

For the model with MCR3, 574 teachers were included in this analysis—132 nonpublic schools and 442 public schools. For MCR4, there were 548 teachers included in the model, 113 nonpublic and 435 public. Note that since we are now analyzing two different types of schools, public and nonpublic, we also tested if the differences between school type were significant or not by testing the interactions of school types with the variables in the model. The asterisks in Table 5 indicate that the variable was significantly different by school type. For both public and nonpublic AP Biology teachers, the variables discussed below proved significant ($p < 0.01$) in the models; the overall test of the model shows ($F(32, 554) = 4.73$) for MCR3 and ($F(64, 483) = 4.97$) for MCR4. For statistical details, please see Tables A3 and A4 in Appendix A.

The contextual variable that was significant in both models that included nonpublic school teachers was class size (η_p^2 were 0.032 and 0.019 for MCR3 and MCR4, respectively). Larger class sizes are associated with lower performances, with a significant difference across school type because nonpublic schools tend to have smaller classes than public schools. Also, smaller class sizes were associated with better exam performance. For class size, the significant difference by school type may be due to the fact that a higher percentage of nonpublic school teachers have smaller classes than teachers in public schools.

The teacher practices variable common to both models with the nonpublic teachers is student researching of information on the Internet; most students tend to use the Internet for this purpose, but it is unclear why classes that do this tend to perform worse than those who do not use the Internet for research. Teachers claim that over 80 percent of their students research information

Table 5

AP Biology Public and Nonpublic School Models: Statistically Significant ($p < 0.10$) Variables

Variables	*MCR3*	*MCR4*
Class size	–	– *
Adequacy of facilities: Laboratories, lab tables, sinks, etc.	+ *	
Computer use: Student researching information on the Internet	–	–
Objective: Learn scientific methods	+ *	
Student activities: Submitting reports on experiments or observations	–	
PAPPD: Attendance at an AP workshop		*
Number of AP Biology classes taught		– *
Percentage of students who take the exam		+ *
Influence of resources: Exemplary syllabi from other AP Biology classes		–
Influence of resources: AP Exam topics and/or scoring rubrics		+
Frequency of class meetings		+
PAPPD: Teach an AP Institute		*
Assessment: Laboratory notebooks or journals		+ *
Assessment: Presentations by students		+
Frequency of feedback: Phrase- or sentence-length descriptions of students' performance		+
Student activities: Apply biology concepts to simulated or real-world problems		+

Note: Explanations of the symbols used in this table are: "+" represents a positive relationship with the variable, "–" represents an inverse relationship with the variable, and "*" represents a significant difference by school type.

on the Internet. It is unclear what the other 20 percent of students are doing, and there were no other items in the survey that could help explain this finding.

Eight of the variables in Table 5 are related to context for teaching: class size, attendance at an AP workshop, number of AP Biology classes taught, percentage of students who take the exam, use of exemplary syllabi from other AP Biology classes, use of AP Exam topics and/or scoring rubrics, frequency of class meetings, and teaching at an AP Institute. Some of these variables are significant by school type, meaning there are differences between public school and nonpublic school teachers. For instance, most public school teachers claim to have attended an AP workshop more than once, while most nonpublic teachers have only attended an AP workshop once; this difference is significant between the two school types. For public schools, the more often teachers attended AP workshops, the better their classes performed. For nonpublic schools, teachers who never attended an AP workshop had classes performing better than those teachers who had attended an AP workshop. It may be that these teachers have other resources at their disposal.

In terms of number of AP Biology classes that were taught, most teachers said they only taught one class. For both public and nonpublic school teachers, the fewer AP Biology classes taught, the better the teachers' classes performed. While only 5 percent of nonpublic school teachers taught more than two AP Biology classes, 10 percent of public school teachers taught more than two. The differences between student performance of public and nonpublic school teachers is statistically significant in that public school classes tend to have a better MCR than nonpublic school classes. In addition, there was a statistically significant difference between public and nonpublic schools with respect to the percentage of students who take the AP Exam; 93 percent of nonpublic school teachers have 75 to 100 percent of their students take the AP Exam as compared to 76 percent of public school teachers. Higher percentages of students taking the AP Exam are associated with higher predicted student success on the exam for both groups.

Most teachers say that, as a resource, exemplary syllabi are somewhat influential (this variable had the largest η_p^2 in the model at 0.025). In fact, the more influential teachers find syllabi to be as a resource, the worse their classes perform. This may be due to new teachers relying on such materials to become familiar with the course, and the data suggests that this seems to be associated with AP teaching experience. As seen in the model with only public school teachers, the more often teachers used the exam topics and scoring rubrics, the better their classes performed. In terms of schedule, there are no nonpublic school teachers who have taught AP Biology as a compressed semester course, while a small percent of public school teachers have. There were statistically significant differences between the schedules of public and nonpublic school teachers, as 30 percent of public school teachers claim to teach AP Biology classes every day for more than an hour as compared to only 7 percent of nonpublic school teachers. Most teachers (97 percent of public school teachers and 98 percent of nonpublic school teachers) have never taught at an AP Institute; when comparing public to nonpublic classes for these teachers, we find public school classes performing better than nonpublic school classes. The numbers for teachers who taught at an AP Institute were too small to compare.

Four variables are related to teacher practices. One teacher-practice variable from the models based on only public school teachers remained significant in the models that added nonpublic school teachers: applying biology concepts to simulated or real-world problems. As with public school teachers, the more often teachers asked students to apply biological concepts in a real-world context, the better the classes did on the AP Exam, but there are significant differences when public and nonpublic school teachers are combined and between the two school types. Classes of most nonpublic school teachers perform worse the less frequently the teacher asks them to apply their biology concepts, whereas classes of public school teachers tend to perform better than nonpublic school classes regardless of how often the teacher asks them to do this. It is unclear as to why this pattern is true for public schools, and there were no other data sources for us to investigate further to try to explain this finding.

The remaining three variables unique to the model with MCR4 are related to teacher practices: providing phrase- or sentence-length feedback, frequency of assessment of student lab notebooks, and student presentations. The more often teachers claim to give feedback, the better their classes perform. LSM shows the significant difference in student performance between those teachers who hardly ever provided such feedback and those who did so between once or twice a month and once or twice a week. A majority of teachers assess students by checking their laboratory notebooks less than once or twice a month. For such teachers, the more often they assessed students this way, the better the students performed. For nonpublic school teachers, those who hardly ever assessed their students' lab notebooks had students performing significantly worse than those who claimed to do this more often. This difference also proved significant when comparing public and nonpublic school teachers. Most teachers assess their students based on student presentations at most "sometimes" per year (80 percent of public and 87 percent of nonpublic school teachers). In general, the more often teachers assessed students this way, the better their classes performed on the AP Exam, and public school classes tend to perform significantly better than nonpublic school classes for this type of assessment.

Comparison of Public and Nonpublic School Models for AP Biology

We have now seen two models for public school teachers for the two AP criteria, as well as two models when we have included nonpublic school teachers. There are no variables that were significant in all four models for AP Biology. However, there were three variables that were significant in three of the four models, all of which are context for teaching variables: frequency of class meetings, percentage of students taking the exam, and class size.

As seen in the public school model, nonpublic school teachers with classes that meet every day, whether for 30–60 minutes or 61–110 minutes per day, also have better performing classes than those teaching on either block scheduling or compressed semesters. See Table 6. Last, the higher the proportion of students in a class taking the AP Exam, the better classes tended to do. Nonpublic schools have a higher percent of students taking the AP Exam than public schools, and this difference is significant between the two types of schools. In fact, more nonpublic schools teachers claimed that all their students are required to take the AP Exam than did public school teachers. In terms of class size, no nonpublic school teachers who responded to this survey had classes of more than 30 students, whereas 6 percent of public school teachers in this sample taught classes of this size. Although nonpublic school teachers tended to have significantly smaller classes than public school teachers, it was true for both school types that class size and student performance were inversely related.

When adding in nonpublic schools, the model accounts for 7 percent less of the total variance than public school alone for the model with MCR3, but 7 percent more of the total variance for the model with MCR4. This finding may be due to the different variables included in each of the models. For instance, public school teachers had more variables related to professional development and resources that were not significant when nonpublic school teachers were added. For the models that included both public and nonpublic schools, assessment and feedback were significant variables that were not found in the public school–only models. It appears that there are differences in public and nonpublic school teaching contexts as well as their teaching practices.

AP U.S. History Public School Teachers with Criteria of MCR3 and MCR4

The model for MCR3 includes 497 teachers, while the model for MCR4 includes 520 teachers. Table 7 lists the variables in the models, with the overall test of the model for MCR3 ($F(35, 461) = 6.30$, $p < 0.01$) and MCR4 ($F(30, 499) = 5.94$, $p < 0.01$). There are six variables in common between the two models; three of these variables are contextual while three are related to teacher practice. Most variables that are unique to the model for MCR3 are mainly context-for-teaching related (see Tables B1 and B2 in Appendix B for details).

Six variables were significant in both models. The first three variables related to school and class context: schedule (η_p^2 were 0.019 and 0.012 for MCR3 and MCR4, respectively), resources (0.032 and 0.028), and percentage of students who take the exam (0.028 and 0.050). Students in classes that met every day for more than an hour tended to be more successful than those who met less often; worst off were students who took the course compressed into the fall semester. In terms of instructional materials and other resources AP teachers may need, the availability of more resources and materials resulted in higher prediction of student success. Classes did significantly worse when teachers felt that they received only "some" or "hardly any" of the resources they needed than when teachers felt that "most" or "nearly all" the resources were available. Classes in which less than 75 percent of students take the AP Exam do significantly worse than those classes that have between 75 percent and 99 percent of students

Table 6

Comparison of AP Biology Public and Nonpublic School Models: Statistically Significant ($p < 0.10$) Variables

Variables	*Public only: MCR3*	*Public only: MCR4*	*Combined public and nonpublic: MCR3*	*Combined public and nonpublic: MCR4*
Frequency of class meetings	–	–		–
Percentage of students who take the exam	+	+		+
Class size	–		–	–

Note: Explanation of the symbols used in this table are: "+" represents a positive relationship with the variable and "–" represents an inverse relationship with the variable.

Table 7

AP U.S. History Public School Models: Statistically Significant ($p < 0.10$) Variables

Variables	*MCR3*	*MCR4*
Frequency of class meetings	–	–
Adequacy of school resources	+	+
Percentage of students who take the exam	+	+
Objective: Develop historical research skills and techniques	X	X
Student activities: Explain reasoning or thinking	+	+
Type and frequency of review activities: Teacher estimate of student time dedicated to study course material on their own	+	+
PAPPD: Review of released AP Exams	+	
PAPPD: Participated in AP Reading(s)	+	
School policies regarding who takes the exam	+	
Educational level attained		+
Control of selecting content, topics, and skills to be taught		+
Objective: Develop skills for supporting stating and claims	+	

Note: Explanation of the symbols used in this table are: "+" represents a positive relationship with the variable, "–" represents an inverse relationship with the variable, and "X" represents a nonlinear relationship with the variable.

taking the exam. This information is consistent with the results about the performance of students encouraged to take the exam. It appears that students attending schools that offer high incentives to take the AP Exam were more likely to take the exam.

Three teacher-practice variables appeared significant in both models of teaching practice; they are related to teacher's stated learning goals, type of activities implemented in the classroom, and review activities provided in the month prior to the AP Exam. Specifically, the variables are developing historical research skills and techniques, having students explain their reasoning or thinking, and the teacher's estimate of the time students dedicated to study for the AP U.S. History Exam on their own.

In terms of teachers' emphasis on developing students' historical research skills and techniques, teachers who claimed to do this either with the least or most emphasis had classes that did better than classes of other teachers; however, there are relatively few teachers at the extremes. Concerning instructional activities, our results indicate that teachers who ask their students to explain their reasoning or thinking in almost every class have classes that do better than those of teachers who do this only several times per year or once or twice a month. In relation to AP preparation in the month prior to the exam, the analysis reveals that the more time the teacher estimates that his or her students study on their own for the AP Exam, the better the teacher's classes performed. Most teachers estimated their students studied on their own about 4 to 9 hours per week. Those teachers who estimated that their students studied at least 4 hours per week had classes that performed better than classes of students who studied less than 4 hours per week. Teachers who estimated that their students studied more than 20 hours per week on their own to prepare for the AP Exam had the highest performing classes.

For MCR3, two variables related to professional development proved significant: reviewing released AP Exams and participating in AP Readings. Most teachers have reviewed released AP Exams more than once, and the classes of these teachers perform better than classes of teachers who have not done this at all or only once. The more teachers reviewed the AP Exams, the better their classes tended to do. Most teachers have not participated in an AP Reading; however, those who have participated in an AP Reading more than once have classes that perform better than those who have not attended a Reading at all or only once.

Regarding school policies with respect to who takes the exam, most teachers say either all students must take the AP Exam or are encouraged to take the AP Exam. Students in classes in which all class members are encouraged to take the AP Exam, regardless of how well they do in the course, tend to do better compared with classes in which students either must take the exam or decide on their own whether or not to take it. Teachers who say that only students who performed well in the class are encouraged to take the AP Exam have the highest performing classes, with significantly better performance than those requiring all students to take the exam. This difference is reasonable since only strong performing students are selected to take the exam. The heavier the emphasis teachers placed on the development of skills for stating and supporting claims, the better their classes performed. Teachers who emphasized this the most had classes that performed better than classes of teachers who did not emphasize it as much.

For MCR4, the two significant context variables were educational level and control of selecting topics and content. The higher the teachers' degree, the better their classes tended to perform; most teachers in this sample have at least a master's degree, as well as being certified. Most teachers have credits beyond a master's degree, and it is these teachers that have the highest student performance, between 0.02 and 0.06 better than classes of teachers with other degrees. LSM shows these differences to be statistically significant. Most teachers tend to feel they have control of selecting the content, topics, and skills to be taught, and classes of these teachers perform better than those of teachers who do not feel they have as much control.

Combined Analysis of AP U.S. History Public and Nonpublic School Teachers for MCR3 and MCR4

Similar to the combined analyses for AP Biology, we ran two models combining public and nonpublic school teachers for AP U.S. History. There were 162 and 111 nonpublic school teachers for the models MCR3 and MCR4, respectively. The variables in Table 8 were retained in the parsimonious model, combining contextual and teaching practices for MCR4 ($F(67, 559) = 4.46$, $p < 0.01$). There were no teacher practices in the model for MCR3, so the model is the same as the contextual model. (Please see Tables B3 and B4 in Appendix B for details.)

Two variables were significant in both models: discussions with colleagues and mentors (η_p^2 were 0.010 and 0.012 for MCR3 and MCR4, respectively) as well as adequacy of school resources (0.026 and 0.0222). When asked about discussions with colleagues and mentors,

Table 8

AP U.S. History Public and Nonpublic School Models: Statistically Significant ($p < 0.10$) Variables

Variables	*MCR3*	*MCR4*
Influence of resources: Discussions with colleagues and mentors	– *	– *
Adequacy of school resources	+	+ *
AP teaching experience	+	
Influence of resources: Exemplary syllabi from other AP Biology classes	X	
PAPPD: Attended an AP Institute		+ *
PAPPD: Consulted for an AP workshop		+ *
PAPPD: Review of released AP Exams		+
Hours of preparation for AP Biology/USH class per week		+ *
Percentage of students who take the exam		+
Class size		–
Student activities: Discuss controversial events or themes		+ *
Student activities: Use tools of analysis to generate hypotheses or develop arguments		+
Objective: Develop historical research skills and techniques		+ *
Provide instruction to small groups of students		– *

Note: Explanations of the symbols used in this table are: "–" represents an inverse relationship with the variable, "*" represents a significant difference by school type, "+" represents a positive relationship with the variable, and "X" represents a nonlinear relationship with the variable.

most teachers say they are "somewhat influential" as a resource, followed by those saying they are "very" or "slightly" influential. Nonpublic school teachers who say these are an "extremely" influential resource have the worst performing classes; it may be that these teachers are new and rely considerably on colleagues and mentors in teaching the course. For nonpublic school teachers, there appears to be an inverse relation between how influential the teachers found such discussions and how well their classes performed. This finding is not true for public school teachers, as the more influential these teachers found discussions with colleagues and mentors, the better their classes performed. In terms of adequacy of resources, most teachers said they have "good" or "excellent" resources. The more teachers claimed to have quality resources, the better their classes tended to perform. Those classes of teachers with poor resources scored significantly worse than those classes with good or excellent resources. For MCR4, while 20 percent of nonpublic school teachers report "poor" or "fair" resources, 34 percent of public school teachers report inadequate resources; the difference between public and nonpublic school resources is statistically significant for MCR4.

There are six context-for-teaching variables unique to the model for MCR4, which tend to be related to professional development and influence of resources. When adding nonpublic school teachers to the model with MCR3, we see that none of the significant variables from the public school teaching practices model are significant. The model for MCR3 includes only four variables on context for teaching, two of which are in common with MCR4. For facilities, the breakdown by means indicate that "good" is good enough, as the responses break down to "poor" and "fair" versus "good" and "excellent."

The other 10 remaining variables are unique to the model for MCR4: Half of them are context-for-teaching related, while the other half are from the teacher practices profile. Three of these variables are related to professional development. Most teachers have never attended an AP Institute (51 percent of public and 59 percent of nonpublic school teachers), and there was a statistically significant difference between public and nonpublic schools for this variable. For public school teachers, there was no mean difference in student performance by the number of times the teacher attended an AP Institute. However, for nonpublic school teachers, the more often they attended an AP Institute, the better their student performance, a residual difference of 0.06 and 0.13 larger than those who never attended. In terms of consulting for an AP workshop, only 3 percent of public and 4 percent of nonpublic school teachers have ever done so. The more they have consulted, the higher their classes' performance. The difference is negligible for public school teachers, but larger for nonpublic school teachers (a difference of 0.07). The last professional development activity is reviewing released AP Exams. Most teachers have done this, and the more often they reviewed released AP Exams the better their classes tended to perform.

The last two context-for-teaching variables deal with student numbers in terms of percentage taking the exam and class size. In general, the higher the percentage of students taking the AP Exam, the better the class performance (this variable had the largest η_p^2 value for the MCR4 model at 0.034, which was even higher than the PSAT/NMSQT effect size). For public school teachers, this difference is between 0.08 to 0.11, and for nonpublic school teachers the differences is 0.08 to 0.09. Only 35 percent of public school teachers have all students in their class take the AP Exam, compared to 71 percent of nonpublic school teachers. However, a large number of public school teachers have between 75 percent and 99 percent of students taking the AP Exam (as compared to between 18 percent and 41 percent of nonpublic school teachers). Public school classes tend to be larger than nonpublic schools. For instance, 57 percent of public school teachers have 21 or more students in their classes while only 31 percent of nonpublic school teachers have classes this size. As expected, the best performers are the ones in the smaller classes (fewer than 15 students), and the worst performers are those in classes with more than 30 students.

Five variables on teacher practice were unique to the model with MCR4. The development of historical research skills and techniques shows that most teachers place between "slightly less" to "slightly more than average" emphasis on it in relation to other learning objectives. There is no clear relation between the amount of emphasis on this objective and student performance; however, classes taught by public school teachers perform worse than classes of nonpublic school teachers when both groups of teachers place average or slightly more than average emphasis on the development of historical research skills and techniques.

Most teachers claim to provide instruction to small groups less than once or twice a month. Classes of public school teachers who did this the least often had the highest performers, while those that did it every class had the lowest performance. This finding may reflect teachers' attention to the individual needs of students, so that teachers who provided small group instruction more often did so because their students needed it. We ran analyses of how often the teachers used small group instruction by class size. For the very large and very small classes, 52 percent of classes of 15 or fewer students as well as 54 percent of classes with more than 30 students, teachers claimed to use small groups very infrequently (hardly ever to once or twice a year). Classes of such teachers scored significantly better (0.08 to 0.18 better) than those teachers who used small groups more frequently (one to two times per month or more). For nonpublic school teachers, classes perform similarly no matter how often their teacher provides small group instruction, and these differences in school type proved significantly different.

There were two variables associated with student activities: using tools of analysis to generate hypotheses or develop arguments, as well as discussing controversial events or themes. Most teachers had students use tools of analysis once to twice a month. The more often teachers had students practice this kind of analysis, the better their classes fared on the AP Exam. For teachers that did this every class period, classes performed significantly better than those of teachers who hardly ever had students participate in such an exercise (0.08 to 0.10 difference in residuals). Teachers tended to have students discuss controversial events or themes from once or twice per month to once or twice per week. Except for teachers who claimed to have students do this every day, the more often students engaged in such discussions, the better they did. An interesting finding among nonpublic school teachers was that those who have students do this almost every class period perform between 0.12 to 0.18 worse than those who do this less often. The final teacher-practice variable was the number of hours teachers would prepare for their AP U.S. History class(es). The more time a public school teacher spends preparing for class, the better their classes perform. This is true for nonpublic teachers except for those who claim to prepare for more than 20 hours; these may be new teachers who, due to inexperience, need to spend this amount of time to prepare for class. The differences in the hours a teacher prepares for AP U.S. History classes is statistically significant when comparing public to nonpublic school teachers.

Comparison of Public and Nonpublic School Models for AP U.S. History

In the context-for-teaching models, adequacy of resources was the only variable that remained significant across all four models (see Table 9). For both public and nonpublic schools, the better the resources, the better the classes performed. The two variables that were in three of the four models were percentage of students taking the exam and developing historical research skills. Although a higher percentage of nonpublic school students tended to take the AP Exam, the more students in the class who took the exam, the better the class performance. Most teachers claimed to place slightly less to slightly more than average emphasis on developing historical research skills and techniques. Teachers who placed average emphasis on developing historical research skills had classes that performed better than classes whose teachers placed either more or less than average emphasis on developing such skills.

The variables unique to models with both public and nonpublic school teachers are context-for-teaching related. Only professional development practices significant for public schools did not hold up when nonpublic school teachers were added to the model. Unlike AP Biology, it appears that there are considerable differences between public school and nonpublic school teachers of AP U.S. History.

When adding nonpublic schools for MCR3, the model accounts for 16 percent less of the total variance than the model for public schools alone. It accounts for 8 percent more of the total variance than the model for MCR4. Since the model for combined public and nonpublic school teachers did not have any significant teacher practices, this finding is understandable, if disappointing. For MCR4, nonpublic school teachers had more variables related to professional development and student activities that were not significant than in the models that only included public school teachers. It appears that there are differences in public and nonpublic schools' teaching contexts, as well as teaching practices across the different criteria for AP U.S. History.

Limitations

The teacher survey was designed to generate a nationally representative sample of teachers in both AP Biology and AP U.S. History. Substantial nonresponse implies that the

Table 9

Comparison of AP U.S. History Public and Nonpublic School Models: Statistically Significant ($p < 0.10$) Variables

Variables	*Public only: MCR3*	*Public only: MCR4*	*Combined public and nonpublic: MCR3*	*Combined public and nonpublic: MCR4*
Adequacy of school resources	+	+	+	+
Percentage of students who take the exam	+	+		+
Objective: Develop historical research skills and techniques	X	X		X

Note: Explanation of the symbols used in this table: "+" represents a positive relationship with the variable and "X" represents a nonlinear relationship with the variable.

obtained sample cannot be so considered. This difficulty was exacerbated by the difficulty in linking students to their teachers in schools with more than one teacher for the particular AP subject. Despite our best attempts to work through the AP Coordinators, we were able only to recover a small fraction of the data from those schools. Furthermore, students without PSAT/NMSQT scores could not be included in the analysis. Finally, as explained below, we eliminated classes with fewer than eight scores from the analysis. Consequently, although the sample of teachers for each subject is comparatively large and the amount of data collected voluminous, our results can only be suggestive of what we might have found had the original survey sample been available.

We also appreciate the fact that the different decisions made during the course of the analysis lead to one of many possible perspectives on the structure of the relationship between teacher practices and student achievement. Among the more critical decisions were: (a) In order to find a more appropriate dependent variable than raw AP scores, we used the difference between the dichotomized score for the student and the corresponding expected value based on the student's PSAT/NMSQT score; (b) in order to conduct a class-level analysis, we employed the mean class residual (MCR); (c) in order to have stable estimates of the dependent variable, we deleted classes with eight or fewer students; and (d) in order to keep the data analysis manageable, we employed a stagewise approach to variable selection, employing parsimonious GLMs. The rationale for each of these decisions is explicated below.

1. We used the difference between the dichotomized score and the expected value based on AP score so that the criterion would more closely resemble a gain score for the student. In comparison to the raw score, the residual should be a more sensitive criterion for revealing possible teacher contributions to student achievement. In creating this residual, we treated the individual PSAT/NMSQT score as a proxy for prior achievement as well as other unmeasured relevant student characteristics. It is certainly possible that by adjusting for PSAT/NMSQT scores we may have removed some criterion variance that could be accounted for by teacher characteristics. Moreover, use of this criterion places a greater "burden" on the subsequent regression analyses, since much of the systematic variation was removed through the preliminary logistic regression.
2. The MCR was used because our analyses are conducted at the teacher/class level. The MCR, constructed by taking the average of the residuals for all the students in a class, is still a rather noisy criterion because the original dichotomous outcome score has relatively little information.
3. The reason for removing classes with eight or fewer students was to assure reasonable stability of the MCR. When classes with fewer than eight students were included in the analysis, we observed substantial variability often due to the result for a single student. We felt that an analysis restricted to more stable MCRs would yield more informative results.
4. The different parsimonious GLMs used as a first step were found using the partial sums of squares (Type III), employing a p-value of 0.15 for inclusion in subsequent models. Some would argue that this p-value was too liberal, as the most common p-value used for identifying significance is 0.05. We used this moderately liberal p-value because of possible collinearity among the variables. We did not want to eliminate potentially interesting variables too early in the analysis. Nonetheless, because of the large number of (correlated) variables used in the models, what emerged as significant—or not—may be due in part to the relationships among the predictors.

Conclusions

The purpose of this study was to examine the relation between AP teachers and students' AP Exam performance. It appears that there is evidence of a relationship between teacher practices and student achievement in AP classes once we have adjusted for the student achievement proxy (PSAT/NMSQT score) and teacher context variables. As summarized below, these teacher practices differ depending on the subject matter as well as the criterion variable used (an AP Exam score of 3 or better, or 4 or better).

For both AP Biology and AP U.S. History, models have been presented incorporating context-for-teaching and teacher-practice variables, for public and nonpublic school teachers, and for different standards of success on the exams (MCR3 and MCR4). There are some similarities in the variables that accounted for differences in the criterion across classes. Two variables included in most of the models are the frequency of class meetings and the percent of students taking the exam. See Tables 10 and 11.

We found that for AP Biology public school teachers, similar context-for-teaching and teacher-practice variables predict success on the AP Exam for both MCR3 and MCR4. For both public and nonpublic school AP Biology teachers, there are more teacher-practice variables unique to the model of MCR4. This finding seems to support the notion that teacher practices are more important than context in producing better class performance. If true, this result suggests that teachers do have an impact on student achievement.

Table 10

Public School Models: Statistically Significant ($p < 0.10$) Variables

	MCR3	*MCR4*
AP Biology	Overall experience PD: Attended an AP Institute PD: Review AP Biology Teacher's Guide IOR: Schedule Percentage of students who take exam Class size OB: Learn scientific methods AS: Tests requiring lengthy responses RA: Estimate of time students study on their own	IOR: Exemplary syllabi IOR: Schedule IOR: AP Exam topics and/or scoring rubrics Percentage of students who take exam Teach test-taking strategies SA: Explain reasoning or thinking CU: Teacher research info on the Web Focus on AP Exam preparation Last month's percentage of class time to prepare for exam RA: Estimate of time students study on their own
	MCR3	*MCR4*
AP U.S. History	PD: Collaborate with mentor teacher PD: Review of released AP Exams PD: Participated in AP Reading(s) IOR: Schedule Adequacy of school resources School policy: Who takes the exam Percentage of students who take exam OB: Skills for supporting claims OB: Historical research skills/techniques SA: Explain reasoning or thinking RA: Estimate of time students study on their own	Educational level attained IOR: Schedule Adequacy of school resources Percentage of students who take exam OB: Historical research skills/techniques SA: Explain reasoning or thinking RA: Estimate of time students study on their own

Note: Explanations of the abbreviations used in this table are AS: assessment, CU: computer use, FB: feedback, IOR: influence of resources, OB: objectives, PD: participation in professional development activities, RA: review activity, and SA: student activity.

Table 11

Public and Nonpublic School Models: Statistically Significant ($p < 0.10$) Variables

	MCR3	*MCR4*
AP Biology	Class size OB: Learn scientific methods AS: Tests requiring lengthy responses SA: Submit reports on experiments CU: Students researching information on Web	PD: Attended an AP workshop IOR: Exemplary syllabi IOR: AP Exam topics and/or scoring rubrics Number of AP Biology classes taught IOR: Schedule Class size Percentage of students who take exam AS: Lab notebooks or journals AS: Presentations by students FB: Phrase/sentence descriptions SA: Design-conduct own AP Biology project CU: Students researching information on Web CU: Simulation and modeling
	MCR3	*MCR4*
AP U.S. History	AP teaching experience IOR: Exemplary syllabi IOR: Discussion with colleagues and mentors Adequacy of school resources	PD: Attended an AP Institute PD: Review of released AP Exams PD: Consulted for an AP workshop Adequacy of school resources Percentage of students who take exam Class size SA: Developing hypotheses/arguments SA: Discuss controversial events/themes Weekly hours of AP preparation time

Note: Explanations of the abbreviations used in this table are AS: assessment, CU: computer use, FB: feedback, IOR: influence of resources, OB: objectives, PD: participation in professional development activities, and SA: student activity.

For U.S. History, there are many more variables unique to each of the four models, specifically within context for teaching. Many of these variables were related to professional development. For each model, the number of contextual variables exceeds the number of teacher-practice variables. In the U.S. History public school model, we do not see more teacher-practice variables when using the more stringent standard of success as we did in the AP Biology analyses. However, in the model with public and nonpublic schools, we do find more teacher-practice variables. Similar to the finding with AP Biology, this seems to support the notion that teacher practices are related to better-than-predicted student success since these variables were significant for predicting MCR4.

Professional development was consistently significant in models for both subjects; however, the types of professional development tended to differ across subject area as well as public versus nonpublic and the two different AP success criteria. School and class context were also significant across the models. For both criteria, resources proved to be important for AP U.S. History teachers, both public and nonpublic. On the other hand, class size and schedule were most important for AP Biology. It may be that other materials/resources could enhance the student learning experience in AP U.S. History, while both public and nonpublic AP Biology teachers can attend better to smaller classes that meet daily.

With regard to teacher practices, for AP U.S. History teachers, emphasizing the skill for developing historical research skills and techniques proved most important. This finding makes sense in terms of preparing students for the document-based question and the stress on higher order thinking. Similarly, for AP Biology, the emphasis on learning scientific methods proved significant, although only for MCR3. Learning the scientific method also stresses higher order thinking and is the basis for conducting biology labs. Although there were no assessment practices in common in the models for AP U.S. History teachers, one assessment practice was included in two of the AP Biology models: tests requiring lengthy written responses. This may be the technique that more successful teachers use to prepare their students for the free-response portion of the AP Biology Exam.

The models for public school teachers had similarities across both content areas: frequency of class meetings, percentage of students who take the exam, and teacher's estimate of the time students dedicate to studying on their own. Schedule refers to the fact that students in schools where class meet every day performed better than those with block or compressed schedules. We see that the percentage of students who take the exam is significant in all models; this finding indicates that classes where more students took the exam tended to perform better than those with fewer students taking the exam. The last variable indicates that teachers who indicated that their students dedicated large amounts of time to preparing for the course tended to having better performing classes than teachers who estimated a low amount of hours dedicated by their students to preparing for the AP class.

As seen in Table 10, the public school analysis also indicates that there seems to be more similarities within the subject areas from the MCR3 and MCR4 models, both for AP Biology and AP U.S. History. That is, most of the practices that appear in the MCR3 model are also included in the MCR4 model. In addition, for AP Biology, the number of teaching practices included in the MCR4 model exceeds that of the MCR3 model. This may indicate that teaching practices have a stronger effect when the criterion for achievement is more stringent. The fact that this does not happen in AP U.S. History may be related to the fact that these models include fewer teaching-practice variables than the AP Biology models. The most notable differences are that AP U.S. History models do not include any assessment-related variables or variables about teacher preparation time. This may reflect the fact that while the majority of AP U.S. History teachers use more elaborated assessments, such as notebooks or questions that require long answers, this is a less common and more differentiated practice among AP Biology teachers. In addition, it may be that because of the added difficulty of labs, prep time makes more of a difference in Biology than in U.S. History classes.

As seen in Table 11, a notable finding is that for AP U.S. History MCR3 models, there is no significant teaching-practice variable. It is also interesting to see that there are no variables significant across all models. The only variable that is significant in three of the context models is class size. It is still the case, though, that professional development and influence of resources are the most common variables in the context model. Similarly, assessment, objective, and student activities are the most common variables in the teaching-practices model (again, assessment only shows up in the AP Biology models). However, in the models that include public and nonpublic schools, feedback is a significant teaching practice, whereas this practice was not significant in the public school-only model analysis. It may be that teachers in nonpublic schools have more time to dedicate to providing elaborate feedback to students and thus that variability allows for differences in the effect that variable has on performance.

In models with the more stringent success criterion (MCR4), the influence of resources was evident. For AP Biology, these variables were use of exemplary syllabi and AP Exam topics or scoring rubrics, while for U.S. History teachers, discussions with colleagues and mentors proved significant. There are many more differences than similarities across these models; however, each provides a unique view of the relationship between teacher practices and student performance. Since there were reasonably plausible explanations of AP performance in the models that include both context-for-teaching and teacher-practice variables, it seems reasonable to conclude that we have identified some variables that are likely to differentiate more and less successful teachers in both subject areas for public and nonpublic school teachers, and for different criteria of success on the AP Exams. Although it is beyond the scope of this report, we also conducted some preliminary analyses looking at prediction models after dividing schools into three categories based on the school's mean PSAT/NMSQT score. We found several variables significant specifically for those classes in schools with low mean PSAT/NMSQT scores. These variables were different for AP Biology than for AP U.S. History, but were mainly context-for-teaching related. The full set of analyses is suggestive of possible differences to be discovered when taking account of the interaction between student performance and overall school context.

Although we targeted a nationally representative sample of AP teachers and students, we obtained a nonrandom sample, and thus our inferences to the general population must be made tentatively. However, the methodology, as well as the results, should still be considered of interest: (a) We have developed an instrument that documents teachers' self-report of their work environment and practices; (b) we developed an approach to link such information with student performance; and (c) we identified context-for-teaching characteristics and teacher practices that accounted for some of the variation in student performance in two subjects, employing two different criteria.

At the same time, this study reveals some of the challenges in linking teacher practices to student achievement: Obtaining representative samples of teachers, accurately characterizing the teaching context and teacher practice, condensing that information into a form suitable for quantitative analysis, developing suitable criteria related to student learning, and dealing with collinearity among predictors.

There are a number of directions in which this research may be extended. These include (a) validating survey responses with actual observed practice, (b) employing different statistical approaches, such as hierarchical linear modeling or structured equation modeling, and (c) cross-validating results. Ultimately, one hopes that it will be possible to identify teacher practices that are consistently associated with elevated levels of achievement. Although observational surveys can never serve as the basis for causal inferences, they can generate well-grounded hypotheses that can be examined through well-designed experimental interventions.

References

Bain, H. P., Lintz, N., & Word, E. (1989, March). *Study of fifty effective teachers whose class average gain scores ranked in the top 15% of each of four school types in Project STAR*. Paper presented at the 1989 annual meeting of the American Educational Research Association, San Francisco, CA.

Beatty, A. (Ed.). (1997). *Learning from TIMSS: Results of the Third International Mathematics and Science Study, summary of a symposium*. Washington, DC: National Academy Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications, Inc.

Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, R., Mirocha, J., & Guitton, G. (1995). *Validating national curriculum indicators*. Santa Monica, CA: RAND Corporation.

Camara, W. J., & Millsap, R. (1998). *Using the PSAT/NMSQT and course grades in predicting success in the Advanced Placement Program*. New York: The College Board.

Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification* (College Board Report No. 99-05). New York: The College Board.

Chaney, B. (1994). *The accuracy of teachers' self-reports on their postsecondary education: Teacher transcript study, schools and staffing survey* (Working Paper Series Rep. No. 94-04). Washington, DC: National Center for Education Statistics.

Clare, L., & Aschbacher, P. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment, 7*, 39–59.

Cohen, D. K., & Hill, H. C. (1998). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, *102*(2), 294–343

Cohen, J. (1969). *Statistical power analyses for the behavioral sciences*. New York: Academic Press.

Consortium for Policy Research in Education. (1999). *A close look at effects on classroom practice and student performance: A report on the fifth year of the Merck Institute for Science Education*. Philadelphia, PA: Author.

Consortium for Policy Research in Education. (2000). *Deepening the work: A report on the sixth year of the Merck Institute for Science Education*. Philadelphia : Author.

Consortium for Policy Research in Education. (2001). *Steady work: A report on the seventh year of the Merck Institute for Science Education*, 1999–2000. Philadelphia: Author.

Consortium for Policy Research in Education. (2002). *A report on the eighth year of the Merck Institute for Science Education*, 2000–2001. Philadelphia: Author.

Corcoran, T. (2003). *The Merck Institute for Science Education: A successful intermediary for education reform*. Philadelphia: Consortium for Policy Research in Education.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, *103*(6), 1013–55.

Finn, J. D. (2002). *School reform proposals: The research evidence*. Tempe, AZ: Arizona State University, Education Policy Studies Laboratory.

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin, 70*, 245–51.

Haertel, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice, 18*(4), 5–9.

Halderson, J. S., & Glasnapp, D. R. (1972). Generalized rules for calculating the magnitude of effect in factorial and repeated measures ANOVA designs. *American Educational Research Journal*, *9*(2), 301–10.

Herman, R. (1999). *An educators' guide to schoolwide reform*. Washington, DC: American Institutes for Research and Educational Research Service.

Kennedy, J. J. (1970). The eta coefficient in ANOVA designs. *Educational and Psychological Measurement, 30*, 885–89.

Kennedy, M. M. (1998). *Form and substance in in-service teacher education* (Research Monograph No. 13). Arlington, VA: National Science Foundation.

Klein, S. P., Hamilton, L., McCaffrey, D., Stecher, B., Robyn, A., & Burroughs, D. (2000). *Teaching practices and student achievement: Report of first-year findings from the "Mosaic" study of systemic initiatives in mathematics and science*. Santa Monica, CA: RAND Corporation.

Krueger, A. B., & Whitmore, D. M. (1999, October). The effect of attending a small class in the early grades on college test-taking and middle school test results: Evidence from Project STAR. *Economic Journal, 111*(468), 1–28.

Kuppermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Education Evaluation and Policy Analysis, 25*(3), 297–98.

Leighton, M. S., Mullens, J. E., Turnbull, B., Weiner, L. K., & Williams, A. S. (1995). M*easuring instruction, curriculum content, and instructional resources: The status of recent work* (Working Paper Series Rep. No. 95-11). Washington, DC: National Center for Education Statistics.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*, 29–45.

Mullens, J. E., Gayler, K., Goldstein, D., Hildreth, J., Rubenstein, M., Spiggle T., et al. (1999). *Measuring classroom instructional processes: Using survey and case study fieldtest results to improve item construction* (Working Paper Series Rep. No. 1999-08). Washington, DC: National Center for Education Statistics.

Olson, L. (2003). The great divide. *Education Week, 22*, 9–10, 13–14, 16, 18.

Paek, P. , Ponte, E., Sigel, I., Braun, H., & Powers, D. (2005). *A portrait of AP teacher practices* (College Board Research Report 2005-7). New York: The College Board.

Robitaille, D. F., & Garden, R. A. (1996). *Research questions and study design* (TIMSS Monograph No. 2). Vancouver, BC: Pacific Educational Press.

Rowan, B., Correnti, R., & Miller, R. J. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospect study of elementary schools* (CPRE Research Report No. RR-051). Philadelphia: Consortium for Policy Research in Education.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (Research Progress Report). Knoxville, TN: University of Tennessee, Value-Added Research and Assessment Center.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.

Smithson, J. L., & Porter, A. C. (1994). *Measuring classroom practice: Lessons learned from efforts to describe the enacted curriculum—The Reform Up Close study* (CPRE Research Rep. No. RR-031). New Brunswick, NJ: Consortium for Policy Research in Education.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*(12) Retrieved May 2, 2005, from http://epaa.asu.edu/epaa/v10n12/.

Williams, T., Levine, D., Jocelyn, L., Butler, P., Heid, C., & Haynes, J. (2000). M*athematics and science in the eighth grade: Findings from the Third International Mathematics and Science Study* (Statistical Analysis Rep. No. NCES 2000-014). Washington, DC: National Center for Education Statistics.

Word, E., Johnston, J., Bain, H. P., Fulton, B. D., Zaharias, J. B., Achilles, D. M., et al. (1990). *The state of Tennessee's student/teacher achievement ratio (STAR) project: Final summary report 1985–1990*. Nashville, TN: Tennessee State University, Center of Excellence for Research in Basic Skills.

Wuensch, K. L. (2006). *Two-way orthogonal independent samples ANOVA: Computations*. Retrieved June 21, 2007, from http://core.ecu.edu/psyc/wuenschk/StatsLessons.htm, document name is Factorial-Computations.doc.

Zwick, R. (2001, October 29). Making the grade: The SAT vs. the GPA. *San Francisco Chronicle*, p. A17.

Appendix A: Type III Tables for AP Biology

Table A1

Type III: AP Biology Public School Teachers, MCR3

Source	*df*	*Type III SS*	*Mean square*	F	Pr > F	η^2
Five-year mean PSAT/NMSQT score (Math and Verbal)	1	1.0556	1.0556	37.58	< 0.0001	0.0627
Overall experience	2	0.2703	0.1352	4.81	0.0086	0.0169
PAPPD: Attended an AP Institute	2	0.3462	0.1731	6.16	0.0023	0.0215
PAPPD: Review of AP Biology Teacher's Guide	2	0.2597	0.1299	4.62	0.0103	0.0162
PAPPD: Review of AP Course Description: Biology	2	0.1735	0.0867	3.09	0.0467	0.0109
Influence of resources: AP Exam topics and/or scoring rubrics	4	0.2745	0.0686	2.44	0.0462	0.0171
Influence of resources: Schedule	4	0.3772	0.0943	3.36	0.0101	0.0234
Percentage of students who take the exam	3	0.5147	0.1716	6.11	0.0005	0.0316
Class size	3	0.3559	0.1186	4.22	0.0059	0.0221
Objective: Learn scientific methods	4	0.2661	0.0665	2.37	0.0521	0.0166
Frequency of assessment: Tests requiring lengthy written responses	4	0.2572	0.0643	2.29	0.0592	0.0160

Table A2

Type III: AP Biology Public School Teachers, MCR4

Source	*df*	*Type III SS*	*Mean square*	F	Pr > F	η^2
Five-year mean PSAT/NMSQT score (Math and Verbal)	1	0.7860	0.7860	31.37	< 0.0001	0.0496
Influence of resources: Exemplary syllabi from other AP Biology classes	4	0.3357	0.0839	3.35	0.0103	0.0218
Influence of resources: AP Exam topics and/or scoring rubrics	4	0.4744	0.1186	4.73	0.0010	0.0305
Influence of resources: Schedule	4	0.4908	0.1227	4.90	0.0007	0.0315
Percentage of students who take the exam	3	0.7416	0.2472	9.87	< 0.0001	0.0469
Teach test-taking strategies	4	0.4215	0.1053	4.21	0.0024	0.0272
Computer use: Teacher researching information on the Internet	1	0.1956	0.1956	7.81	0.0055	0.0128
Focus on multiple-choice, free-response, or both portions of the test	2	0.1753	0.0876	3.50	0.0311	0.0115
Percentage of class time dedicated to prepare for the AP Exam during the month prior to the AP Exam	4	0.2295	0.0573	2.29	0.0591	0.0150
Type and frequency of review activities: Teacher estimate of student time dedicated to study course material on their own	4	0.3815	0.0953	3.81	0.0047	0.0247

Table A3

Type III: AP Biology Public and Nonpublic School Teachers, MCR3

Source	*df*	*Type III SS*	*Mean square*	F	Pr > F	η^2
Five-year mean PSAT/NMSQT score (Math and Verbal)	1	1.7409	1.7409	58.83	< 0.0001	0.0836
School type	1	0.0080	0.0080	0.27	0.6018	0.0004
Class size	3	0.6203	0.2067	6.99	0.0001	0.0315
Objective: Learn scientific methods	4	0.3746	0.0936	3.17	0.0137	0.0193
Assessment: Tests requiring lengthy written responses	4	0.3603	0.0900	3.04	0.0168	0.0185
Student activities: Submit reports on experiments or observations	4	0.3173	0.0793	2.68	0.0309	0.0164
Computer use: Student researching information on the Internet	1	0.1107	0.1107	3.74	0.0535	0.0058
School type and adequacy of facilities: Laboratories, lab tables, sinks, etc.	3	0.2053	0.0684	2.31	0.0750	0.0106
School type and objective: Learn scientific methods	4	0.2773	0.0693	2.34	0.0538	0.0143

Note: When "school type" is included, the results show the interaction of school type with a context for teaching or teacher practice variable.

Table A4

Type III: AP Biology Public and Nonpublic School Teachers, MCR4

Source	*df*	*Type III SS*	*Mean square*	F	Pr > F	η^2
Five-year mean PSAT/NMSQT score (Math and Verbal)	1	1.1443	1.1443	48.06	<0.0001	0.0566
School type	1	0.0148	0.0148	0.62	0.4306	0.0008
Objective: Communicate biological concepts effectively	3	0.2218	0.0739	3.11	0.0263	0.0115
Assessment: Laboratory notebooks or journals	4	0.2716	0.0679	2.85	0.0234	0.0140
Assessment: Presentations by students	4	0.2214	0.0553	2.32	0.0556	0.0115
Feedback: Phrase or sentence-length descriptions of their performance	4	0.3028	0.0757	3.18	0.0135	0.0156
Computer use: Student researching information on the Internet	1	0.1314	0.1314	5.52	0.0192	0.0068
Computer use: Simulation and modeling	1	0.1277	0.1277	5.36	0.0210	0.0067
PAPPD: Attended an AP workshop	2	0.2243	0.1121	4.71	0.0094	0.0116
Influence of resources: Exemplary syllabi from other AP Biology classes	4	0.4853	0.1213	5.10	0.0005	0.0248
Influence of resources: AP Exam topics and/or scoring rubrics	4	0.3167	0.0791	3.33	0.0106	0.0163
Number of AP Biology classes taught	4	0.3854	0.0963	4.05	0.0031	0.0198
Influence of resources: Schedule	4	0.4363	0.1090	4.58	0.0012	0.0224
Percentage of students who take the exam	3	0.2376	0.0792	3.33	0.0195	0.0123
Class size	3	0.3678	0.1226	5.15	0.0016	0.0189
School type and assessment: Laboratory notebooks or journals	4	0.2496	0.0624	2.62	0.0343	0.0129
School type and PAPPD: Attended an AP workshop	2	0.1460	0.0730	3.07	0.0475	0.0076
School type and PAPPD: Taught in an AP Institute	1	0.1550	0.1550	6.51	0.0110	0.0081
School type and number of AP Biology classes taught	3	0.2572	0.0857	3.60	0.0135	0.0133
School type and percentage of students who take the exam	3	0.2230	0.0743	3.12	0.0257	0.0116
School type and class size	2	0.2263	0.1131	4.75	0.0090	0.0117

Note: When "school type" is included, the results show the interaction of school type with a context for teaching or teacher practice variable.

Appendix B: Type III Tables for AP U.S. History

Table B1

Type III: AP U.S. History Public School Teachers, MCR3						
Source	***df***	***Type III SS***	***Mean square***	**F**	**Pr > F**	η^2
Five-year mean PSAT/NMSQT score (Verbal)	1	0.8903	0.8903	31.09	< 0.0001	0.0436
PAPPD: Collaborated with mentor teacher	2	0.1436	0.0718	2.51	0.0825	0.0073
PAPPD: Review of released AP Exams	2	0.3009	0.1504	5.25	0.0055	0.0152
PAPPD: Participated in AP Reading(s)	2	0.1931	0.0965	3.37	0.0351	0.0098
Influence of resources: Schedule	4	0.3824	0.0956	3.34	0.0104	0.0192
Adequacy of school resources	3	0.6458	0.2152	7.52	< 0.0001	0.0320
School policies regarding who takes the exam	3	0.2104	0.0701	2.45	0.0630	0.0107
Percentage of students who take the exam	3	0.5534	0.1844	6.44	0.0003	0.0276
Objective: Develop skills for supporting stating and claims	3	0.3606	0.1202	4.20	0.0060	0.0181
Objective: Develop historical research skills and techniques	4	0.3031	0.0757	2.65	0.0330	0.0153
Student activities: Explain reasoning or thinking	4	0.2546	0.0636	2.22	0.0656	0.0129
Type and frequency of review activities: Teacher estimate of student time dedicated to studying course material on their own	4	0.4299	0.1074	3.75	0.0051	0.0216

Table B2

Type III: AP U.S. History Public School Teachers, MCR4						
Source	***df***	***Type III SS***	***Mean square***	**F**	**Pr > F**	η^2
Five-year mean PSAT/NMSQT score (Verbal)	1	0.6750	0.6750	28.21	< 0.0001	0.0400
Educational level attained	4	0.2411	0.0602	2.52	0.0405	0.0147
Influence of resources: Schedule	4	0.1991	0.0497	2.08	0.0822	0.0121
Adequacy of school resources	3	0.3134	0.1044	4.37	0.0048	0.0190
Percentage of students who take the exam	3	0.8454	0.2818	11.78	< 0.0001	0.0496
Control of selecting content, topics, and skills to be taught	3	0.1536	0.0512	2.14	0.0944	0.0094
Objective: Develop historical research skills and techniques	4	0.2387	0.0596	2.49	0.0422	0.0145
Student activities: Explain reasoning or thinking	4	0.2909	0.0727	3.04	0.0171	0.0176
Type and frequency of review activities: teacher estimate of student time dedicated to studying course material on their own	4	0.3566	0.0891	3.73	0.0053	0.0215

Table B3

Type III: AP U.S. History Public and Nonpublic School Teachers, MCR3

Source	*df*	*Type III SS*	*Mean square*	F	Pr > F	η^2
Five-year mean PSAT/NMSQT score (Verbal)	1	1.7548	1.7548	51.57	< 0.0001	0.0585
School type	1	0.1160	0.1160	3.41	0.0652	0.0041
AP teaching experience	2	0.2805	0.1402	4.12	0.0166	0.0098
Influence of resources: Exemplary syllabi from other AP U.S. History classes	4	0.2440	0.0610	1.79	0.1285	0.0086
Influence of resources: Discussions with colleagues and mentors	4	0.2934	0.0733	2.16	0.0724	0.0103
Adequacy of school resources	3	0.7577	0.2525	7.42	< 0.0001	0.0261
School type and Influence of resources: Discussions with colleagues and mentors	4	0.2905	0.0726	2.13	0.0750	0.0102

Note: When "school type" is included, the results show the interaction of school type with a context for teaching or teacher practice variable.

Table B4

Type III: AP U.S. History Public and Nonpublic School Teachers, MCR4

Source	*df*	*Type III SS*	*Mean square*	F	Pr > F	η^2
Five-year mean PSAT/NMSQT score (Verbal)	1	0.6108	0.6108	24.16	< 0.0001	0.0274
School type	1	0.0130	0.0130	0.52	0.4723	0.0006
PAPPD: Attended an AP Institute	2	0.1191	0.0595	2.36	0.0958	0.0055
PAPPD: Review of released AP Exams	2	0.1349	0.0674	2.67	0.0703	0.0062
PAPPD: Consulted for an AP workshop	2	0.3376	0.1688	6.67	0.0014	0.0153
Adequacy of school resources	3	0.4916	0.1638	6.48	0.0003	0.0222
Percentage of students who take the exam	3	0.7517	0.2505	9.91	< 0.0001	0.0335
Class size	3	0.2699	0.0899	3.56	0.0142	0.0123
Student activities: Use tools of analysis to generate hypotheses or develop arguments	4	0.3784	0.0946	3.74	0.0051	0.0171
Student activities: Discuss controversial events or themes	4	0.5601	0.1400	5.54	0.0002	0.0252
Hours of preparation for AP U.S. History class per week	3	0.2167	0.0722	2.86	0.0365	0.0099
School type and PAPPD: Attended an AP Institute	2	0.3463	0.1731	6.85	0.0012	0.0157
School type and PAPPD: Consulted for an AP workshop	2	0.2577	0.1288	5.10	0.0064	0.0117
School type and Influence of resources: Discussions with colleagues and mentors	4	0.2641	0.0660	2.61	0.0347	0.0120
School type and Objective: Develop historical research skills and techniques	4	0.2642	0.0660	2.61	0.0346	0.0120
School type and Provide instruction to small groups of students	4	0.2507	0.0626	2.48	0.0431	0.0114
School type and Student activities: Discuss controversial events or themes	4	0.5641	0.1410	5.58	0.0002	0.0253
School type and Hours of preparation for AP U.S. History class per week	3	0.3256	0.1085	4.29	0.0052	0.0148

Note: When "school type" is included, the results show the interaction of school type with a context for teaching or teacher practice variable.

Appendix C: Survey of AP Biology Teachers

Section 1—Instructional and Assessment Practices

	Please circle one number for each item				
1. In comparison to the other objectives listed below, how much emphasis do you place on each of the following for AP Biology? Helping students:	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
a. Learn facts and terminology	1	2	3	4	5
b. Understand key concepts	1	2	3	4	5
c. Learn scientific methods	1	2	3	4	5
d. Develop scientific reasoning skills	1	2	3	4	5
e. Communicate biological concepts effectively	1	2	3	4	5
f. Develop interest in biology	1	2	3	4	5

2. How often do you do each of the following with your AP Biology students?	***Hardly ever***	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Lecture	1	2	3	4	5
b. Teacher-led whole-group discussions	1	2	3	4	5
c. Provide instruction to small groups of students	1	2	3	4	5
d. Provide instruction to individual students	1	2	3	4	5
e. Provide summaries of key concepts to accompany class notes	1	2	3	4	5
f. Teach test-taking strategies	1	2	3	4	5
g. Make group assignments	1	2	3	4	5
h. Use additional materials (e.g., films) to illustrate a biological theory/concept	1	2	3	4	5

3. How often do you do use the following kinds of assessments with your AP Biology students?	***Hardly ever***	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Multiple-choice tests	1	2	3	4	5
b. Tests requiring sentence- or paragraph-length responses	1	2	3	4	5
c. Tests requiring lengthy written responses	1	2	3	4	5
d. Laboratory notebooks or journals	1	2	3	4	5
e. Presentations by students	1	2	3	4	5
f. Independent research/projects by students	1	2	3	4	5

4. **How often do students receive each of the following kinds of feedback on their tests or assignments for your AP classes?**	**Please circle one number for each item**				
	Hardly ever	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Numerical or letter grades	1	2	3	4	5
b. Phrase- or sentence-length descriptions of their performance	1	2	3	4	5
c. Paragraph-length descriptions of strengths and weaknesses	1	2	3	4	5
d. Page-length descriptions of strengths and weaknesses	1	2	3	4	5
e. Discussion of areas needing improvement	1	2	3	4	5
f. Comparison of performance with that of the class as a whole	1	2	3	4	5

5. ***How often do you do each of the following with your AP Biology students?***	**Please circle one number for each item**				
	Hardly ever	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Conduct an experiment	1	2	3	4	5
b. Lead other students in systematic observations	1	2	3	4	5
c. Submit reports on experiments or observations	1	2	3	4	5
d. Discuss current issues and events related to AP Biology	1	2	3	4	5
e. Design and conduct their own science projects	1	2	3	4	5
f. Discuss controversial theories and innovations	1	2	3	4	5
g. Participate in various competitions (e.g., science fairs)	1	2	3	4	5
h. Work on biology exercises or problems	1	2	3	4	5
i. Explain reasoning or thinking	1	2	3	4	5
j. Apply biology concepts to simulated or real-world problems	1	2	3	4	5

6. ***Are computers used in your AP Biology class(es) in any of the following ways?***	***No***	***Yes***
a. Researching information on the Internet (students)	1	2
b. Researching information on the Internet (teacher)	1	2
c. Simulation and modeling	1	2
d. Data analysis	1	2

7. On average, how many hours per week do you spend preparing for your AP Biology class(es)?
   1. 0–3 hours per week
   2. 4–9 hours per week
   3. 10–15 hours per week
   4. More than 15 hours per week

8. About how many hours each week do you expect a student to spend doing AP Biology homework (including assigned reading)?
   1. Less than 5 hours per week
   2. 5–10 hours per week
   3. More than 10 hours per week

Section 2—Content Coverage

9. In teaching AP Biology, would you rather…
   1. cover each potential topic on the examination, even if only briefly, or
   2. cover some topics very thoroughly, even if this means not covering certain topics at all?

	Please circle one number for each item				
10. In comparison to the other topics listed below, how much emphasis do you place on each of the following in your AP Biology class(es)?	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
a. Chemistry of Life (e.g., water, enzymes)	1	2	3	4	5
b. Cells (e.g., membranes, subcellular organization)	1	2	3	4	5
c. Cellular Energy Processes (e.g., photosynthesis)	1	2	3	4	5
d. Heredity (e.g., inheritance patterns)	1	2	3	4	5
e. Molecular Genetics (e.g., gene regulation)	1	2	3	4	5
f. Evolutionary Biology (e.g., evidence of evolution)	1	2	3	4	5
g. Diversity of Organisms (e.g., phylogenetic classification)	1	2	3	4	5
h. Structure and Function of Plants and Animals (e.g., response to the environment)	1	2	3	4	5
i. Ecology (e.g., population dynamics)	1	2	3	4	5

11. In question 10 (above), which are the most difficult topics for students to learn? Write the letter of the topic in the spaces below.

___ most difficult

___ second most difficult

___ third most difficult

12. In comparison to the other themes listed below, how much emphasis do you place on each of the following in your AP Biology class(es)?	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
a. Science as a Process	1	2	3	4	5
b. Evolution	1	2	3	4	5
c. Energy Transfer	1	2	3	4	5
d. Continuity and Change	1	2	3	4	5
e. Relationship Between Structure and Function	1	2	3	4	5
f. Regulation	1	2	3	4	5
g. Interdependence in Nature	1	2	3	4	5
h. Science, Technology, and Society	1	2	3	4	5
i. Ecology (e.g., population dynamics)	1	2	3	4	5

Section 3—Test-Specific Instructional Activities and Practices

13. When preparing students for the AP Biology Examination, do you typically focus attention…
    1. more on the free-response portion of the examination?
    2. more on the multiple-choice portion of the examination?
    3. about equally on both portions of the examination?

14. About what proportion of classroom time is directly related to helping students pass the AP Exam (e.g., reviewing AP Biology practice exams) …	**Please circle one number for each item**				
	Less than 20%	***21–40%***	***41–60%***	***61–80%***	***More than 80%***
a. throughout the school year?	1	2	3	4	5
b. in the month before the AP Exam?	1	2	3	4	5

15. In the month before the AP Exam, how many hours per week do you…	***None***	***Less than 4 hours***	***4–9 hours***	***10–20 hours***	***More than 20 hours***
a. review material for the AP Exam after school?	1	2	3	4	5
b. administer or help students review old AP Exams?	1	2	3	4	5
c. think most students participate in student-led study groups outside of class time without the teacher?	1	2	3	4	5
d. think most students spend studying course material on their own, including practice tests?	1	2	3	4	5

Section 4—School Context

16. How were you assigned to teach AP Biology? Circle only one number.
    1. It was assigned to me.
    2. I volunteered to teach it.

17. How many AP Biology classes are you teaching this year?
   1. One
   2. Two
   3. Three
   4. Four
   5. Five or more

18. Which schedule option best describes the AP course you are teaching in the 2002-03 academic year?
   1. A 30–60 minute session every school day throughout the year
   2. A 61–110 minute session every school day throughout the year
   3. A 61–110 minute session every other school day throughout the school year
   4. The complete course compressed in the fall 2002 semester (with or without review in spring 2003)
   5. The complete course compressed into the spring 2003 semester

19. Indicate the adequacy of the following resources for AP Biology at your school.	**Please circle one number for each item**			
	Poor	***Fair***	***Good***	***Excellent***
a. The facilities (laboratories, lab tables, sinks, etc.)	1	2	3	4
b. The supplies (specimens, etc.)	1	2	3	4
c. Instructional materials or other resources	1	2	3	4

20. To what extent do the following practices describe the situation in your school?	***Strongly Disagree***	***Disagree***	***Agree***	***Strongly Agree***
a. I am encouraged to experiment with my teaching.	1	2	3	4
b. I have a wide degree of autonomy in selecting course content.	1	2	3	4
c. I am encouraged to coordinate the content of my courses with other teachers in my department.	1	2	3	4
d. There is a strong commitment to AP courses in my department.	1	2	3	4

21. Does your school have any special procedures or criteria for enrollment for AP Biology class(es)?
   1. No, enrollment is completely open → (skip to question 22)
   2. Yes (continue to item 21a)

21a. If you answered "yes" above, please indicate the degree to which each of the following is a factor in deciding student enrollment in your AP Biology class(es).	***Not a factor***	***A minor factor***	***A major factor***
a. Completion of a prerequisite course (such as Honors Biology).	1	2	3
b. Achievement of required grades in prior course(s)	1	2	3
c. Recommendation by teachers	1	2	3
d. Earning a qualifying score on PSAT/NMSQT (or other test)	1	2	3
e. Meeting requirements of school-designed admission policy	1	2	3
f. Self-nomination	1	2	3
g. Recommendation by parent or guardian	1	2	3
h. Recommendation by guidance counselor/school administrator	1	2	3
i. Entering through vertical teaming	1	2	3

22. Are there initiatives at your school to increase the enrollment of minority students in AP Biology (or other AP classes)?

1. No
2. Yes

Yes →	No →
Our school employs the following initiatives: Please mark all that apply. ☐ Recruitment by teachers ☐ Meetings with parents ☐ Special mailings or communications ☐ Recruitment by guidance counselor	No initiatives exist because… Please mark all that apply. ☐ Most students in this school are minority students ☐ We have few, if any, minority students in this school ☐ Minority enrollment in AP classes is sufficient already

23. Which best describes students who take the AP Biology Examination at your school? Circle only one response.

1. All students who take the course must also take the AP Exam.
2. Only those students who do well in the course are encouraged to take the AP Exam.
3. All students who take the course are encouraged to take the AP Exam.
4. Students who take the course are left to decide whether to take the AP Exam.

24. On average, what percentage of students in your AP Biology class(es) takes the AP Biology examination?

1. Less than 50% of students
2. Between 51–74% of students
3. Between 75–99% of students
4. 100% of students

Section 5—Classroom Context

	Please circle one number for each item			
25. How much control do you feel you have in your AP Biology class(es) in selecting each of the following ?	***Little or no control***	***Some control***	***Substantial control***	***Complete control***
a. Textbook(s)	1	2	3	4
b. Supplemental instructional materials	1	2	3	4
c. Content, topics, and skills to be taught	1	2	3	4
d. Teaching techniques	1	2	3	4

26. What is the average class size (number of students) in your AP Biology class(es) this year?

1. Fewer than 15 students
2. 16–20 students
3. 21–30 students
4. More than 30 students

Section 6—Your Professional Development Experiences and Training

	Please circle one number for each item		
27. In what AP professional development activities have you participated within the last five years?	***No***	***Yes, once***	***Yes, more than once***
a. Attended an AP workshop (1–2 day events)	1	2	3
b. Attended an AP Institute (week, summer)	1	2	3
c. Collaborated with mentor teacher	1	2	3
d. Reviewed released AP Exams	1	2	3
e. Reviewed AP Biology Teacher's Guide	1	2	3
f. Reviewed AP Course Description: Biology	1	2	3
g. Took college-level course in Biology or other related subject	1	2	3
h. Networked with AP Biology teachers at different schools	1	2	3
i. Participated in AP Reading(s)	1	2	3
j. Consulted for an AP workshop (event for 1–2 days)	1	2	3
k. Taught in an AP Institute (event for 1 week or longer)	1	2	3

28. How much influence has each of the following resources had on your teaching of AP Biology?	***Not at all influential***	***Slightly influential***	***Somewhat influential***	***Very influential***	***Extremely influential***
a. Exemplary syllabi from other AP Biology classes	1	2	3	4	5
b. AP Exam essay topics and/or scoring rubrics	1	2	3	4	5
c. Supplementary texts, workbooks, etc.	1	2	3	4	5
d. Biology Lab Manual and/or Teacher's Version of the Biology Lab Manual	1	2	3	4	5
e. Discussions with colleagues and mentors	1	2	3	4	5
f. Teaching resources through the Internet (e.g., sample lessons, readings, etc.)	1	2	3	4	5
g. Conversations through the Internet about teaching and learning	1	2	3	4	5

29. Please tell us about the general areas in which you have the need for further education/training in AP Biology.	**Please circle one number for each item**			
	Not an important training need	***Somewhat important training need***	***Important training need***	***Critical training need***
a. Understanding specific areas of course content	1	2	3	4
b. Developing specific skills (e.g., analytical writing, advanced problem solving, using the computer)	1	2	3	4
c. Learning alternative methods for presenting content or developing skills	1	2	3	4
d. Understanding the concepts behind the AP syllabus topics/labs and having alternative instructional strategies for teaching them	1	2	3	4
e. Communicating the AP content and target skills to students with different levels of preparation	1	2	3	4
f. Preparing students for the AP Exam	1	2	3	4
g. Covering the course content in the time available: What can be dropped or modified?	1	2	3	4
h. Accurately assessing student performance and proficiency levels during the AP course	1	2	3	4
i. Integrating new technologies into my AP teaching	1	2	3	4

Section 7—About You

30. Counting this school year, how long have you been teaching?
    1. 0–3 years
    2. 4–6 years
    3. 7–10 years
    4. More than 10 years

31. Counting this school year, how many years have you been teaching AP Biology?
    1. 0–3 years
    2. 4–6 years
    3. 7–10 years
    4. More than 10 years

32. In which academic (school) years did you teach AP U.S. Biology? (Check all that apply)
    - ☐ 1998-99
    - ☐ 1999–2000
    - ☐ 2000-01
    - ☐ 2001-02
    - ☐ 2002-03

33. What is the highest level of education you have attained?
    1. Bachelor's degree
    2. Bachelor's degree plus further credits
    3. Master's degree
    4. Master's degree plus further credits
    5 Doctorate or professional degree (e.g., Ph.D., Ed.D., J.D., M.D.)

34. What was/were your college major(s)?
    1. Biology
    2. Other science (e.g., physiology, chemistry)
    3. Other

35. What type of teaching certificate do you have? Please mark only one (your highest certification).
    1. I don't have a certificate.
    2. Regular or standard state certificate offered in the state.
    3. Advanced professional certificate (e.g., National Board for Professional Teaching Standards).

36. What is your age?
    1. 25 or under
    2. 26–35
    3. 36–45
    4. 46–55
    5. 56–65
    6. 66 or older

37. What is your ethnicity?
    1. African American or Black
    2. American Indian/Native American
    3. Asian American/Asian Indian/Pacific Islander
    4. Caucasian (non-Hispanic)
    5. Latino, Latin American, Puerto Rican, Hispanic, Chicano

38. Are you…
    1. Male
    2. Female

39. Do you have any comments for us regarding your experience as an AP Biology teacher? Is there anything you do as an AP Biology teacher that you feel is especially noteworthy? Please use the space below.

Appendix D: Survey of AP U.S. History Teachers

Section 1—Instructional and Assessment Practices

	Please circle one number for each item				
1. In comparison to the other objectives listed below, how much emphasis do you place on each of the following for AP U.S. History? Helping students:	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
a. Learn facts, dates, events, and terminology	1	2	3	4	5
b. Understand themes	1	2	3	4	5
c. View history as multiple perspectives/stories	1	2	3	4	5
d. Develop skills for stating and supporting claims	1	2	3	4	5
e. Develop historical research skills and techniques	1	2	3	4	5
f. Develop interest in U.S. History	1	2	3	4	5

2. How often do you do each of the following with your AP U.S. History students?	***Hardly ever***	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Lecture	1	2	3	4	5
b. Teacher-led whole-group discussions	1	2	3	4	5
c. Provide instruction to small groups of students	1	2	3	4	5
d. Provide instruction to individual students	1	2	3	4	5
e. Provide summaries of key concepts to accompany class notes	1	2	3	4	5
f. Teach test-taking strategies	1	2	3	4	5
g. Make group assignments	1	2	3	4	5
h. Use additional materials (e.g., films or art) to illustrate a historical period	1	2	3	4	5

3. How often do you use the following kinds of assessments with your AP U.S. History students?	***Hardly ever***	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Multiple-choice tests	1	2	3	4	5
b. Tests requiring sentence- or paragraph-length responses	1	2	3	4	5
c. Using document-based evidence to organize an essay (e.g., preparing for the DBQ)	1	2	3	4	5
d. Presentations by students	1	2	3	4	5
e. Independent research/projects by students	1	2	3	4	5

	Please circle one number for each item				
4. How often do students receive each of the following kinds of feedback on their tests or assignments for your AP classes?	***Hardly ever***	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Numerical or letter grades	1	2	3	4	5
b. Phrase or sentence-length descriptions of their performance	1	2	3	4	5
c. Paragraph-length descriptions of strengths and weaknesses	1	2	3	4	5
d. Page-length descriptions of strengths and weaknesses	1	2	3	4	5
e. Discussion of areas needing improvement	1	2	3	4	5
f. Comparison of performance with that of the class as a whole	1	2	3	4	5

5. How often are your AP U.S. History students asked to do each of the following?	***Hardly ever***	***Several times a year***	***Once or twice a month***	***Once or twice a week***	***Almost every class session/ period***
a. Submit thematic essays on specific historical topics (e.g., slavery, suffrage)	1	2	3	4	5
b. Use tools of analysis to generate hypotheses or develop arguments	1	2	3	4	5
c. Discuss current issues and events related to AP U.S. History	1	2	3	4	5
d. Discuss controversial events or themes	1	2	3	4	5
e. Analyze documents or evaluate essays	1	2	3	4	5
f. Participate in various competitions (e.g., reenactments, debates)	1	2	3	4	5
g. Work on history exercises or problems	1	2	3	4	5
h. Explain reasoning or thinking	1	2	3	4	5

6. Are computers used in your AP U.S. History class(es) in any of the following ways?	***No***	***Yes***
a. Researching information on the Internet (by students)	1	2
b. Researching information on the Internet (teacher)	1	2

7. On average, how many hours per week do you spend preparing for your AP U.S. History class(es)?
    1. 0–3 hours per week
    2. 4–9 hours per week
    3. 10–15 hours per week
    4. More than 15 hours per week

8. About how many hours each week do you expect a student to spend doing AP U.S. History homework (including assigned reading)?
   1. Less than 5 hours per week
   2. 5–10 hours per week
   3. More than 10 hours per week

Section 2—Content Coverage

9. In teaching AP U.S. History, would you rather…
   1. cover each potential topic on the examination, even if only briefly, or
   2. cover some topics very thoroughly, even if this means not covering certain topics at all?

	Please circle one number for each item				
10. In comparison to the other topics listed below, how much emphasis do you place on each of the following in your AP U.S. History class(es)?	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
a. Discovery and Settlement of the New World, 1492–1650	1	2	3	4	5
b. America and the British Empire, 1650–1754	1	2	3	4	5
c. Colonial Society in the Mid-Eighteenth Century	1	2	3	4	5
d. The American Revolution, 1775–1783	1	2	3	4	5
e. National and Economic Expansion	1	2	3	4	5
f. Age of Jackson, 1828–1848	1	2	3	4	5
g. Creating an American Culture	1	2	3	4	5
h. Civil War	1	2	3	4	5
i. New South and the Last West	1	2	3	4	5
j. Industrialization and Corporate Consolidation	1	2	3	4	5
k. Intellectual and Cultural Movements	1	2	3	4	5
l. National Politics, 1877–1896: The Gilded Age	1	2	3	4	5
m. The First World War	1	2	3	4	5
n. New Era: The 1920s	1	2	3	4	5
o. Depression: 1929–1933	1	2	3	4	5
p. The Second World War	1	2	3	4	5
q. Truman and the Cold War	1	2	3	4	5
r. Kennedy's New Frontier; Johnson's Great Society	1	2	3	4	5
s. The United States Since 1974	1	2	3	4	5

11. In question 10 (above), which are the most difficult topics for students to learn? Write the letter of the topic in the spaces below.

    ___ most difficult

    ___ second most difficult

    ___ third most difficult

12. In comparison to the other subtopics listed below, how much emphasis do you place on each of the following in your AP U.S. History class(es)?	**Please circle one number for each item**				
America and the British Empire, 1650–1754	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
a. Chesapeake country	1	2	3	4	5
b. Growth of New England	1	2	3	4	5
c. Restoration colonies	1	2	3	4	5
d. Mercantilism; the dominion of New England	1	2	3	4	5
e. Origins of slavery	1	2	3	4	5
Intellectual and Cultural Movements	***Less than average emphasis***	***Slightly less than average emphasis***	***About average emphasis***	***Slightly more than average emphasis***	***More than average emphasis***
f. Education: colleges and universities	1	2	3	4	5
f. Education: scientific advances	1	2	3	4	5
g. Professionalism and the social sciences	1	2	3	4	5
h. Realism in literature and art	1	2	3	4	5
i. Mass culture: use of leisure	1	2	3	4	5
j. Mass culture: publishing and journalism	1	2	3	4	5

Section 3—Test-Specific Instructional Activities and Practices

13. When preparing students for the AP U.S. History Examination, do you typically focus attention…

1. more on the free-response portion of the examination?
2. more on the multiple-choice portion of the examination?
3. about equally on both portions of the examination?

14. About what proportion of classroom time is <u>directly</u> related to helping students pass the AP Exam (e.g., reviewing AP U.S. History practice exams)…	***Less than 20 percent***	***21–40 percent***	***41–60 percent***	***61–80 percent***	***More than 80 percent***
a. throughout the school year?	1	2	3	4	5
b. in the month before the AP Exam?	1	2	3	4	5

	Please circle one number for each item				
15. <u>In the month before the AP Exam</u>, how many hours per week do you…	***None***	***Less than 4 hours***	***4–9 hours***	***10–20 hours***	***More than 20 hours***
a. review material for the AP Exam after school?	1	2	3	4	5
b. administer or help students review old AP Exams?	1	2	3	4	5
c. think most students participate in student-led study groups outside of class time without the teacher?	1	2	3	4	5
d. think most students spend studying course material on their own, including practice tests?	1	2	3	4	5

Section 4—School Context

16. How were you assigned to teach AP U.S. History? Circle only one number.
    1. It was assigned to me.
    2. I volunteered to teach it.

17. How many AP U.S. History classes are you teaching this year?
    1. One
    2. Two
    3. Three
    4. Four
    5. Five or more

18. Which schedule option <u>best</u> describes the AP course you are teaching in the 2002-03 academic year?
    1. A 30–60 minute session <u>every</u> school day throughout the year
    2. A 61–110 minute session <u>every</u> school day throughout the school year
    3. A 61–110 minute session <u>every other</u> school day throughout the school year
    4. The complete course compressed in the fall 2002 semester (with or without review in spring 2003)
    5. The complete course compressed into the spring 2003 semester

19. Please indicate which statement most accurately represents how well your school system provides you with the instructional materials and other resources you need to teach your AP U.S. History class(es)?
    1. I get hardly any of the resources I need.
    2. I get some of the resources I need.
    3. I get most of the resources I need.
    4. I get nearly all of the resources I need.

	Please circle one number for each item			
20. To what extent do the following practices describe the situation in your school?	***Strongly disagree***	***Disagree***	***Agree***	***Strongly agree***
a. I am encouraged to experiment with my teaching.	1	2	3	4
b. I have a wide degree of autonomy in selecting course content.	1	2	3	4
c. I am encouraged to coordinate the content of my courses with other teachers in my department.	1	2	3	4
d. There is a strong commitment to AP courses in my department.	1	2	3	4

21. Does your school have any special procedures or criteria for enrollment for AP U.S. History class(es)?

1. No, enrollment is completely open → (skip to question 22)
2. Yes (continue to item 21a)

21a. If you answered "yes" above, please indicate the degree to which each of the following is a factor in deciding student enrollment in your AP U.S. History class(es).	***Not a factor***	***A minor factor***	***A major factor***
a. Completion of a prerequisite course (such as Honors History)	1	2	3
b. Achievement of required grades in prior course(s)	1	2	3
c. Recommendation by teachers	1	2	3
d. Earning a qualifying score on PSAT/NMSQT (or other test)	1	2	3
e. Meeting requirements of school-designed admissions policy	1	2	3
21b. If you answered "yes" above, please indicate the degree to which each of the following is a factor in deciding student enrollment in your AP U.S. History class(es).	***Not a factor***	***A minor factor***	***A major factor***
f. Self-nomination	1	2	3
g. Recommendation by parent or guardian	1	2	3
h. Recommendation by guidance counselor/school administration	1	2	3
i. Entering through vertical teaming	1	2	3

22. Are there initiatives at your school to increase the enrollment of minority students in AP U.S. History (or other AP classes)?

1. No
2. Yes

If Yes:

Our school employs the following initiatives:

Please mark all that apply.

- ☐ Recruitment by teachers
- ☐ Meetings with parents
- ☐ Special mailings or communications
- ☐ Recruitment by guidance counselor

If No:

No initiatives exist because…

Please mark all that apply.

- ☐ Most students in this school are minority students
- ☐ We have few, if any, minority students in this school
- ☐ Minority enrollment in AP classes is sufficient already

23. Which <u>best</u> describes students who take the AP U.S. History Examination at your school? Circle only one response.

1. All students who take the course <u>must</u> also take the AP Exam.
2. Only those students who do well in the course are <u>encouraged</u> to take the AP Exam.
3. All students who take the course are <u>encouraged</u> to take the AP Exam.
4. Students who take the class are left to decide whether to take the AP Exam.

24. On average, what percentage of students in your AP U.S. History class(es) takes the AP U.S. History Examination?

1. Less than 50% of students
2. Between 51–74% of students
3. Between 75–99% of students
4. 100% of students

Section 5—Classroom Context

	Please circle one number for each item			
25. How much control do you feel you have in your AP U.S. History class(es) in selecting each of the following?	***Little or no control***	***Some control***	***Substantial control***	***Complete control***
a. Textbook(s)	1	2	3	4
b. Supplemental instructional materials	1	2	3	4
c. Content, topics, and skills to be taught	1	2	3	4
d. Teaching techniques	1	2	3	4

26. What is the average class size (number of students) in your AP U.S. History class(es) this year?
    1. Fewer than 15 students
    2. 16–20 students
    3. 21–30 students
    4. More than 30 students

Section 6—Your Professional Development Experiences and Training

27. In what AP professional development activities have you participated within the last five years?	***No***	***Yes, once***	***Yes, more than once***
a. Attended AP workshops (1–2 day events)	1	2	3
b. Attended AP Institute (week, summer)	1	2	3
c. Collaborated with mentor teacher	1	2	3
d. Reviewed released AP Exams	1	2	3
e. Reviewed AP U.S. History Teacher's Guide	1	2	3
f. Reviewed AP Course Description: U.S. History	1	2	3
g. Took college-level course in U.S. History or related course	1	2	3
h. Networked with AP U.S. History teachers at other schools	1	2	3
i. Participated in AP Reading(s)	1	2	3
j. Consulted for an AP workshop (event for 1–2 days)	1	2	3
k. Taught in an AP Institute (event for 1 week or longer)	1	2	3

	Please circle one number for each item				
28. How much influence has each of the following resources had on your teaching of AP U.S. History?	***Not at all influential***	***Slightly influential***	***Somewhat influential***	***Very influential***	***Extremely influential***
a. Exemplary syllabi from other AP U.S. History classes	1	2	3	4	5
b. AP Exam essay topics and/or scoring rubrics	1	2	3	4	5
c. Supplementary texts, workbooks, etc.	1	2	3	4	5
d. Discussions with colleagues and mentors	1	2	3	4	5
e. Teaching resources through the Internet (e.g., sample lessons, readings, etc.)	1	2	3	4	5
f. Communications through the Internet about teaching and learning	1	2	3	4	5

29. Please tell us about the general areas in which you have the need for further education/training in AP U.S. History.	*Not an important training need*	*Somewhat important training need*	*Important training need*	*Critical training need*
a. Understanding specific areas of course content	1	2	3	4
b. Developing specific skills (e.g., analytical writing, using the computer)	1	2	3	4
c. Learning alternative methods for presenting content or developing skills	1	2	3	4
d. Understanding the concepts behind the AP syllabus topics and having alternative instructional strategies for teaching them	1	2	3	4
e. Communicating the AP content and target skills to students with different levels of preparation	1	2	3	4
f. Preparing students for the AP Exam	1	2	3	4
g. Covering the course content in the time available: what can be dropped or modified?	1	2	3	4
h. Accurately assessing student performance and proficiency levels during the AP course	1	2	3	4
i. Integrating new technologies into my AP teaching	1	2	3	4

Section 7—About You

30. Counting this school year, how long have you been teaching?
    1. 0–3 years
    2. 4–6 years
    3. 7–10 years
    4. More than 10 years

31. Counting this school year, how many years have you been teaching AP U.S. History?
    1. 0–3 years
    2. 4–6 years
    3. 7–10 years
    4. More than 10 years

32. In which academic (school) years did you teach AP U.S. History? (Check all that apply)
    - ☐ 1998-99
    - ☐ 1999–2000
    - ☐ 2000-01
    - ☐ 2001-02
    - ☐ 2002-03

33. What is the highest level of education you have attained?
    1. Bachelor's degree
    2. Bachelor's degree plus further credits
    3. Master's degree
    4. Master's degree plus further credits
    5. Doctorate or professional degree (e.g., Ph.D., Ed.D., J.D., M.D.)

34. What was/were your college major(s)?
    1. History
    2. Other social science (e.g., political science, education, economics)
    3. Other

35. What type of teaching certificate do you have? Please mark only one (your highest certification).
    1. I don't have a teaching certificate.
    2. Regular or standard state certificate offered in the state
    3. Advanced professional certificate (e.g., National Board for Professional Teaching Standards)
    4. Other teaching certificate

36. What is your age?
    1. 25 or under
    2. 26–35
    3. 36–45
    4. 46–55
    5. 56–65
    6. 66 or older

37. What is your ethnicity?
    1. African American or Black
    2. American Indian/Native American
    3. Asian American/Asian Indian/Pacific Islander
    4. Caucasian (non-Hispanic)
    5. Latino, Latin American, Puerto Rican, Hispanic, Chicano

38. Are you….
    1. Male
    2. Female

39. Do you have any comments for us regarding your experience as an AP U.S. History teacher? Is there anything you do as an AP U.S. History teacher that you feel is especially noteworthy? Please use the space below.

CollegeBoard
connect to college success™