CollegeBoard
connect to college success™

AP® Use in Admissions: A Response to Geiser and Santelices[1]

Wayne Camara and Michalis Michaelides
March 11, 2005

Overview

In their 2004 paper "The Role of Advanced Placement and Honors Courses in College Admissions," Saul Geiser and Veronica Santelices of the University of California, Berkeley, address the use of Advanced Placement Program® (AP®) and honors courses as a criterion for admission at the University of California system and suggest that the policy for awarding bonus points to such courses "has little, if any, validity with respect to the prediction of college outcomes" (p. 24). They find that the number of AP or honors courses taken is not a statistically significant predictor of college outcomes, while performance on AP Examinations is strongly related to college performance.

Policy of Awarding Extra Weight to AP, Honors, IB, and Concurrent Community College Courses

In 1982, the University of California instituted a policy of awarding one bonus point to AP and honors courses taken in the last two years of high school. Similar policies of considering advanced courses in college admissions exist in other institutions, particularly selective ones. Admissions officers may consider students' advanced-level courses by examining the number of these courses on transcripts and/or through the bonus weight given to them in the calculation of the high school GPA, as in the case of the University of California.

The authors identify a number of reasons why this policy is an important issue that deserves scrutiny. Some of them are discussed below:

(a) Access to AP or honors level courses may not be equal for all students; there may be disparities, often related to socioeconomic variables. Results from Table 1 in the paper show that the disparities in access to advanced-level courses at the school level are "not as great as perhaps might be expected" (p. 8); for example, schools in the upper API quintile offer on average 14.5 AP courses as opposed to 10.2 AP courses offered by the lowest API quintile schools. No statistical tests are carried out to compare advanced-course offerings across the school API categories. Results from Table 2, which is based on the self-reported responses of the California SAT® population, indicate that the representation of minorities and less advantaged populations in the categories with five or more "AP/Honors subjects taken" is less than that of their counterparts. About 19 percent of the sample report taking five or more advanced-level courses; 26.3 percent of the sample take one to four such courses, and less advantaged subgroups are slightly overrepresented compared to their share in the overall sample. They are also slightly overrepresented in about half of the sample that reported no AP/honors course work. However, the authors cite research from the CSU Institute for Education Reform that points to schools' internal policies for AP participation (e.g., tracking) rather than the availability of the AP courses as a reason for the observed disparities.

(b) The policy itself may encourage schools to offer more rigorous courses, and students to enroll in them, but the resources available to schools may not be adequate to ensure high quality of these courses; or students may casually take the course, without evidence of mastery of the material—in the case of AP courses, for example, enrolled students may not take the end-of-course exam. For admissions purposes, enrollment in such advanced-level courses during the senior high school year suffices, because admissions applications and decisions are made before the end of those courses. Thus, there is no control over student performance and no guarantee that the student had a "truly" rigorous, college-level experience in the course.

(c) More importantly, the authors cite the lack of research on the validity of advanced-level courses as an admissions criterion. The AP Program was developed to enable placement into sequent college courses and/or for granting college credit, and research continues to support this use (e.g., Dodd et al., 2002; Morgan & Crone, 1993; Morgan & Ramist, 1998); however, the use of AP, IB, honors or concurrent community college course work in admissions decisions has not been validated. The authors' attempt to fill the gap on the predictive validity of such courses in the literature is commended. It should be noted that the authors' purpose is to examine the use of AP and other honors courses in admissions to the University of California system with data from the particular, selective population of students who enrolled in the system.

The Predictors

Multiple predictors are used to build the models presented in the study. Most of them are known to be positively correlated:

[1]We would like to thank Rick Morgan, Neil Dorans, Shelby Haberman, and the AP Program staff at ETS for their input.

high-school GPA, SAT combined score, SAT Subject Tests™ scores, parental education level, school API quintile, and number of AP/honors courses taken. The authors provide no simple correlations between predictors or between predictors and outcomes. There is no data that will allow the reader to understand the relationship between AP/Honors/IB and concurrent community college courses with other predictors or outcomes used in the study. The relationships for AP course taking, AP Examinations, and AP grades are not reported with any outcome or other predictor variables. At a minimum, the relationships between these variables must be reported to support the conclusions and assertions the authors make throughout the paper. The magnitude of estimated regression coefficients cannot be clearly determined when collinearity is present. Instead, the authors interpret the contributions of each variable in predicting the outcomes as effects in models assuming no collinearity.

"Number of AP/Honors courses" is the main variable analyzed and discussed in the paper. It is constructed by a count of AP, IB, honors, and community college concurrent courses. At best it is a composite variable composed of four different types of courses that are not equally represented across UC students. Treating these four different types of courses interchangeably is a major weakness of the study, especially given some of the findings; the authors acknowledge the limitation but nevertheless use this variable for their main analyses. In each of the models, they predict college outcomes by discipline using a count of AP/Honors/IB/community college course work taken in any subject. For the Fall 2001 student cohort, they add each of these types of courses separately in the multiple regression model that predicts first-year UCGPA from the HSGPA, demographic, and SAT Reasoning Test™ and Subject Tests variables (Table 6). The numbers of AP and IB courses have statistically significant standardized regression coefficients, albeit small, while the numbers of other honors courses and courses taken at community colleges do not attain significance. The authors argue that because the addition of each of these advanced-level courses does not add much in the explanatory power of the model (AP courses add about 0.1 percent in R^2 while the other course types add less than that), "individually or in combination, AP, IB and other honors-level course work contributes little to the prediction of college performance" (p. 17). However, the value and usefulness of a predictor variable does not depend alone on the variance accounted for or the incremental explanatory power it adds to the model. Rosenthal (1990) describes how treatments with significant effects may have very low correlations and squared correlations with outcomes. Effects sizes for each predictor and model should be reported and increasingly used in such social science studies. Again, the authors fail to report descriptive data and simple correlations that are required to examine the inferences the authors make regarding AP courses and examination grades.

The high school GPA, as well as the college GPA, is an unreliable variable, although typically used in studies of predictive validity. There are no common grading standards across schools or across courses in the same school. The contribution of grades from advanced-level courses is arbitrary if greater weight is assigned to these courses, while at the same time grading may still involve "curving" or instructors may be more lenient or reward students who choose to take a challenging course. It is not clear in the paper how the "AP Exam Scores" variable (Table 7) was defined, whether for example an average was computed if a student had multiple AP Exam scores. AP Exam scores are ordinal, and even though reported on a scale of 1 to 5, they are not comparable. A score of 3 or above often translates to "qualified for advanced-level college course work" but 3s in AP Calculus BC, AP Environmental Science, or AP Comparative Government and Politics, to use three different exams as examples, do not correspond to the same standards. Therefore, adding or averaging AP scores is only a crude approximation for the underlying construct of performance on AP courses and may reduce the incremental validity in the analyses.

The Modeling Technique

Multiple regression is often employed in studies that involve prediction. The technique supports certain types of claims and rests on a number of assumptions. The authors do not acknowledge these issues. First, the analysis is based on observational data. Claims that infer causal effects based on regression coefficients are not supported by this kind of modeling. Even though background and academic variables are used as controls in the regression equation, it is inappropriate to infer effects of variables, especially when no assumptions of the model are discussed. If basic assumptions are violated, then the appropriateness of the model is questionable and any inferences based on the model are dubious. The authors provide no supporting evidence or claims regarding linearity, homoscedasticity, or the interval nature of the data. Multicollinearity is also a serious concern that is completely ignored. The variable of interest, "Number of AP/Honors" is expected to correlate positively with most if not all of the other predictors in the model, as mentioned earlier. The contribution of this variable may have already been accounted for by its correlates. As a result, it is not surprising that it does not turn out to be statistically significant. The authors' failure to account for the assumptions of their modeling technique is a major weakness of the study and the interpretations attributed to their analyses.

A strength of the study is that it examines the validity of an existing policy. The authors properly mention that their purpose is not to assess the value or effectiveness of the AP or other honors programs. They do not consider the validity of these programs for placement or credit-granting policies. They use high school data from grades 10 and 11, and not 12, since these years of data are available at the time of admissions—in footnote 15 they note that including twelfth-grade data in the analysis did not produce different results. However, the variables they use to build their models do not replicate the admissions process. The University of California considers a capped version of a weighted high school GPA (see footnote 13), but the predictor entering the regression models is the unweighted variant and the number of AP/honors courses is treated as a separate variable (in the actual admissions process, the latter indirectly manifests itself in a weighted GPA). This decision was made based on the results of Table 3, which merely shows the R^2 for predicting college GPA from SAT combined scores, SAT Subject Tests scores, and the various versions of the high school GPA.[2] Because the models with the unweighted variant exhibited slightly higher R^2, the authors employ that variant for their subsequent analyses, even though they introduce additional variables in the model. In addition, AP course taking in the eleventh grade is quite distinct from AP course taking among twelfth-graders. Several advanced math, computer science, and other courses are overwhelmingly comprised of twelfth-graders, and omission of these students would very likely result in a much smaller and unrepresentative sample of students for several AP Examinations. If the authors do find that exclusion of all seniors still results in a representative sample of students in every AP course, they should provide that data in the appendix.

The academic variables are introduced because they are in fact used in the admissions process, but then the authors include demographic variables in the same prediction models. Does UC actually use parental education and family income in admissions decisions—favoring students who come from more advantaged households? If this is not used in admissions, then why is it used in the prediction equation? And the incremental validity of AP and honors courses is considered only AFTER the contributions of "Parents' Education," "Family Income," and "School API Quintile" are included.[3] The authors comment that the introduction of "additional demographic variables into the regression analysis does not, in short, help improve or explain the null relationship between AP/honors course work and college grades" (p. 14), but the inclusion of such variables appears to be designed to reduce the incremental validity of course rigor and is not relevant to the stated purpose of the study. If the authors seek to illustrate that family income and parental education are related to test scores and course rigor, that is a separate issue. Camara and Schmidt (1999) illustrated that parental education and family income are highly related to test performance on a wide variety of measures, as well as high school grades, course rigor, and graduation rates. However, inclusion of socioeconomic or demographic variables in a regression model used to evaluate the incremental contributions of various predictors is inappropriate unless the admissions policy explicitly includes these factors. Rather than clarifying the policy issues, the current study is obfuscating the central purpose of the study, to evaluate admissions policies, with subgroup differences. Inclusion of socioeconomic and demographic variables in this specific analyses appears intended to reduce the variance accounted for by the predictors and should be examined separately.

The Findings

The main claim of the study is that the number of AP/honors courses that a student takes in high school is not a statistically significant predictor of college performance. Existing research shows that academic intensity and quality of high school curriculum (often defined in terms of AP and other honors courses) are the most important factors in preparation for college degree completion (Adelman, 1999). A number of possible explanations of why their claim may not be warranted have been mentioned above and are summarized and extended here: the "effect" of the advanced-level courses is already included in the high school GPA, or in the SAT score variables, since taking advanced-level classes may well result in higher test scores, particularly in SAT Subject Tests. The more AP/honors courses a student takes, the more advanced courses he/she will take in the first two years of college, and as a result will be faced with more challenging material, which may lead to lower freshman or sophomore GPA compared to his/her classmates who take the lower level courses during the same period. The relationship between number of advanced-level courses and college grades may not be linear; is it reasonable to expect as much difference between a student with 5 such courses and a student with 10, as with a student with no such courses and a student with 5? The relationship may well be curvilinear. Students who enroll in advanced-level high school courses choose to take a rigorous curriculum, and the distinctions between students who take many such courses (diverse in content and in type: AP, IB, honors, etc.) may not account for a large portion of variance. Especially since all the data analyzed come from the students already accepted in the University of California system, a highly selected population, which implies a restricted range for the variables examined.

Another important finding is the significance of the "AP Exam scores" variable in Table 7, despite the inclusion of the same set of control variables in the model as in other models. This

[2]It would be interesting to see what the estimates for the regression coefficients for each predictor are in the models on Table 3.

[3]As in Table 6, for example.

variable is not very precise given the ordinal nature of the AP reporting scale that has only 5 scores. Averaging across ordinal and noncomparable AP Exams score scales is also an approximation of the construct. In this variable only students who took AP Exams are included, i.e., other honors level students and AP students who did not take the corresponding exam do not contribute information. The statistical significance and magnitude of the regression coefficients for "AP Exam scores" are second only to the best predictor "Unweighted HSGPA" of the model. Assuming appropriateness of the model, this finding points to a strong relationship between AP and college performance even after controlling for HSGPA, SAT Reasoning Test, and SAT Subject Tests scores and demographic variables and suggests that AP scores demonstrate predictive validity in an admissions context.

The authors use the results presented in Table 6 to justify their decision to lump all the advanced-level courses together in the rest of their analyses. We believe that they could probably disaggregate the AP course work in their data since some of their results signify a differentiation of the AP performance compared to other advanced-level courses (e.g., large and significant effect of AP Exam scores in Table 7, small and significant effect of AP courses in Table 6). Given that the data included AP Exam scores for each student were available (to create the model for Table 7), the authors could easily create a variable for the "Number of AP Exams taken" and study it separately from the other advanced-level courses. However, they make claims such as "AP course work, by itself, contributes almost nothing to the prediction of college performance" (p. 17). Their analyses do not examine AP course work by itself, only grouped together with other honors courses, and hence do not warrant those claims.

The above claim is a misinterpretation of the results, but is in part promoted by some approximations provided in the paper. First, AP course offerings are presented as the predominant type of advanced-level courses in Table 1: 72 percent of such courses are AP courses. This is an approximation because the various types of advanced-level courses are not necessarily exclusive in a school, as the authors note. Moreover, this estimate does not take into consideration the size of classes. The rest of the analyses are done with student-level data, while the 72 percent figure is based on class-level data. Second, it is not known what percentage of AP students take the AP Exam. An estimate is cited on page 4 from a report by the Commission on the Future of the Advanced Placement Program: over a third of the AP students do not sit for the exam. In footnote 5 the authors state that they will provide an estimate in their paper, but they only provide a very rough approximation of 56 percent in footnote 17. According to projected AP enrollments in the Participation Survey conducted by the College Board,[4] 66.6 percent of enrolled students took the exam in 2000 in California, and the percentage has been gradually increasing to 74.9 percent in 2004. There are a few arithmetic errors in the report: the sum of the N column on Table 7 should be 14,922, and the sample sizes mentioned in footnote 18 do not match the numbers in the corresponding tables.

Finally, the authors state that their research is designed to examine the admissions policies related to providing additional credit for AP, honors, IB, and concurrent community college courses. However, they never provide any data that would allow the reader to clearly examine the relationship among the predictors and a single predictor with outcome variables. The methods used obscure and hide these relationships. The authors include socioeconomic variables in their models that are not part of the UC admissions process and then report that rigorous courses have a marginal impact on college success.

Summary and Conclusions

Geiser and Santelices emphasize that their study "is not intended as an assessment of the value or effectiveness of the Advanced Placement or International Baccalaureate programs, nor of other honors-level course work offered by high schools in either the U.S. or California" (p. 23). They argue, however, that after "controlling for other academic and socioeconomic factors, the number of AP and other honors-level courses taken in high school bears little or no relationship to students' later performance in college" (p. 18), a claim that is inconsistent with existing research on the importance of academic rigor (e.g., Adelman, 1999). They also support that "AP Exam scores are strongly related to college performance" (p. 19), and they maintain that students who sit in AP courses or other honors courses and do not take the exam could explain the discrepancy between their two findings.

When the U.S. National TIMSS Report was released in 1998, data were presented illustrating that the performance of high school students taking "advanced" physics and calculus was among the mid-range or below of countries participating in that study. Just as with the Geiser and Santelices study, this TIMSS report combined honors, AP, and other courses into that advanced group of courses, and often made claims that implied even AP students performed below average international comparisons. Gonzalez et al. (2001) replicated the TIMSS study, but they chose students who were in AP courses that were taught in schools where there were *actual* AP classes.[5] When they examined the performance of AP students enrolled in these courses, they found AP students, whether scoring 3 or better or less than 3, performed significantly better than the students in the composite "advanced physics and calculus" group that

[4]Available upon request.

[5]The study sample consisted of students "in schools registered with the College Board as having AP Calculus or Physics courses, and intact AP classes were selected for testing in these schools" (Gonzalez et al., 2001, p. 4).

had been formed. In the original TIMSS study, advanced AP students obtained an average achievement score of 442 ranking them fifteenth of 16 countries where the international average was 501. When AP students were examined alone, their average achievement score was 573, which would have placed them at the top of the 16 nations, significantly better than all nations but France. When students who received AP grades of 3 or better were examined, their average achievement was 586 (for AP Calculus AB) and 633 (for AP Calculus BC).[6] In AP Physics, the results were similar. TIMSS reported U.S. students in advanced physics had an average achievement of 423, placing them last among all participating nations, whereas students enrolled in AP Physics had an average achievement of 529 in the Gonzalez et al. study (2001), which would place the U.S. average fourth among the 15 nations. Average achievement for students with AP Physics grades of 3 or better were 586, 600, 572, for the three AP Physics courses, well above the international average of 501 and statistically equivalent to the first-place nation (Norway's average achievement of 581).

In the current review we discussed a number of other reasons why the authors' claims extend beyond their data. When their results appear to run contrary to other research that has demonstrated the importance of academic rigor in predicting college success, the authors have a responsibility to provide direct comparisons among predictors and outcomes in a straightforward manner, as well as provide estimates of effect sizes when claiming large differences or discounting the impact of any factor. We also expect other researchers will want to have access to the UC data for purposes of replication because these results are so different than those found from national samples, and we expect UC will make that data available to explore these and other differences.

In their concluding section, the authors do extend their findings to the policy domain. They discuss three policy alternatives for revising the admissions procedure for the University of California. In their discussions they consider equity, practicality, and supplementary and unintended consequences of using AP and honors-level courses in admissions, such as "maintaining an incentive for students to take rigorous, higher level course work while minimizing disparities" [p. 22], beyond the predictive validity findings. As a result, the options of requiring minimum AP Exam scores, considering AP/honors in the local context, and reducing the weight placed upon AP/honors course work, even though plausible, appear difficult to implement. The paper would be much more useful if the collinearity among predictors was directly addressed, the contradictory findings between this study and other research on academic rigor (Adelman, 1999; Gonzalez, et al., 2001; etc.) were explored, and factors that are not mechanistically used in admissions policies (socioeconomic and demographic) were not included in regressions equations that are attempting to evaluate variance components of factors in the admissions process.

References

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

Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification.* College Board Research Report RR 99-5. New York, NY: College Board.

Dodd, B. G., Fitzpatrick, S. J., De Ayala, R. J., & Jennings, J. A. (2002). *An investigation of the validity of AP grades of 3 and a comparison of AP and non-AP student groups.* College Board Research Report no. 2002-9. New York, NY: College Board.

Geiser, S., & Santelices, V. (2004). *The role of Advanced Placement and honors courses in college admissions.* Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley. Downloaded from http://cshe.berkeley.edu/publications/papers/papers.html on Dec. 22, 2004.

Gonzalez, E. J., O'Connor, K. M., & Miles, J. A. (June 2001). How well do Advanced Placement students perform on the TIMSS Advanced Mathematics and Physics tests? Technical Report of the International Study Center, Lynch School of Education at Boston College. Boston, MA: International Study Center.

Morgan, R., & Crone, C. (1993) *Advanced Placement Examinees at the University of California: An investigation of the freshman-year courses and grades of examinees in biology, calculus AB, and chemistry.* ETS Statistical Report 93-210. Princeton, NJ: Educational Testing Service.

Morgan, R., & Ramist, L. (1998). *Advanced Placement students in college: An investigation of course grades at 21 Colleges.* ETS Statistical Report no. 98-13. Princeton, NJ: Educational Testing Service.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist*, *45*, 775–777.

[6]See Gonzales et al. (2001) exhibits 3 and 5. The international average was 501. The U.S. average for all advanced students was 442, while France had the highest average achievement score of 557. When students enrolled in AP courses were examined, their average achievement was 573. The differences between students who scored 3 and above on AP versus below 3 were not that great, and both groups clearly outscored all national averages. For Calculus AB, students with an AP grade of 3 or better averaged 586 (versus 565 for those with scores of 1 to 2). For Calculus BC, students with an AP grade of 3 or better averaged 633 (versus 564 for AP students with grades of 1 to 2).