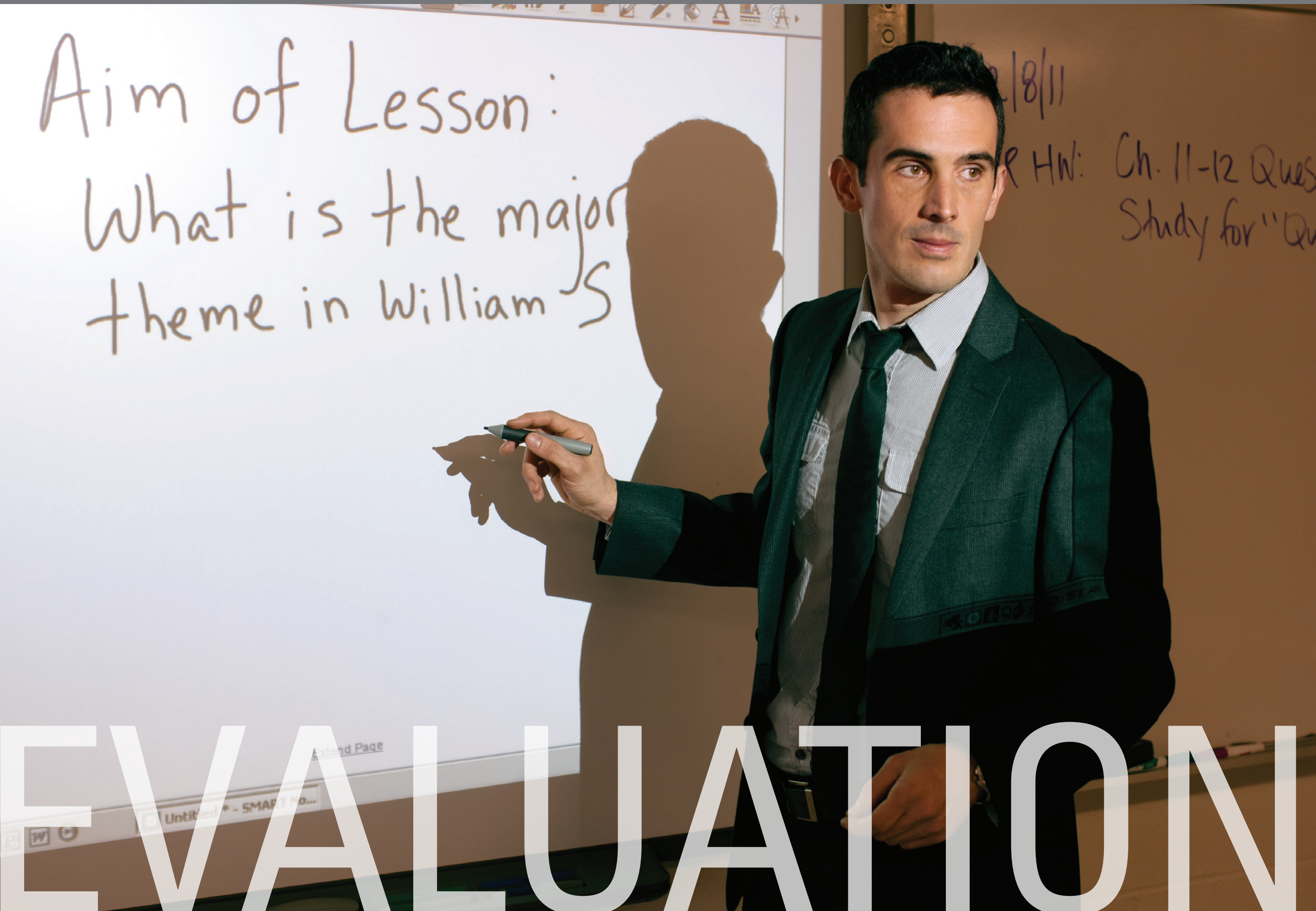
RESEARCH REPORT 2012-8

CollegeBoard

An Analysis of the Relationship Between School-Level AP® Professional Development Activity and Subsequent Student AP Performance

By Vytas Laitusis



Vytas Laitusis is a research scientist at the College Board.

The author wishes to acknowledge the contributions of Thanos Patelis and Jennifer Bausmith for their feedback on earlier drafts of this manuscript.

About the College Board

Contents

TION

Executive Summary

The overarching purpose behind this evaluation was to gauge the impact of AP® professional development (PD) on AP student outcomes in a state with a significant rate of PD implementation. The evaluation attempted to predict the level of student AP performance by the number of AP professional development events attended by teachers in that school in the prior year, while controlling for some socioeconomic status (SES), teacher, and school effects. The outcomes predicted by the number of PD events attended were defined as the average AP score obtained for that school as well as the percentage of AP Exam takers scoring 3 or above. A similar analysis was also performed for AP courses comprising the STEM disciplines (Biology, Chemistry, Computer Science, Environmental Science, Calculus AB, Calculus BC, Physics B, Physics C: Mechanics, Physics C: Electricity and Magnetism, and Statistics). The controlling factors (covariates) used in the analyses were average household income (a proxy for SES), the percentage of students taking AP in the school (school effect), and the average number of years teaching AP (teacher effect).

The results were as follows:

- After controlling for average household income (SES), level of AP activity, and teacher experience, schools with higher levels of teachers participating in AP PD were more likely to have higher levels of overall average AP performance (average exam score and average percentage of exams with scores of 3 or above) the following year.
- In addition to the number of PD events attended, teacher experience was also a statistically significant predictor of subsequent overall AP performance.
- For STEM-related AP Exams only, the level of AP PD attended by teachers in the school was also a statistically significant predictor of subsequent AP performance. Teacher experience was also significant and more predictive for AP STEM course performance than overall AP performance (reported in the previous bullet).

Given some data access restrictions that did not allow for teachers to be directly linked to their students, drawing direct causal inferences between AP PD and positive student outcomes is not warranted. In addition, there was no information collected regarding implementation of skills learned from the PD. However, the results do point to the potential unique contribution that a school's level of AP PD activity may have on subsequent student success on AP Exams while accounting for a set of school environmental factors.

Introduction

The continuing education of teachers has long been considered a cornerstone to improving our education system (Borko & Putnam, 1995; Darling-Hammond, 1993). Although the field of PD is varied in terms of approach and apparent outcomes, Desimone (2009) has identified a set of well-accepted characteristics of professional development (content focus, active learning, coherence, duration, and collective participation) that serve as a foundation for these efforts. Although there is some value placed on the direct impact of PD on teachers in the areas of job satisfaction, increased content, or pedagogical knowledge, students should ultimately be the beneficiaries of this training. However, this indirect effect of PD on student outcomes raises significant challenges for measuring any potential impact of the PD program. If PD is to affect student outcomes, success is contingent upon the acquisition of skills and strategies by the teacher. Once teachers possess these skills, they must implement them effectively in the classroom. In addition, although it seems plausible to expect some indirect impact of a teacher's newly acquired skills on students, it is equally important to note the complexity of the system with which student learning takes place. Educators do not teach and students do not learn in a vacuum; they are part of a complex dynamic of social, financial, and educational factors. In assessing the efficacy of any program, it is important to acknowledge this complexity.

The relative level of PD engagement within a school or some amount of collective participation has been shown to have positive effects by promoting an atmosphere of shared understanding among colleagues regarding goals, methods, problems, and solutions (Ball, 1996). In addition, research has clearly demonstrated that PD should be a sustained effort as opposed to a singular event in order to be effective (Guskey & Yoon, 2009). This finding highlights the importance of a school or district engaging in and maintaining an active program of PD. There is also evidence that the benefits of PD are not necessarily apparent immediately and that there is often a lag from the time PD is delivered to an effective manifestation of that PD in the classroom (Guskey, 2000). Evaluations need to take this potential latency into account when planning impact studies so the effects on students have adequate time to manifest themselves. It also points to the importance of evaluation studies to follow up on outcomes initially and then continue over an extended period of time.

The current study investigated the relationship of a school's level of AP PD activity on subsequent student performance on the AP Exam while controlling for teacher, school, and SES effects. In an effort to account for factors outside of AP PD participation that might influence student AP performance, the analyses included proxies for these effects such as years of experience (teacher), level of AP activity (school), and average household income (SES). Teacher experience was defined as the mean number of years teaching AP. Research has demonstrated that more years of teaching experience can have a positive impact on student performance though these results tend to be mixed (Wenglinsky, 2002; Rice, 2010). A majority of the positive effects are generally seen when comparing new teachers to those who have several years of experience. The benefit of experience does not increase at the same rate that experience increases (Rice, 2010; Goe, 2007; Wenglinsky, 2002). For example, the effect of experience when comparing a new teacher to a teacher with five years' experience is more pronounced than when comparing a teacher with 20 years' experience to another with 25. The relationship between SES and educational outcomes has been a pervasive topic for several decades since it was highlighted in the work of Coleman (1966). For the purposes of this study, a proxy for SES (average household income) was used. Average household income was taken from the National Center for Education Statistics (NCES) database and provides some information on the economic standing of the students in the school.

The current study examined the relationship between the level of AP PD activity undertaken in a school and the subsequent utility of that PD in predicting student outcomes while accounting for student, teacher, and school effects. More specifically, the current study examines the level of AP PD activity and subsequent AP performance both in terms of average grade and percentage of students scoring a 3 or above (which is often used as an indicator of success). The analyses focused on AP outcomes for all subject areas as well as those only related to STEM courses.

Following an overview of the AP professional development program, this report will describe the methodology used to examine these outcomes. Two sets of analyses will be reported. The first will simply examine the differences in student outcomes given the level of school engagement in AP PD. The second analysis will also utilize the level of school AP PD activity and incorporate other potentially mediating school, teacher, and student effects into a prediction model of subsequent student AP performance. The report will conclude with a discussion of the study's broader implications and limitations.

AP Professional Development

In order to reach a large constituency and provide a variety of PD offerings, the AP Program enlists the support of a large network of educators to conduct a wide array of workshops and institutes. AP workshops are led by an accomplished member of the AP community and/or a College Board–endorsed consultant. The workshops typically guide participants through the skills students must master in the AP classroom and the most innovative teaching practices to help engage students. These PD events often concentrate on specific subject-area topics and offer curricular assistance for new, intermediate, and experienced AP teachers.

Some recent research has supported the notion that this AP PD is perceived by teachers as an effective factor in their teaching and that this perception is maintained after implementing the newly obtained strategies in their own classrooms (Godfrey, 2009; Laitusis & Barry, in press). Although the positive perceptions of teachers are valuable, evidence relating to the efficacy of this PD on student performance is also of ultimate concern. Bausmith and Laitusis (in press) investigated the impact of the AP Achievement Institute I (APAI I) on subsequent student AP performance. The study matched a set of students to their participating teachers before their engagement in PD and then matched a subsequent cohort of students to these same teachers after the PD. The results indicated that although the teacher classrooms had become more diverse and represented lower levels of student prior achievement (as indicated by state assessment data), student performance outcomes on AP were maintained from the prior year (Bausmith & Laitusis, in press). Although these results seem promising and student prior achievement was taken into account, other factors such as school and teacher effects that may influence outcomes were not incorporated. The current study attempts to address the greater level of complexity inherent in an educational setting by incorporating school, teacher, and students effects into a PD model predicting student AP success.

Method

The study utilized the AP PD activity of a set of teachers in a Midwestern state for the 12-month period leading up to an academic year. Then the AP data from the students who attended the schools where these teachers taught were matched using data from the following spring administration. The original sample used for analysis consisted of 309 schools with AP Exam data. In order to increase the stability of any statistical analysis, only schools with at least five or more test-takers were retained. This resulted in a reduced sample of

247 schools (i.e., 80%). Due to the reliance on background data obtained from additional extant databases (NCES), there were several additional schools that were removed from the regression analysis, resulting in a data set with 197 schools. Since the purpose of the second analysis was to investigate the relationship between the level of AP PD and student AP outcomes, only schools that had PD participation information in the database were retained for the regression analysis, which further reduced the sample to 136 schools.

The number of AP PD events was aggregated from an internal database of teachers who had registered for College Board–sponsored AP PD offerings during a 12-month time period leading up to the school year. If a school did not show any record of teachers attending AP PD during the experimental window, they were assumed to have not participated in PD. It is unknown whether any of the teachers in these schools also participated in other forms of PD or attended AP PD that was not recorded in the database. The PD events represented a fairly broad range of topics and ranged in duration from one to five days. The unit of analysis representing AP PD activity for this study was the number of AP PD events attributed to a school in a particular 12-month period leading up to an administration year. Outcome data (student AP scores) were then obtained for this subsequent administration year. For example, the level of school AP PD activity from September 2004 through August 2005 was used to predict student AP performance in the spring of 2006 administration. Since matched teacher–student data could not be obtained, the unit of analysis was at the school level so all teacher PD participation and student performance were aggregated up to the school level.

In addition to examining the relationship between the level of AP PD and overall AP performance, the study also examined student performance on a subset of the exams relating to the STEM disciplines (Biology, Chemistry, Computer Science, Environmental Science, Calculus AB, Physics B, Physics C: Electricity and Magnetism, Physics C: Mechanics, and Statistics).

Two sets of analyses were undertaken to examine the relationship of AP PD activity and student performance on AP. First, two one-way analyses of variance (ANOVAs) were undertaken to examine whether student AP performance measured as mean AP Exam score and percentage of exams with a score of 3 or above was statistically significantly different across three levels of teacher activity in AP PD. Second, to control for other variables that have been shown to be related to student performance, a regression analysis was used to examine the relationship of AP PD activity on student AP performance when teacher experience, SES levels, and school AP activity were controlled. AP activity was an indicator of the relative size of the AP program at the school and was computed as the percentage of students at the school taking at least one AP Exam. The unit of analysis for all indicators was at the school level.

Results

Although the schools with no AP PD data for the year in question were not retained for the regression (second) analysis, it was still of interest to understand the characteristics of this group of teachers and how they compared to the sample used in the study. In order to accomplish this, a one-way ANOVA was performed comparing three groups of schools: (1) schools with no known AP PD activity, (2) schools with low AP PD activity, and (3) schools with higher AP PD activity. A second ANOVA was conducted using the percentage of AP Exam scores at or above a score of 3 as the performance outcome. The categories representing the level of AP PD activity were determined by taking a median split of the PD activity variable, which turned out to be 3 or fewer events for low and more than 3 for high activity. Student performance in terms of mean exam score as well as mean percentage scoring 3 or above are displayed in Table 1 for each of these groups.

Table 1.

Mean Student AP Performance Outcomes on All Exams by Prior Year AP PD Participation

Outcome	AP PD Events	*n*	*Mean*	*SD*	*SE*
AP Exam Score (All Exams)	None	61	1.82	0.55	0.07
	Low	67	1.98	0.52	0.06
	High	69	2.33	0.48	0.06
	Total	197	2.05	0.55	0.04
Percentage Scoring 3+ (All Exams)	None	61	0.22	0.20	0.03
	Low	67	0.26	0.22	0.03
	High	69	0.41	0.20	0.02
	Total	197	0.30	0.22	0.02

The results of the analyses indicated a statistically significant effect between groups for AP Exam (all exams) score, $F(2,194) = 16.82$, $p < .01$ indicating a statistically significant difference with respect to student performance across the different categories of previous school AP PD activity. An examination of the effect size of $\eta^2 = .15$ indicates that this effect is in the moderate range (Cohen, 1988). A set of post-hoc analyses using the Scheffe procedure indicated that significant differences were noted between the "high" PD category and both of the other two categories ("none" and "low"). Therefore, there were no significant differences with respect to AP scores between the "no activity" and "low activity" schools the following year. Only schools with a higher level of PD activity statistically differentiated themselves with respect to mean AP score the following year.

Similar to the findings for mean AP score, there was also a statistically significant difference between groups for student performance represented by the percentage of exams with scores of 3 or above across levels of AP PD, $F(2,194) = 15.89$, $p < .01$, $\eta^2 = .15$ indicating a significant difference with respect to student performance based on previous school AP PD activity. A set of post-hoc analyses using the Scheffe procedure indicated the same pattern of results presented for mean AP Exam score. Only schools with a higher level of PD activity statistically differentiated themselves with respect to the percentage of exams with AP Exam scores of 3 or higher.

A similar set of analyses were also performed on student AP outcomes for STEM-related courses. In these analyses only schools with five or more AP students and scores on at least five or more STEM exams were retained. The descriptive results are summarized in Table 2.

Table 2.

Mean Student AP Performance Outcomes on STEM Exams by Prior Year AP PD Participation

Outcome	AP PD Events	*n*	*Mean*	*SD*	*SE*
AP Exam Score (STEM)	None	29	1.71	0.73	0.14
	Low	41	1.80	0.65	0.10
	High	54	2.29	0.64	0.09
	Total	124	1.99	0.71	0.06
Percentage scoring 3+ (STEM)	None	29	0.20	0.25	0.05
	Low	41	0.22	0.23	0.04
	High	54	0.42	0.23	0.03
	Total	124	0.30	0.25	0.02

Similar to the results presented for all exams, there appears to be higher AP STEM exam performance as number of AP PD events increases for the school in the prior year. The between groups difference for STEM performance by mean score and mean percent scoring 3 or above were both statistically significant $F(2,121) = 9.65$, $p <= .01$, $\eta^2 = .14$ and $F(2,121) = 11.48$, $p <= .01$, $\eta^2 = .16$, respectively. A post-hoc analysis indicated the same pattern reported previously for all exams; statistically significant differences were seen only between the group representing the high number of PD events and the other two categories ("none" and "low") and no significant difference between the "none" and "low" PD activity groups.

The results presented so far seem to suggest an association between a highly active professional development environment and positive student performance outcomes. However, the results presented thus far only show the relation between teacher-activity levels in AP PD and student AP performance outcomes in isolation without accounting for the complexity of the educational environment that may also influence student exam performance. The dynamics inherent in any education system are varied and complex. In an attempt to address at least some aspect of this complexity, another set of analyses was performed to examine the role of AP PD in this system by accounting for some of the variables that have been shown to relate to student performance. The variables that were used in these analyses include (a) teacher experience (number of years teaching AP-level course work), SES (average household income), and school (level of AP activity) factors.

Before the regression models were run, each of the variables was checked to see if it satisfied the statistical assumptions underlying this form of analysis. Checks for normality, multicollinearity, and both univariate and multivariate outliers were performed for each independent variable (years teaching AP, household income, AP activity, and number of AP PD events) in the model. The results of these checks revealed moderate levels of skewness for all of the variables with the exception of household income. In order to correct for this lack of normality, a log-base-10 transformation was performed on each of the skewed independent variables.

A sequential regression analysis was undertaken in two steps. Teacher experience, parental income, and school AP activities were entered first. This was done to control for the effect of these variables on student AP performance. The next step in the model added an additional predictor (AP PD events) to the previous three. AP PD events was our independent variable of greatest interest in that we wanted to see if this variable could explain any student performance above and beyond that explained by the background (teacher, SES, and AP activity) variables. Therefore, any differences in the model between step 1 and step 2 were said to represent the influence of AP PD events above and beyond that of the background variables.

Of initial interest was whether the model tested was statistically significant. Table 3 presents the results of the regression model. The significance test for the first two steps of the model were both significant, with $F(3,131) = 3.87$, $p <= .01$ and $F(4,130) = 8.96$, $p <= .01$ for steps 1 and 2, respectively. Also of note was how much of the variance attributed to the dependent measure could be explained by the predictors. A measure of this is obtained in the multiple correlation coefficient (R^2). In this example the adjusted R^2 was reported, because it is generally regarded as a more conservative estimate as it takes into consideration the number of predictors and the sample size, which may artificially inflate the multiple correlation statistic. The initial step with background variables only produced an $R^2 = .06$, indicating that 6% of the variance of student AP mean score was accounted for by these variables. With the addition of the number of AP PD events the R^2 increased to .19, causing a change in R^2 of .13. All of these R^2 values including the change in R^2 were significant at the $p <= 0.01$ level of significance. This indicates that the number of AP PD events attended in the prior year significantly adds to the prediction of subsequent AP performance above and beyond the teacher, SES, and school effects used in this study.

Table 3.

Summary of Regression Parameters Predicting AP Exam Performance for All Exams

	Variable	*B*	*SE B*	*β*	*t*	*Sig (p)*
$R^2 = .06$						
Step 1	(Constant)	1.84	0.24		7.66	0.00
	Household Income	-0.01	0.02	-0.06	-0.69	0.49
	AP Activity	0.08	0.15	0.04	0.52	0.60
	Experience	0.63	0.20	0.26	3.14	0.00
$R^2 = .19$						
Step 2	(Constant)	1.27	0.25		5.02	0.00
	Household Income	0.01	0.02	0.04	0.45	0.65
	AP Activity	-0.16	0.15	-0.09	-1.08	0.28
	Experience	0.51	0.19	0.22	2.75	0.01
	AP PD Events	0.50	0.11	0.42	4.73	0.00

While discussion so far has focused on the model as a whole, also of note is the relative importance or role each of the independent variables has in predicting student outcomes. Table 3 provides a summary of these predictors and their relative contribution for explaining student mean performance for all AP subject areas. Because of the different scales associated with each of the predictor variables, standardized regression coefficients were evaluated (β in Table 3). Each β is displayed with its corresponding level of statistical significance. An inspection of this table indicates that two variables — “experience” and number of “AP PD events” — make statistically significant contributions to the explanation of student performance on the AP Exam in the subsequent year. The β for experience and AP PD of .22 and .42, respectively, also shed light on the relative importance of each of these predictors in the model. Although experience is a statistically significant factor, the level of AP PD in a school had a greater level of association with subsequent mean AP score while accounting for the previously mentioned teacher, SES, and school background effects in the model.

A second regression analysis was performed for student performance on AP Exams only in the STEM disciplines (Biology, Chemistry, Computer Science, Environmental Science, Calculus AB, Calculus BC, Physics B, Physics C: Electricity and Magnetism, Physics C: Mechanics, and

Statistics). For these analyses all predictors and AP PD participation variables were the same as before. The only modification to the previous model presented was with respect to the outcome measures (performance on all AP Exams was replaced by performance on only STEM exams).

The STEM model was initially tested for overall statistical significance. The significance tests for the first two steps of the model indicated that both were statistically significant, with $F(3,90) = 6.52$, $p <= .01$ and $F(4,89) = 11.22$, $p <= .01$ for steps 1 and 2, respectively. The initial step of the model, which included only the background variables (household income, AP activity, and teacher experience), resulted in an adjusted $R^2 = .15$. This indicated that 15% of the variance of student scores on AP STEM exams was accounted for by these background variables. With the addition of the number of AP PD events, the R^2 increased to .31, causing a change in R^2 of .16. All of these R^2 values, including the change in R^2, were significant at the $p <= .01$ level.

Each β was also tested for its level of statistical significance. Table 4 provides a summary of these predictors and their relative contribution to explaining student mean performance on AP STEM exams.

Table 4.

Summary of Regression Parameters Predicting AP Exam Performance for STEM Exams

	Variable	*B*	*SE B*	*β*	*t*	*sig*
$R^2 = .15$						
Step 1	(Constant)	1.54	0.38		4.01	0.00
	Household Income	-0.02	0.02	-0.06	-0.64	0.52
	AP Commitment	0.33	0.26	0.13	1.28	0.21
	Experience	1.26	0.32	0.38	3.93	0.00
$R^2 = .31$						
Step 2	(Constant)	0.74	0.39		1.90	0.06
	Household Income	0.01	0.02	0.05	0.50	0.62
	AP Commitment	-0.01	0.24	0.00	-0.04	0.97
	Experience	1.12	0.29	0.33	3.83	0.00
	AP PD Events	0.67	0.15	0.44	4.58	0.00

An inspection of this table indicates that the variables "experience" and "AP PD events" significantly contribute to an explanation of student performance on AP STEM exams in the subsequent year. The β for experience and AP PD of .33 and .44, respectively, are similar to the findings reported for all exams in that both were significant and AP PD had the greater weight. The main difference in this model is the increased importance of teacher experience in the model.

The regression results reported here using the mean grade for all exams and the mean grade for STEM exams as dependent variables were also replicated using the percentage of students scoring 3 or above and are reported in Appendixes A and B. The results for the percentage scoring 3 or above for both the "all exams" and "only STEM" models were consistent with those reported just for mean grade.

Conclusions

The results of this study provide some incremental evidence that a school culture that includes a significant engagement in professional development activities is associated with higher student performance. The model that was tested indicated that a school's level of AP PD activity can predict student outcomes above teacher experience and other factors, such as household income and a simple level of AP activity (percentage of students taking AP in the school). The influence of experience also seemed to play a significant role, and the effects seemed to increase in the STEM-related exams. However, the level of PD activity was still the most significant contributor. Given the unit of analysis and the inability to directly attribute student test scores to specific teachers, it is difficult to say for certain that the specific PD events were the cause, or whether the outcomes were unduly influenced by some artifact of a school culture with a robust professional development component. In addition, although the professional development and teacher-experience predictors in the model were significant, there was still a majority of the variance unaccounted for in the model, indicating that there is much more that could help inform our understanding of student performance.

Research in professional development is characterized by significant levels of variability over what works and what does not (Guskey & Yoon, 2009). However, the field is beginning to identify elements that generally seem to be indicative of positive PD outcomes. The variations in many of these studies are likely due to the complex nature of the environment where the intervention is taking place. This report highlighted some of this complexity and made an attempt to incorporate a certain level of it into the design. The education system in practice and the individuals who comprise it are significantly more complex. Any research in this area needs to appreciate this complexity and understand and incorporate it into further work.

Limitations

Given the previous discussion regarding the complexities inherent in teachers imparting knowledge gained through professional development to their students, it is unwise to draw too many inferential claims from these results. The grain size for the analysis is moderate (school level) and the sample was limited to a geographical region and affected by some limitations inherent in the data. Since the PD is not directly linked from teachers to students, it may be more appropriate to couch the results in terms of a culture of PD affecting student outcomes as opposed to any causal effect regarding the number of PD events and its impact on student performance. However, even with these limitations, the results do begin to address some of the complexities inherent in the educational process while pointing to some potential positive effects of PD and the influence of other factors such as experience.

Another factor in determining the effectiveness of a training program is an understanding of the program's level of implementation. Perceived program effectiveness can be influenced by both the degree and quality of this implementation. Unfortunately, implementation data were not available for this study. Understanding mediating factors such as level of support from administration, alignment with organizational goals, and level of collaboration with peers would provide a context for assessing the degree to which a program is realizing its stated goals. Ideally any evaluation of this sort would benefit from data that directly linked teachers engaged in PD activities to their students' subsequent performance over the course of several years, along with additional controls for confounding variables.

Future Work

Given the limitations of the study, there are several areas that can be addressed that would add to the rigor of the study. Further research could begin to unpack some of the relationships described in this study by increasing the specificity of the analysis to individual students and teachers and linking them together in the analysis. This would allow for greater specificity in terms of who attended the PD and the outcomes of the students they teach. Since the PD was ongoing, there would be some difficulty in producing a clear pre–post comparison but the use of a control group to draw comparisons may be more practical and would increase the rigor. Further research could also better address the climate within which learning takes place, leading to a more complete understanding of how strategies learned by teachers are implemented and what some of the barriers for this implementation may be in the classroom. The study also did not collect information on how any of the PD was implemented in the classroom. Subsequent work in this area would benefit from a greater understanding of the level of implementation and characterization of challenges that may confront this implementation.

Finally, any single study should not be viewed as providing conclusive evidence for any relationship under investigation. Rather, these results should be viewed as an additional source of incremental evidence toward a more complete understanding of an otherwise complex dynamic.

References

Ball, D. L. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan*, *77*(7), 500–508.

Bausmith, J., & Laitusis, V. (in press). *The impact of AP Achievement Institute I on students' AP performance*. (College Board Research Notes RN-5505). New York: The College Board.

Borko, H., & Putnam, R. T. (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms & practices* (pp. 35–66). NY: Teachers College, Columbia University.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coleman, J. S. (1966). *Equality of educational opportunity* (Report No. OE-3800). Washington, DC: National Center for Education Statistics.

Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan*, *74*(10), 752–761.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Godfrey, K. (2009). *Advanced Placement Summer Institute: A survey of Florida AP teachers* (College Board Research Notes RN-39). New York: The College Board.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

Guskey, T., (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.

Guskey, T. & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495–500.

Laitusis, V., & Barry, C. (in press). Pre-AP® Science Vertical Teams Summer Institute evaluation summary: A one year follow-up (College Board Research Notes RN 5504). New York: The College Board.

Rice, J. (2010). *The impact of teacher experience examining the evidence and policy implications* (CALDER Brief 11). Washington, DC: The Urban Institute.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, *10*(12), 1–13.

Appendix A

Table A1.

Regression Results for Percentage Scoring 3+ (All AP Exams)

	Variable	*B*	*SE B*	*β*	*t*	*sig*
$R^2 = .08$						
Step 1	(Constant)	0.22	0.10		2.23	0.03
	Household Income	0.00	0.01	-0.01	-0.10	0.92
	AP Commitment	0.08	0.06	0.11	1.28	0.20
	Experience	0.29	0.08	0.29	3.50	0.00
$R^2 = .20$						
Step 2	(Constant)	-0.01	0.11		-0.08	0.93
	Household Income	0.01	0.01	0.08	1.04	0.30
	AP Commitment	-0.02	0.06	-0.02	-0.28	0.78
	Experience	0.25	0.08	0.25	3.15	0.00
	AP PD Events	0.20	0.04	0.40	4.60	0.00

Appendix B

Table B1.

Regression Results for Percentage Scoring 3+ on AP STEM Exams

	Variable	*B*	*SE B*	*β*	*t*	*sig*
$R^2 = .15$						
Step 1	(Constant)	0.07	0.14		0.51	0.61
	Household Income	0.00	0.01	-0.05	-0.48	0.63
	AP Commitment	0.08	0.09	0.08	0.84	0.40
	Experience	0.48	0.12	0.40	4.10	0.00
$R^2 = .31$						
Step 2	(Constant)	-0.23	0.14		-1.63	0.11
	Household Income	0.01	0.01	0.07	0.71	0.48
	AP Commitment	-0.05	0.09	-0.05	-0.54	0.59
	Experience	0.43	0.11	0.35	4.03	0.00
	AP PD Events	0.25	0.05	0.46	4.77	0.00

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

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

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: collegeboard.org/research

CollegeBoard

CollegeBoard

12b-5506