

RESEARCH IN REVIEW 2012-2



Research on College Board Programs and Services: Annotated Bibliography 2012



RESEARCH

College Board authors:

Carol Barry, Assistant Research Scientist

Jennifer Bausmith, Senior Director

Siobhan Cooney, Assistant Research Scientist

Vytas Laitusis, Research Scientist

Haifa Matos-Elefonte, Assistant Research Scientist

Mary McKillip, Assistant Research Scientist

Sunny Niu, Associate Research Scientist

Anita Rawls, Assistant Research Scientist

About the College Board

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

For further information, visit www.collegeboard.org.

© 2012 The College Board. College Board, Advanced Placement Program, AP, SAT, SpringBoard and the acorn logo are registered trademarks of the College Board. AP Potential, College Readiness System, EXCEerator, My College QuickStart, SAT Subject Tests and Skills Insight are trademarks owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners.

Visit the College Board on the Web: www.collegeboard.org.

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

RESEARCH

Contents

Methods	3
Advanced Placement Program®	5
EXCEerator™	15
Professional Development	19
PSAT/NMSOT®	23
SAT®	31
SpringBoard®	41
Glossary of Terms	43

Methods

This annotated bibliography contains summaries of research studies examining a number of College Board assessments and programs. To be included in the bibliography, each study needed to meet a number of criteria. First, articles must have been published (as a College Board research report, in an external journal, or as an ETS research report). Conference presentations or proceedings were not included because these materials often do not undergo as much scientific scrutiny as fully published papers. Also, the articles must represent research as either a validity study or an impact/evaluation study. All SAT® studies must have used samples from 2005 or later to account for the new test design, and Advanced Placement Program® (AP®) studies should have been published in the last 20 years. In addition, all studies included in the bibliography must have met at least a moderate level of scientific evidence, as defined by the U.S. Department of Education's Institute for Education Sciences. This definition includes well-designed, quasi-experimental studies and correlational designs that control for selection bias such as prior achievement levels.

All studies are grouped by assessment/program, and key findings that represent a synthesis across the studies are presented. Key findings refer to results of the most rigorous work to date regarding the College Board program in question. The strength of the research design and validity of the findings vary according to the program in question, and are discussed in more detail in the summaries in this report. Many additional research studies were reviewed for the various programs, but those that did not meet these criteria were not included in this bibliography.

Advanced Placement Program®

The Advanced Placement Program (AP), developed and offered by the College Board, is a program designed to offer high school students the opportunity to complete college-level work while still in high school. Specifically, students enroll in AP courses and complete the required course work. They then have the opportunity to take an end-of-course AP Exam. Subsequently, colleges and universities may grant the student course credit, advanced placement, or both on the basis of successful exam scores (i.e., 3 or above).

Key Findings:

- Key Finding 1:** AP examinees, particularly those taking two or three AP Exams, were more likely to attend a four-year institution than non-AP examinees.
Chajewski, Mattern, & Shaw (2011)..... 6
- Key Finding 2:** AP examinees, particularly those earning course credit or scoring a 3 or higher, attended more selective institutions and had higher college-level GPAs and higher freshman-year retention rates than non-AP examinees.
Hargrove, Godin, & Dodd (2008) 8
Murphy & Dodd (2009)..... 12
Mattern, Shaw, & Xiong (2009)..... 10
- Key Finding 3:** AP examinees, especially those scoring a 3 or higher, were more likely to graduate from college than non-AP examinees; the finding held across race/ethnicity and income groups.
Dougherty, Mellor, & Jian (2006)..... 7
Hargrove, Godin, & Dodd (2008) 8
- Key Finding 4:** AP examinees, especially those scoring a 5, earned higher grades in introductory and subsequent college-level course work, or in courses in the same subject area as the exam, compared to non-AP examinees.
Morgan & Klaric (2007)..... 11
Murphy & Dodd (2009)..... 12
Sadler & Tai (2007)..... 13
- Key Finding 5:** AP STEM examinees were more likely to graduate with a STEM major than to choose a nonscience major.
Tai, Liu, Almarode, & Fan (2010) 14
- Key Finding 6:** Students taking AP courses were not notably different from students not taking AP courses in first-semester college grades or in retention to the second year.
Klopfenstein & Thomas (2009)..... 9

Advanced Placement Program

Chajewski, M., Mattern, K., & Shaw, E. J. (2011). AP participation and college enrollment. *Education Measurement: Issues and Practice*, 30, 16–27.

Brief Description:

This study examined the relationship between AP Exam participation and enrollment in a four-year postsecondary institution. A positive relationship was expected, given that the primary purpose of offering AP courses is to allow students to engage in college-level academic work while in high school and potentially receive college credit by earning qualifying scores on the corresponding AP Exam.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The study finds that AP Exam participation was related to college enrollment, even after controlling for student demographic and ability characteristics and high school-level predictors. The study controlled for the number of AP Exam titles offered; average high school PSAT/NMSQT® critical reading, mathematics, and writing score; and students' gender, ethnicity, and prior academic performance as indexed by their composite PSAT/NMSQT score. The odds of attending a four-year postsecondary institution increased by at least 171% for all three AP participation groups (those taking either one AP Exam, two or three AP Exams, or four or more AP Exams) compared to students who took no AP Exams. The largest increase in odds was observed for those students taking two or three AP Exams. This group's odds of attending a four-year institution increased by 224%.

Research Design:

Logistic regression with statistical controls: PSAT/NMSQT; number of AP Exams offered at the high school; average high school PSAT/NMSQT critical reading, mathematics, and writing scores; gender; and race/ethnicity.

Sample Size and Characteristics:

High school seniors in the class of 2007 who took College Board tests were matched with National Student Clearinghouse data, and the final working sample includes 1,523,546 students from 17,142 high schools. Data were split into two subsets of randomly selected cases, so as to duplicate and thus provide cross-validation of the findings ($N_1 = 761,740$ and $N_2 = 761,806$). Estimates and statistics presented here are based on the primary (N_1) dataset.

Validity:

Moderate internal validity.

Moderate external validity.

Advanced Placement Program

Dougherty, C., Mellor, L., & Jian, S. (2006). *The relationship between Advanced Placement and college graduation* (National Center for Educational Accountability: 2005 AP Study Series, Report 1). Austin, Texas: National Center for Educational Accountability.

Brief Description:

This study explored the relationship between college graduation rates and student participation and achievement in AP courses and exams. Students were assigned to one of four categories based on their AP experience: (1) received a 3 or higher on at least one AP Exam; (2) took but did not received a 3 or higher on at least one AP Exam; (3) took AP course, not AP Exam; and (4) took no AP course or exam. Student ethnicity (African American, Hispanic, and white) and socioeconomic status (low-income versus non-low-income) were also taken into account.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The results indicated that students who took and received a 3 or higher on at least one AP Exam were more likely to graduate from college than students in the other three categories. Results indicated that 63% of students who earned a 3 or better on at least one AP Exam graduated from college in five years or less, whereas only 17% of students who did not participate in AP graduated in five years or less. After controlling for prior academic achievement, student-level variables, and school-level variables, the study found that students who earned a 3 or better on at least one AP Exam still had a higher probability of graduating from college than students in the other three categories. Specifically, African American and Hispanic students who achieved a 3 or higher on an AP Exam had a 28% higher probability of graduating from college than African American and Hispanic students who did not participate in AP. White students who received a 3 or higher on an AP Exam had a 33% higher probability of graduating from college than white students who did not participate in AP. If they received a 3 or higher on an AP Exam, low-income students had a 26% higher probability of graduating from college than students who did not participate in AP. Non-low-income AP students had a 34% increase in probability of graduating from college over non-low-income students who did not participate in AP. To control for school-related variables, a school-level regression analysis was also run, and similar results were found. Overall, students who received a 3 or higher on one or more AP Exams were more likely to graduate from college in five years or less than students who did not pass the exam or who did not participate in AP. Regression coefficients were statistically significant ($p < .001$) for Hispanic, white, low-income, and non-low-income students who received a 3 or higher on at least one AP Exam, but not for African American students.

Research Design:

Correlational/regression design with statistical controls.

Sample Size and Characteristics:

Statewide 1994 cohort of 67,412 eighth-grade students from Texas; students graduated from high school in 1998 and within 12 months enrolled in a Texas public college or university.

Validity:

Moderate internal validity.

Moderate external validity.

Advanced Placement Program

Hargrove, L., Godin, D., & Dodd, B. G. (2008). *College outcomes comparisons by AP and non-AP high school experiences* (College Board Research Report 2008-3). New York: The College Board.

Brief Description:

The study explored the performance of five cohorts of students (1998–2002) who graduated from public high schools in Texas and subsequently attended a public college or university in Texas. Analyses were broken into two phases. The first phase examined the impact of AP course taking on college outcomes by course for AP English Language, English Literature, Calculus, Biology, Chemistry, U.S. History, and Spanish Language. College outcomes included (1) first- and fourth-year grade point averages, (2) first- and fourth-year credit hours earned, and (3) four-year graduation status. Outcomes were compared across students who varied by three types of AP (course only, exam only, and both course and exam) and two types of non-AP (dual enrollment only and other course only) experiences in high school. Phase 2 looked at similar outcomes as Phase 1 but used aggregated AP data across subject areas.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Results of Phase 1 showed that students who took one or more AP courses and exams generally outperformed groups that had taken the AP course only, as well as groups that had taken other courses, for all college outcomes in most cohort years and for most subjects after using the SAT score and free or reduced-price lunch (FRPL) status as statistical controls. Students who took the AP course and exam tended to outperform (1) students who took the AP course only, and (2) students who only took dual enrollment and other courses on the following outcomes: GPA, college graduation rate, and “most credit hours earned” on most courses for most cohort years. In Phase 2, aggregated AP results showed that students who took one or more AP courses and exams significantly outperformed the “AP course only” and “other courses” groups on all college outcomes in all years, after using the SAT score and FRPL status as statistical controls. Students who took the AP course and exam significantly outperformed (1) students who took the AP course only, and (2) students who only took dual enrollment and other courses on the following outcomes: GPA, college graduation rate, and “most credit hours earned.”

Research Design:

MANOVA, ANOVA, and logistic regression analyses were conducted. Variations in comparison groups relating to prior achievement and SES were statistically accounted for in analyses.

Sample Size and Characteristics:

Five cohorts ranging from 58,899 to 62,709 Texas public high school graduates attending a Texas public institution of higher education through their first year; and four cohorts ranging from 38,907 to 42,199 students through their fourth year at a Texas public higher education institution.

Validity:

Moderate internal validity.
Moderate external validity.

Advanced Placement Program

Klopfenstein, K., & Thomas, K. (2009). The link between Advanced Placement experience and early college success. *Southern Economic Journal*, 75(3), 873–891.

Brief Description:

This paper examined the extent to which AP course taking predicts early college grades and retention in Texas. Controlling for a broad range of student, school, and curricular characteristics, the authors found that AP course taking does not reliably predict first semester college grades or retention to the second year when controlling for students' non-AP course work. They showed that failing to control for the students' non-AP curricular experience, particularly in mathematics and science, leads to positively biased AP coefficients.

Key Findings (including specifics related to statistical significance and magnitude of effect):

First, AP course credits have a statistically significant positive, albeit small, effect on the likelihood of persistence when student, family, high school characteristics, and colleges attended were considered. However, these positive and significant findings are not significant for all but Hispanic students when non-AP courses taken were also considered. Furthermore, the entire AP effect on retention for the average Hispanic student was driven by AP science; the Texas Pre-Freshmen Engineering Program (TexPREP), which suggests that taking AP science might have been a possible cause. Second, AP course credits also had statistically significant positive and large effects on *first year GPA* for white students when student, family, high school characteristics, and colleges attended were considered. The effects were much larger, albeit insignificant, for minorities. However, the magnitudes dropped substantially but remained significant for white students when non-AP courses taken were also considered. Furthermore, three of the most popular AP courses — AP Calculus, English, and History — had no effect on first-semester GPA for any group. Instead, AP Government, Economics, and Psychology were found to have positive effects. Such effects are most likely capturing some unobserved characteristics of the high schools that could offer an AP curriculum of such breadth and the students who choose to take AP courses outside the core.

Research Design:

Logit and OLS regressions with statistical controls: student, family, high school characteristics, non-AP courses taken, and universities attended.

Sample Size and Characteristics:

$N = 19,801$ whites, 3,126 blacks and 5,240 Hispanics as freshmen in Texas public universities in 1999.

Validity:

Moderate internal validity.

Moderate external validity.

Advanced Placement Program

Mattern, K. D., Shaw, E. J., & Xiong, X. (2009). *The relationship between AP Exam performance and college outcomes* (College Board Research Report 2009-4). New York: The College Board.

Brief Description:

This study explored the relationship between student achievement in AP courses and exams and college success. College success was defined as first-year GPA (FYGPA), institutional selectivity, and retention. Students were categorized as having no AP score, scoring a 1 or 2 on an AP Exam, or scoring a 3 or higher. The relationship between college success and AP achievement was investigated for four AP Exams: English Language, Biology, Calculus AB, and U.S. History.

Key Findings (including specifics related to statistical significance and magnitude of effect):

In general, higher performance on AP Exams corresponded to higher FYGPAs, attendance at more selective institutions, and higher second-year retention rates. Separate paired contrasts and ANCOVAs were run for each AP Exam; however, the results were comparable across the four exams. For the paired contrasts, students scoring a 3 or higher on an AP Exam had a statistically significantly ($< .01$) higher FYGPA, attended statistically significantly ($< .01$) more selective institutions, and were statistically significantly ($< .01$) more likely to return for their second year of college than students with no AP experience or students scoring a 1 or 2. The ANCOVA analyses were conducted to control for prior academic achievement (high school GPA [HSGPA] and SAT scores). After controlling for HSGPA and SAT scores, all paired contrasts remained statistically significant; however, the effect sizes decreased in magnitude. Effect sizes ranged from small (about .2) to moderate (about .5) and tended to be small for FYGPA and moderate for institutional selectivity and retention across the numerous comparisons. One difference from the paired contrasts that emerged in the ANCOVA was that students who received a 1 or 2 on the AP Exam had statistically significantly lower FYGPAs than non-AP students; however the effect size was minimal (< 0.1). A possible interpretation of these results is that AP courses better prepare students to be successful at the college level.

Research Design:

Paired contrasts and ANCOVA.

Sample Size and Characteristics:

Data were from the SAT Validity Study database and matched back to College Board databases. Students in this study came from a total of 99 institutions. The total number of students used in each analysis varied based on the AP Exam. The sample size ranged from 71,377 for AP Biology to 93,775 for AP U.S. History.

Validity:

Moderate internal validity.

Moderate external validity.

Advanced Placement Program

Morgan, R., & Klaric, J. (2007). *AP students in college: An analysis of five-year academic careers* (College Board Research Report 2007-4). New York: The College Board.

Brief Description:

This study examined the academic careers of students who took AP Exams compared to students who did not. Specifically, this study followed college students for five years, examining performance and subsequent course work, graduation rates, and college major.

Key Findings (including specifics related to statistical significance and magnitude of effect):

For most AP course areas, students with scores of 5 on AP Exams had statistically significantly higher grades in college-level intermediate courses than non-AP students, even after controlling for prior achievement ($p < 0.05$); students with scores of 3 or 4 on AP Exams had at least comparable grades in intermediate courses compared to non-AP students (for all but one subject area), even after controlling for prior achievement. The effect sizes varied across subject areas, but ranged in size from small to moderate. Further, AP students generally graduated significantly earlier than non-AP students, even after controlling for prior achievement ($p < 0.001$); this finding holds across gender and race/ethnic groups.

Additionally, with the exception of students who took the AP English Language and AP English Literature Exams, AP students took a greater number of courses in the academic area related to their AP Exam than did non-AP students. Although significance tests were not reported, the AP students took anywhere from approximately one to seven more courses in the subject area than non-AP students. Of those students who graduated, the percentages of AP students majoring in an area closely related to their AP Exam were higher than those for non-AP students. It is important to note, however, that no statistical control was included in this analysis.

Research Design:

Correlational/regression design with statistical controls for examination of performance and graduation rates; nonexperimental, for examination of subsequent course work and college major; comparison of existing groups.

Sample Size and Characteristics:

The sample consisted of 72,457 students attending 27 collegiate institutions. The 200 institutions that received the largest number of AP scores were identified and contacted, having first been categorized according to location, selectivity, and whether public or private. The 27 institutions in the study provided five years of academic data for students from the incoming class of 1994.

Validity:

Low to Moderate internal validity.
Moderate external validity.

Advanced Placement Program

Murphy, D., & Dodd, B. (2009). *A comparison of college performance of matched AP and non-AP student groups* (College Board Research Report 2009-6). New York: The College Board.

Brief Description:

This study examined the relationships between college performance (GPA and number of credit hours) and AP achievement. AP students were assigned to the following categories: (1) AP credit, meaning that students took an AP Exam and received college credit, or (2) AP no credit, meaning that students took an AP Exam but did not score well enough to receive college credit. AP students were compared to two different groups of students: non-AP students (i.e., students from the same cohort, but without AP credit) and concurrent students (i.e., high school students concurrently enrolled in a college course).

This study expands upon:

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

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

Key Findings (including specifics related to statistical significance and magnitude of effect):

Overall, AP students consistently outperformed non-AP students. Specifically, students who had taken at least one AP Exam took more credit hours in their first year, in their related subject area, and in college overall than non-AP students. AP students also had consistently higher GPAs in the subject area related to the AP Exam than non-AP students. Furthermore, students with credit statistically significantly outperformed AP students who did not receive credit for the AP, and also outperformed their matched counterparts without AP credit by having statistically significantly higher GPAs and number of credits. In the comparison of AP students to the concurrent group, AP students took more credit hours in their first year of college than the concurrent students.

Research Design:

Matched AP students to non-AP students and concurrent groups (on SAT scores and high school rank); two-way (AP status x credit status) MANOVA followed by univariate ANOVAs with planned comparisons.

Sample Size and Characteristics:

Data were obtained from students at The University of Texas at Austin. Four years (1998 to 2001) of entering student cohorts were analyzed to replicate results. Sample sizes were 5,910, 6,345, 6,467, and 6,219 for the 1998, 1999, 2000, and 2001 academic years, respectively.

Validity:

Moderate internal validity.

Low external validity: The sample consisted of students from one public university.

Advanced Placement Program

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

Brief Description:

This study examines the performance of college students taking introductory college biology, chemistry, and physics. Survey data from 8,594 students at 55 randomly chosen colleges and universities finds that those having received a score of three or higher on an AP science exam but retaken the introductory course earned somewhat higher college science grades, but not enough to assume prior mastery. Moreover, half of this performance difference appears to be related to demographics and high school course work, and not to students' AP course work.

Key Findings (including specifics related to statistical significance and magnitude of effect):

There are large differences between students who enroll in AP courses and others who don't. AP students earned college course grades about four points (out of a 100-point scale) higher than the course average of a B-. However, the AP advantages were cut in half when demographics and prior academic achievement were accounted for. Furthermore, once demographics and prior academic achievement were considered, students who scored a 3 on an AP Exam earned 1.5 points higher than non-AP students on average, students who scored 4 on an AP Exam earned 3.4 points higher than non-AP students on average, and students who scored a 5 on an AP Exams earned 4.6 points higher than non-AP students on average. The study discusses possible explanations for why students who achieved a high score on an AP Exam did not consistently attain levels of performance commensurate with stated College Board expectations: nonequivalency of the AP Exams and college science attainment measures, overgenerosity in AP Exam scoring, lack of sufficient content coverage in AP Exams, weak methodology in AP score validation, sample size of the current study, missing students who placed out in the current study, and bored AP students who were retaking introductory courses.

Research Design:

OLS (ordinary least squares) regressions with statistical controls: race/ethnicity, parental education, type of high school, mean education level at home ZIP code, SAT scores, highest mathematics course, mathematics grade, English grade, science grades, prep course taken and grades, honor course taken.

Sample Size and Characteristics:

$n = 8,594$ college freshmen in 2005.

Validity:

Moderate internal validity.

Moderate external validity.

Advanced Placement Program

Tai, R. H., Liu, C. Q., Almarode, J. T., & Fan, X. (2010). Advanced Placement course enrollment and long-range educational outcomes. In P. M. Sadler, G. Sonnert, R. H. Tai, & K. Klopfenstein (Eds.), *AP: A critical examination of the Advanced Placement program* (pp. 109–137). Cambridge, MA: Harvard Education Press.

Brief Description:

This study examined whether AP students (i.e., students who took an AP Exam in science or mathematics) were more likely to earn STEM-related college degrees than non-AP students, when controlling for achievement level of the students as well as the students' backgrounds.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Students who took AP Calculus were approximately four times more likely to graduate with a physical science major than with a nonscience major. Students who took an AP science exam were more than twice as likely to graduate with a life science major than with a nonscience major. These results were statistically significant ($p < 0.001$).

Research Design:

Regression design with statistical controls.

Sample Size and Characteristics:

The sample used in this study was a national sample of 3,938 students who attended eighth grade in 1988 and who graduated from a four-year college or university by 2000.

Validity:

Moderate internal validity.

Moderate external validity.

EXCELErator™

The EXCELErator™ Initiative is designed to help underrepresented groups gain access to the higher education pipeline, by increasing graduation rates, reducing dropout rates, increasing student participation in AP courses and exams, and increasing student participation in the SAT.

Key Findings:

Key Finding 1: Students in EXCELErator schools had higher graduation rates and lower dropout rates than students in non-EXCELErator schools.

Holtzman & Stancavage (2011) 16

Simon, Harnett, Tunik, Boyer, Cunningham, Nagler, Mastrorilli, & Thomas (2009)..... 17

Key Finding 2: Students in EXCELErator schools had higher SAT participation rates and no appreciable declines in performance compared to students in non-EXCELErator schools.

Holtzman & Stancavage (2011) 16

Simon, Harnett, Tunik, Boyer, Cunningham, Nagler, Mastrorilli, & Thomas (2009)..... 17

Key Finding 3: Students in EXCELErator schools had higher sophomore and much higher junior PSAT/NMSQT participation rates, although performance scores declined as participation increased compared to students in non-EXCELErator schools.

Holtzman & Stancavage (2011) 16

Simon, Harnett, Tunik, Boyer, Cunningham, Nagler, Mastrorilli, & Thomas (2009)..... 17

Key Finding 4: Students in EXCELErator schools had higher Advanced Placement® participation rates, although the percentage of students scoring a 3 or higher declined by the third year of implementation compared to students in non-EXCELErator schools.

Holtzman & Stancavage (2011) 16

Simon, Harnett, Tunik, Boyer, Cunningham, Nagler, Mastrorilli, & Thomas (2009)..... 17

Key Finding 5: The effect of EXCELErator on student state assessments is unclear and warrants further study, as some results suggest a decline for students in EXCELErator schools compared to students in non-EXCELErator schools.

Holtzman & Stancavage (2011) 16

Simon, Harnett, Tunik, Boyer, Cunningham, Nagler, Mastrorilli, & Thomas (2009)..... 17

EXCELerator

Holtzman, D. J., & Stancavage, F. (2011). *College readiness systems longitudinal evaluation: EXCELerator program impact, year 2 report.* Washington, DC: American Institutes for Research.

Brief Description:

This study examined the implementation and impact of the EXCELerator initiative in four school districts over the course of three years. All of the EXCELerator schools were matched to other regular (i.e., the primary focus of the school was not vocational, special, or alternative education), noncharter, currently open schools.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Results of this evaluation showed that schools that implemented the EXCELerator Initiative had higher graduation rates and lower dropout rates than students in non-EXCELerator schools, with the effect increasing over time. Students in EXCELerator schools had higher Advanced Placement participation rates, though the percentage of students scoring 3 or higher declined by the third year of implementation compared to students in non-EXCELerator schools. Students in EXCELerator schools had higher SAT participation rates and no appreciable declines in performance compared to students in non-EXCELerator schools. The effect of the EXCELerator program on student state assessments is unclear and warrants further study as some results suggest a decline for students in EXCELerator schools compared to students in non-EXCELerator schools.

Research Design:

Quasi-experimental (strong). Comparative interrupted time series.

Sample Size and Characteristics:

In the first year of the EXCELerator Initiative, there were 12 high schools. In the second year, 16 additional high schools began with EXCELerator. In the third year, 21 high schools and 45 middle schools joined the initiative. EXCELerator schools were matched on the composite scores for implementation across the years while controlling for school enrollment size, percentage of African American students, percentage of Hispanic students, and urbanicity.

Validity

Moderate internal validity: Schools were matched on prior achievement.

Low external validity: Schools were drawn from only four school districts.

EXCELerator

Simon, A.J., Harnett, S., Tunik, J., Boyer, D., Cunnington, M., Nagler, E., Mastrorilli, T., & Thomas, L. (2009, May). *EXCELerator schools evaluation final report*. New York: Metis Associates.

Brief Description:

This study examined the implementation and impact of the EXCELerator Initiative in 27 secondary schools in five urban school districts after two years of implementation. The implementation aspect of the study included surveys of district coaches, school administrators, teachers, and other pedagogical staff in participating EXCELerator schools as data sources. The impact aspect of the study examined student achievement after implementation of the initiative.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The schools faced various challenges, such as variability in levels of implementation from school to school as well as competing district policies, values, and priorities. Overall, many of the key components of the EXCELerator Initiative were implemented as planned. The schools were building a college-going culture with rigor, high expectations, access, opportunity, and inspiration of hearts and minds. Schools that implemented the EXCELerator Initiative did not have any significant effects on student achievement. However, the EXCELerator schools did significantly increase the number of students participating in college readiness exams (such as the PSAT/NMSQT, AP, and SAT). The magnitude of the effect of participating in the EXCELerator program for these schools was small. This may have been an effect of the complexity of the implementation, as described in the full report.

Research Design:

The impact aspect of the study used a quasi-experimental design (strong) with various statistical matching techniques to create groups of nonparticipating comparison schools, and/or comparison students from nonparticipating schools, that were as similar as possible to EXCELerator students and schools.

Sample Size and Characteristics:

In the first year of implementation, there were 11 schools from three districts. In the second year of implementation, there were 16 schools from four districts, and one of the four districts also funded the participation of four additional schools (also known as “mirror schools”) within the district. EXCELerator schools and the mirror schools had high proportions of minority students from low-income households and often had larger-than-average populations of special education and English language learners (ELLs). The students were matched on gender, race/ethnicity, ELL, special education, and baseline achievement scores.

Validity

Moderate internal validity: Matched schools were developed for all EXCELerator schools using a variety of indicators, including percentage of students passing the state exam.

Low external validity: A small number of schools implemented EXCELerator for one or two years, though the sample was diverse.

Professional Development

College Board Professional Development (PD) offers hundreds of workshops each year, training thousands of teachers with more than 1,400 trained consultants. The College Board uses this capacity and experience to manage and deliver both the required professional development and project evaluation to a wide range of districts.

Key Findings:

- Key Finding 1:** Students of teachers engaged in PD were more diverse and demonstrated lower levels of prior achievement in the year following PD, but they were still able to maintain consistent levels of AP performance.
Bausmith & Laitusis (in press) 20
- Key Finding 2:** The number of AP Exams taken in a school was significantly influenced by the number of AP teachers in the school.
Patterson & Laitusis (2006) 21

Professional Development

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

Brief Description:

The AP Achievement Institute I (APAI I) is a four-day professional development program offered to teachers and administrators. The program is designed to help teachers develop effective AP instructional strategies for a diverse student body, and to help district, school, and curriculum leaders strengthen the district's infrastructure to support AP students and teachers.

This study examined the impact of APAI I on student AP achievement on English language arts and social studies course exams. Students' AP Exam scores from the 2009 and 2010 administrations were examined for all participating teachers' students who took the AP English Language, English Literature, U.S. History, World History, European History, Comparative Government and Politics, U.S. Government and Politics, or Human Geography Exams. For students whose teachers participated in APAI I, achievement on AP Exams was compared before and after the APAI I professional development.

Key Findings (including specifics related to statistical significance and magnitude of effect):

After the APAI I, more Hispanic students of lower prior ability in absolute counts were enrolled in AP courses taught by the APAI teachers than before APAI I (574 and 509, respectively), although the increase was not statistically significant. Students in 2009-10 also took more AP Exams than in 2008-09 (601 and 523, respectively), with much of this increase represented by Hispanic students. Despite this increase, there were no statistically significant differences between the students before and after APAI I. In summary, students taking AP in the second year were more diverse, representing lower levels of prior achievement while maintaining consistent AP Exam performance.

Research Design:

Nonexperimental ANCOVA design. Multiple cohorts with control for prior achievement using state assessment data.

Sample Size and Characteristics:

There were 12 teachers who received the APAI I professional development and taught one of the specified AP courses in both the 2008-09 and 2009-10 school years.

Validity

Moderate internal validity: Students' prior achievement was statistically controlled for by using state assessment data.

Low external validity: Teachers in the sample were limited to participation in PD in a single district.

Professional Development

Patterson, B., & Laitusis, V. (2006). *AP professional development in Florida: Effects on AP Exam participation* (College Board Research Note RN-27). New York: The College Board.

Brief Description:

The study examined the impact of teacher professional development for AP courses on student AP Exam participation. It specifically looked at the impact of teacher participation in two types of professional development: the AP Summer Institute and AP half-day workshops. The analysis was completed at the school level and controlled for socioeconomic level of the school district, school size, and AP program size.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The number of AP Exams taken in a school was significant and positively impacted by the number of AP teachers in the school for both cohorts at the 0.05 level. The number of days of AP professional development in the three years prior to the academic years was also significant and positively impacted the number of AP Exams in a school at the 0.05 level in 2001–2002 and the 0.10 level in 2003–2004. The number of AP Exams taken in a school slightly decreased as total student enrollment increased, and this difference was significant at the 0.05 level for the 2001–2002 cohort and significant at the 0.10 level for the 2003–2004 cohort. Finally, the number of AP Exams taken in a school significantly decreased as the number of days of AP Summer Institute increased in the three years prior to 2001–2002. The effect size for the above model was 0.72 for the 2001–2002 cohort and 0.78 for the 2003–2004 cohort. Of the variables used in this study, the number of AP teachers in the school explained most of the variance in number of AP Exams taken, with partial R-squared values of 0.69 and 0.77 for 2001–2002 and 2003–2004, respectively

Research Design:

Nonexperimental. Multiple cohort design.

Sample Size and Characteristics:

In 2001–2002, 333 AP Exam takers from 317 of Florida's public schools, along with 309 teachers attending the AP Summer Institute, were included. In 2003–2004, 351 AP Exam takers from 327 of Florida's public schools, along with 343 teachers attending the AP Summer Institute, were included.

Validity

Moderate internal validity: Statistical controls for district SES, school size, and AP program size.

Low external validity: The study was specific to schools with teachers attending the AP Summer Institute and the students in those schools taking AP Exams within one state.

PSAT/NMSQT®

The Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) is a measure of student aptitude for college cosponsored by the College Board and the National Merit Scholarship Corporation. The exam is taken by students in high school, most of whom are sophomores or juniors. Students taking the exam receive access to various tools to help improve academic skills and identify potential colleges and universities (PSAT/NMSQT Skills Insight™ and My College QuickStart™, respectively). In addition, junior test-takers are automatically entered in to the National Merit Scholarship Competition. Schools use the PSAT/NMSQT scores to help identify students for participation in AP courses via AP Potential™ and use the Summary of Answers and Skills (SOAS) as guidance for instruction.

Key Findings:

- Key Finding 1:** Student's PSAT/NMSQT writing scores were strongly related to their AP scores.
Ewing, Camara, Millsap, & Milewski (2007)..... 24
- Key Finding 2:** PSAT/NMSQT scores were moderately to strongly related to multiple measures of student academic achievement.
Milewski & Sawtell (2006)..... 27
- Key Finding 3:** Students who took the PSAT/NMSQT and then retaken the PSAT/NMSQT or then took the SAT tended to receive higher scores on the subsequent exam.
Proctor & Kim (2010)..... 28
- Key Finding 4:** PSAT/NMSQT scores can be used to identify students who are on track toward college readiness.
Proctor, Wyatt, & Wiley (2010) 29
Tierney, Bailey, Constantine, Finkelstein, & Hurd (2009)..... 30
- Key Finding 5:** PSAT/NMSQT scores are valid measures for selection of National Merit Scholars and are predictive of first year of college GPA.
Marini, Mattern, & Shaw (2011a) 25
Marini, Mattern, & Shaw (2011b) 26

PSAT/NMSQT

Ewing, M., Camara, W., Millsap, R., & Milewski, G. (2007). *Updating AP Potential expectancy tables involving PSAT/NMSQT writing* (College Board Research Note RN-35). New York: The College Board.

Brief Description:

Research has shown a moderate-to-strong correlation between PSAT/NMSQT scores and AP Exam scores. The purpose of the study was to recompute the expectancy tables between the PSAT/NMSQT and AP for AP Exams that included writing scores after changes were made in 2006 to the writing scale of the PSAT/NMSQT.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Probabilities of an AP score greater than or equal to 3 and greater than or equal to 4 by PSAT/NMSQT score are provided. Statistical significance and magnitude of effect were not applicable. Results showed that one or more PSAT/NMSQT scores were moderately to strongly correlated to scores on all AP Examinations, with four exceptions: AP German Language, Spanish Language, Studio Art: Drawing, and Studio Art: 2-D Design.

Research Design:

Nonexperimental design. To recompute the expectancy tables, the old PSAT/NMSQT scores from the 2000 and 2001 test administrations were placed on the new 2006 PSAT/NMSQT score scale using the conversion table displayed in Table 3. This conversion table was applied exactly as shown except for the conversion from 80 (on the old scale) to 77–80 (on the new scale), wherein the midpoint value of 78.5 was used for the new scale. Once the conversion table was applied, the expectancy tables were recomputed following the same procedures that were outlined in previous research.

Sample Size and Characteristics:

For this study, the data analyzed included sophomores and juniors who completed the PSAT/NMSQT in October 2000 or October 2001 and took one or more AP Exams 19 months later (i.e., either in May 2002 or May 2003) ($n = 1,035,696$).

Validity:

Low internal validity: Correlational data was used, but there were no controls for selection bias.

High external validity: Data was based on a national sample.

PSAT/NMSQT

Marini, J., Mattern, K., & Shaw, E. (2011a). *Examining the linearity of the PSAT/NMSQT-FYGPA relationship* (College Board Research Report 2011-7). New York: The College Board.

Brief Description:

Overall, the research sought to provide evidence to support the current use of the PSAT/NMSQT to identify National Merit Scholars and as a predictor of student success in college. In particular, this study extended previous research by examining the linear relationship between the PSAT/NMSQT Selection Index (critical reading + mathematics + writing) and student performance in college as measured by the first-year GPA (FYGPA). Moreover, the authors wanted to validate the use of the exam in differentiating between high-performing students.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The study confirmed that the relationship between PSAT/NMSQT and FYGPA was linear and increasing significantly with slight deviations at the lower end of the score scale; such deviations do not impact National Merit Scholarship decisions. The PSAT/NMSQT significantly differentiated among high-scoring students in terms of FYGPA. Using graphical analysis and a power polynomial approach, the study found that most absolute differences between the linear and quadratic model predictions at each PSAT/NMSQT selection index ($n = 0.001$), but it did not add any practical significance for inclusion.

Research Design:

Nonexperimental design. This correlational study examined the relationship between PSAT/NMSQT scores and FYGPA. The authors used a graphical analysis and the power polynomial test to determine whether the relationship was linear or curvilinear.

Sample Size and Characteristics:

The study consisted of first-time, first-year students entering 177 colleges and universities throughout the United States in the fall of 2006, 2007, or 2008. Students in the sample had valid PSAT/NMSQT scores and a valid FYGPA ($n = 444,193$). The sample mostly consisted of white (64.4%) females (54.5%) who were attending very large (55.1%), public (69.8%), moderately selective (62.7%) institutions.

Validity:

Low internal validity: No controls were used to infer causality.

High external validity: The study was based on a nationally representative sample of students enrolled in four-year institutions.

PSAT/NMSQT

Marini, J., Mattern, K., & Shaw, E. (2011b). *Examination of college performance by National Merit Scholarship program recognition level* (College Board Research Report 2011-10). New York: The College Board.

Brief Description:

This study examined the aptness of the selection process used for National Merit Scholarship winners, which includes the initial screening criteria of PSAT/NMSQT scores. If students perform well on the PSAT/NMSQT, they are entered into the scholarship competition, where they have the possibility of earning a variety of awards (e.g., commended student, semifinalist, or finalist). The study compared the college performance of National Merit Scholars with that of other college students who did not receive an award. College performance was measured by FYGPA and retention to the second year of college.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The National Merit Scholarship Program level of recognition was positively related to PSAT/NMSQT scores, HSGPA, SAT scores, FYGPA, and second-year retention. There was a statistically significant difference between recognition level and FYGPA, and second-year retention. The statistically significant differences between recognition levels had a small to medium effect size (0.038).

Research Design:

Weak quasi-experimental design. This study classified students by National Merit Scholarship recognition level and then compared the average FYGPA and second-year retention rate. ANOVA was used to test for differences in FYGPA between the five recognition levels with subsequent effect sizes. A chi-squared statistic was used to test for group differences on the categorical retention rate variable.

Sample Size and Characteristics:

The study consisted of first-time, first-year students entering 177 colleges and universities across the United States in the fall of 2006, 2007, or 2008. Students in the sample participated in the PSAT/NMSQT and SAT, and a self-reported HSGPA, FYGPA, and second-year retention information ($n = 386,011$).

Validity:

Low internal validity: No controls were used to infer causality.

Moderate external validity: Study participants were limited to students with high school and higher education data available to the College Board.

PSAT/NMSQT

Milewski, G., & Sawtell, E. A. (2006). *Relationships between PSAT/NMSQT scores and academic achievement in high school* (College Board Research Report 2006-6). New York: The College Board.

Brief Description:

This study investigated relationships between scores on the verbal (as it was then known), mathematics, and writing sections of the PSAT/NMSQT, the PSAT/NMSQT composite (verbal + mathematics + writing scores), and the following indicators of academic achievement in high school: years of study, participation in specific mathematics and English language arts courses, HSGPA, academic intensity, and participation and performance in AP courses.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The results indicated that students with more years of study (across all academic areas) obtained higher mean PSAT/NMSQT scores. Correlations between verbal, mathematics, writing, and composite PSAT/ NMSQT scores and HSGPA were medium to large. The correlations with PSAT/NMSQT scores (by section and for the composite) were medium to large for academic intensity in mathematics and science ($r = .45$ to $.59$), medium to large for academic intensity in humanities and social science ($r = .44$ to $.52$), and large for overall academic intensity ($r = .53$ to $.61$). This relationship was also supported by the large multiple correlation between PSAT/NMSQT composite scores and the two academic intensity variables (mathematics/science and humanities/social science), which was $.62$ ($R^2 = .38$).

Research Design:

Nonexperimental design. Correlations between PSAT/NMSQT scores and the various academic achievement variables.

Sample Size and Characteristics:

The analysis began with a data set that contained all of the students who graduated in May or June 2002 and participated in at least one College Board program. This data set was reduced to include only the students who took the PSAT/NMSQT in October 2000 during their junior year and the SAT sometime before they graduated in May or June 2002. The reduced data set that was ultimately used for this study was composed of 857,375 students.

Validity:

Low internal validity: Correlational data but no controls for selection bias.
High external validity: Data was based on a national sample.

PSAT/NMSQT

Proctor, T., & Kim, Y. (2010). *Score change for 2007 PSAT/NMSQT test-takers: An analysis of score changes for PSAT/NMSQT test-takers who also took the 2008 PSAT/NMSQT test or a spring 2008 SAT test (College Board Research Note: RN-41).* New York: The College Board.

Brief Description:

The purpose of this paper was to provide information about how students' scores changed when they retook the PSAT/NMSQT as juniors, or took the SAT in the spring after they took the PSAT/NMSQT as juniors.

Key Findings (including specifics related to statistical significance and magnitude of effect):

On average, sophomore PSAT/NMSQT test-takers who repeated the PSAT/NMSQT as juniors improved their critical reading score by 3.3 points, their mathematics score by 4.0 points, and their writing score by 3.3 points. For students who took the PSAT/NMSQT as sophomores and again in October 2008 as juniors, the correlations between the scores were 0.85 for critical reading, 0.87 for mathematics, and 0.84 for writing. On average, junior PSAT/NMSQT test-takers who took their first SAT as a junior received SAT critical reading scores that were 17.5 points higher, SAT mathematics scores that were 15.8 points higher, and SAT writing scores that were 22.5 points higher. For students who took the PSAT/NMSQT as juniors and their first SAT as juniors, the correlations between the PSAT/NMSQT scores and the SAT scores were 0.87 for critical reading, 0.88 for mathematics, and 0.83 for writing.

Research Design:

Nonexperimental design. To study the change in scores from PSAT/NMSQT to PSAT/NMSQT or SAT, analyses were performed that examined the percentage of students who obtained ranges of changes in scores, average scores, score change, and correlations across testing occasions. These analyses were disaggregated by gender and racial/ethnic groups.

Sample Size and Characteristics:

For the analysis of sophomore-to-junior PSAT/NMSQT score changes, 710,595 examinees were selected who took the PSAT/NMSQT both as sophomores in October 2007 and as juniors in October 2008, and had valid scores on all three sections of the PSAT/NMSQT for both testing occasions. For the analysis of the junior PSAT/NMSQT to junior SAT score changes, 585,947 examinees were selected who took the PSAT/NMSQT as juniors in October 2007 and took their first SAT in March, May, or June of 2008.

Validity:

Low internal validity: Correlational data but no controls for selection bias.

High external validity: Data was based on a national sample.

PSAT/NMSQT

Proctor, T., Wyatt, J., & Wiley, A. (2010). *PSAT/NMSQT indicators of college readiness*. (College Board Research Report 2010-4). New York: The College Board.

Brief Description:

This study extended the work of Wiley, Wyatt, & Camara (2010), who developed an indicator of college readiness using HSGPA, SAT scores, and an academic readiness indicator to create a PSAT/NMSQT test score benchmark. This benchmark was used to identify students who were on track toward college readiness when they completed high school.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Students who scored 55 or above as 11th-grade PSAT/NMSQT test-takers had a very high likelihood of becoming college ready on that SAT section. The same pattern held true for students who scored 55 or higher on the 10th-grade PSAT/NMSQT. On the overall test, juniors who obtained a composite score of 160 or above had a very high likelihood of eventually meeting the SAT benchmark of college readiness. Sophomores who obtained a 155 or above had a very high likelihood of meeting the junior PSAT/NMSQT benchmarks and being on track to be college ready by high school graduation. Overall, 45% of 2008 10th-grade PSAT/NMSQT test-takers met the 11th-grade PSAT/NMSQT benchmarks, and 55% of 11th-grade PSAT/NMSQT test-takers went on to meet or exceed the SAT benchmark.

Research Design:

Nonexperimental design. In the first analysis, benchmark scores for 10th- and 11th-grade PSAT/NMSQT test-takers were created. In the case of the 11th-grade PSAT/NMSQT benchmark scores, logistic regression was used to obtain the minimum junior PSAT/NMSQT score associated, with a 65% probability of obtaining the SAT college readiness benchmark. In the second analysis, contingency tables were established to show the percentage of students who went on to meet or exceed the SAT college readiness benchmark by PSAT/NMSQT score band.

Sample Size and Characteristics:

First, to analyze the score changes between the junior PSAT/NMSQT and the junior SAT, 585,947 examinees were selected who took the PSAT/NMSQT (as juniors) in October 2007 and their first SAT in March, May, or June of 2008. The second data set was composed of 710,595 students who completed the PSAT/NMSQT in 2007 of their sophomore year and 2008 of their junior year. Last, students who had valid scores on all three test sections were selected. This resulted in 1,517,231 students in 10th grade and 1,545,856 students in 11th grade.

Validity:

Low internal validity: Correlational data but no controls for selection bias.
High external validity: Data was based on a national sample.

PSAT/NMSQT

Tierney, W. G., Bailey, T., Constantine, J., Finkelstein, N., & Hurd, N. F. (2009). *Helping students navigate the path to college: What high school can do* (NCEE 2009-4066). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>.

Brief Description:

This report is one of many “practice guides” developed by the Institute for Education Sciences (IES), the research branch of the U.S. Department of Education. This guide is intended to help schools and districts develop practices to increase access to higher education. It contains specific steps on how to implement recommendations that are targeted at school- and district-level administrators, teachers, counselors, and related education staff. The guide also indicates the level of research evidence demonstrating that each recommended practice is effective.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The expert panel that developed this review recommends “utilize[ing] assessment measures throughout high school so that students are aware of how prepared they are for college, and assist them in overcoming deficiencies as they are identified” (p. 20) such as the PSAT/NMSQT and SAT tests.

Research Design:

Literature review.

Sample Size and Characteristics:

Not applicable.

Validity:

Low internal validity.

High external validity.

SAT®

As the nation's most widely used college admission test, the SAT is the first step toward higher education for students of all backgrounds. It's taken by more than two million students every year and is accepted by virtually all colleges and universities.

Key Findings:

Key Finding 1: SAT scores (based on revised SAT) were significantly related to college grade point average, and when used in conjunction with HSGPA, provided incremental predictive power over high school grades alone.

<i>Mattern & Patterson (2011a)</i>	32
<i>Mattern & Patterson (2011b)</i>	33
<i>Mattern & Patterson (2011c)</i>	34
<i>Patterson & Mattern (2011)</i>	39

Key Finding 2: SAT scores were significantly related to FYGPA for males, females, and students across racial/ethnic subgroups, although there were some variations in predictive validity.

<i>Mattern, Patterson, Shaw, Kobrin, & Barbuti (2008)</i>	38
---	----

Key Finding 3: SAT scores (based on the revised SAT) were positively related to college retention rates.

<i>Mattern & Patterson (2011d)</i>	35
<i>Mattern & Patterson (2011e)</i>	36
<i>Mattern & Patterson (2011f)</i>	37

SAT

Mattern K., & Patterson B. (2011a). *Validity of the SAT for predicting second-year grades: 2006 SAT validity sample* (College Board Statistical Report 2011-1). New York: The College Board.

Brief Description:

This study evaluated the validity of the SAT for predicting second-year college GPA.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Unadjusted correlations with second-year cumulative GPA ranged from 0.26 to 0.34 for the three SAT sections (moderate effect sizes), while corrected correlations ranged from 0.49 to 0.53. Unadjusted correlations with second-year GPA ranged from 0.23 to 0.31 for the three SAT sections (moderate effect sizes), while corrected correlations ranged from 0.44 to 0.49. Of the three SAT sections, the writing section had the highest correlation with both second-year GPA and second-year cumulative GPA.

When controlling for HSGPA, positive relationships remained between SAT scores and second-year GPA and cumulative GPA. The incremental validity of SAT scores over HSGPA was 0.07 and 0.08 for second-year GPA and second-year cumulative GPA, respectively. The best predictor of both second-year GPA and second-year cumulative GPA was a combination of HSGPA and SAT scores.

Research Design:

This study used a correlational (nonexperimental) design. The study looked at SAT and HSGPA, and their correlations with second-year college GPA and second-year cumulative GPA. Second-year cumulative GPA was defined as the average of course grades earned during the student's first and second years of college.

Sample Size and Characteristics:

The study sample included second-year students at 66 four-year colleges and universities. Students who had valid SAT scores, HSGPA, and first- and second-year GPAs were included, resulting in a sample size of 80,958.

Validity:

Moderate internal validity: The study used a correlational design and corrected correlations for restriction of range.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Mattern K., & Patterson B. (2011b). *Validity of the SAT for predicting third-year grades: 2006 SAT validity sample* (College Board Statistical Report 2011-3). New York: The College Board.

Brief Description:

This study evaluated the validity of the SAT for predicting third-year college GPA.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Unadjusted correlations with third-year cumulative GPA ranged from 0.27 to 0.36 for the three SAT sections (moderate effect sizes), while corrected correlations ranged from 0.50 to 0.56. Unadjusted correlations with fourth-year GPA ranged from 0.18 to 0.27 for the three SAT sections (small to moderate effect sizes), while corrected correlations ranged from 0.38 to 0.43. Of the three SAT sections, the writing section had the highest correlation with both third-year GPA and third-year cumulative GPA.

When controlling for HSGPA, positive relationships remained between SAT scores and third-year GPA and cumulative GPA. The incremental validity of SAT scores over HSGPA was 0.06 and 0.09 for third-year GPA and third-year cumulative GPA, respectively. The best predictor of both third-year GPA and third-year cumulative GPA was a combination of HSGPA and SAT scores.

Research Design:

This study used a correlational (nonexperimental) design. The study looked at SAT and HSGPA, and their correlations with third-year college GPA and third-year cumulative GPA. Third-year cumulative GPA was defined as the average of course grades earned at any time from the first year through the third year.

Sample Size and Characteristics:

The study sample included third-year students at 60 four-year colleges and universities. Students who had valid SAT scores, HSGPA, and first- through third-year GPAs were included, resulting in a sample size of 63,736.

Validity:

Moderate internal validity: The study used a correlational design and corrected correlations for restriction of range.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Mattern, K., & Patterson, B. (2011c). *Validity of the SAT for predicting fourth-year grades: 2006 SAT validity sample* (College Board Statistical Report 2011-7). New York: The College Board.

Brief Description:

This study evaluated the validity of the SAT for predicting fourth-year college GPA.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Unadjusted correlations with fourth-year cumulative GPA ranged from 0.26 to 0.35 for the three SAT sections (moderate effect sizes), while corrected correlations ranged from 0.48 to 0.54. Unadjusted correlations with fourth-year GPA ranged from 0.15 to 0.24 for the three SAT sections (small to moderate effect sizes), while corrected correlations ranged from 0.33 to 0.39. Of the three SAT sections, the writing section had the highest correlation with both fourth-year GPA and fourth-year cumulative GPA.

When controlling for HSGPA, positive relationships remained between SAT scores and fourth-year GPA and cumulative GPA. The incremental validity of SAT scores over HSGPA was 0.04 and 0.08 for fourth-year GPA and fourth-year cumulative GPA, respectively. The best predictor of both fourth-year GPA and fourth-year cumulative GPA was a combination of HSGPA and SAT scores.

Research Design:

This study used a correlational (nonexperimental) design. The study looked at SAT and HSGPA, and their correlations with fourth-year college GPA and fourth-year cumulative GPA. Fourth-year cumulative GPA was defined as the average of course grades earned at any time from the first year through the fourth year.

Sample Size and Characteristics:

The study sample included fourth-year students at 55 four-year colleges and universities. Students who had valid SAT scores, HSGPA, and first- through fourth-year GPAs were included, resulting in a sample size of 56,939.

Validity:

Moderate internal validity: The study used a correlational design and corrected correlations for restriction of range.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Mattern K., & Patterson, B. (2011d). *The relationship between SAT scores and retention to the fourth year: 2006 SAT validity sample* (College Board Statistical Report 2011-6). New York: The College Board.

Brief Description:

This study examined the relationship between performance on the SAT and fourth-year college retention rates.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Results indicate that SAT scores correlated positively with fourth-year retention, with 88% of high performers (SAT total scores ranging from 2100 to 2400) returning but only 42% of low performers (SAT total scores ranging from 600 to 890) returning. The mean SAT score (critical reading + mathematics + writing) for students returning for their fourth year of college was 1727 compared to 1611 for nonreturners; this pattern of SAT means for returners and nonreturners held across subgroups. HSGPA also correlated positively with fourth-year retention. Furthermore, the positive relationship between SAT scores and retention rates still held within HSGPA levels. For students with a HSGPA of "A," those who had SAT total scores from 900 to 1190 had an average retention rate of 63%, whereas those with SAT total scores from 2100 to 2400 had an average retention rate of 89%. Although retention rates varied by subgroups and institutional characteristics, these differences were minimized when taking SAT performance into account. No tests of statistical significance were reported.

Research Design:

Nonexperimental design. Mean SAT scores were computed and then compared for returners (students who returned for the fourth-year) and nonreturners. Retention rates were computed by student academic characteristics (SAT, HSGPA) as well as by student characteristics (gender, race/ethnicity, parental income, and highest parental education).

Sample Size and Characteristics:

The sample used for this study consisted of 78,640 students attending 59 colleges and universities in the U.S. Students in the sample were first-time, first-year students who entered college in fall 2006.

Validity:

Moderate internal validity: The study used a correlational design.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Mattern, K., & Patterson, B. (2011e). The relationship between SAT scores and retention to the second year: 2007 SAT validity sample (College Board Statistical Report 2011-4). New York: The College Board.

Brief Description:

This study examined the relationship between performance on the SAT and second-year college retention rates. This research replicated a previously conducted study (Mattern & Patterson, 2009) but used a more current sample.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Results indicated that SAT scores correlated positively with second-year retention, with 95% of high performers (SAT total scores ranging from 2100 to 2400) returning but only 65% of low performers (SAT total scores ranging from 600 to 890) returning. The mean SAT score (critical reading + mathematics + writing) for students returning for their second year of college was 1699, compared to 1562 for nonreturners; this pattern of SAT means for returners and nonreturners held across subgroups. HSGPA also correlated positively with second-year retention; however, the positive relationship between SAT scores and retention rates still held within HSGPA levels. For example, among students with a HSGPA of "A," those who had SAT total scores from 900 to 1190 had an average retention rate of 61%, whereas those with SAT total scores from 2100 to 2400 had an average retention rate of 96%. Although retention rates varied by subgroups and institutional characteristics, these differences were minimized when taking SAT performance into account. No tests of statistical significance were reported.

Research Design:

Nonexperimental design. Mean SAT scores were computed and then compared for returners (students who returned for the second year) and nonreturners. Retention rates were computed by student academic characteristics (SAT, HSGPA) as well as by student characteristics (gender, race/ethnicity, parental income, and highest parental education).

Sample Size and Characteristics:

The sample used for this study included of 164,362 students attending 109 colleges and universities in the U.S.; institutions were diverse with respect to region, public/private, size, and selectivity. Students in the sample were first-year students entering college in fall 2007.

Validity:

Moderate internal validity: The study used a correlational design.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Mattern, K., & Patterson, B. (2011f). *The relationship between SAT scores and retention to the third year: 2006 SAT validity sample* (College Board Statistical Report 2011-2). New York: The College Board.

Brief Description:

This study examined the relationship between performance on the SAT and third-year college retention rates.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Results indicated that SAT scores correlated positively with third-year retention, with 93% of high performers (SAT total scores ranging from 2100 to 2400) returning but only 42% of low performers (SAT total scores ranging from 600 to 890) returning. The mean SAT score (critical reading + mathematics + writing) for students returning for their third year of college was 1722, compared to 1599 for nonreturners; this pattern of SAT means for returners and nonreturners held across subgroups. HSGPA also correlated positively with third-year retention; however, the positive relationship between SAT scores and retention rates still held within HSGPA levels. For example, among students with a HSGPA of "A," those who had SAT total scores from 900 to 1190 had an average retention rate of 68%, whereas those with SAT total scores from 2100 to 2400 had an average retention rate of 94%. Although retention rates varied by subgroups and institutional characteristics, these differences were minimized when taking SAT performance into account. No tests of statistical significance were reported.

Research Design:

Nonexperimental design. Mean SAT scores were computed and then compared for returners (students who returned for the third year) and nonreturners. Retention rates were computed by student academic characteristics (SAT, HSGPA) as well as by student characteristics (gender, race/ethnicity, parental income, and highest parental education).

Sample Size and Characteristics:

The sample consisted of 89,381 students attending 66 colleges and universities in the U.S.; these institutions were diverse with respect to region, public/private, size, and selectivity. Students in the sample were first-time, first-year students who entered college in fall 2006.

Validity:

Moderate internal validity: The study used a correlational design.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Mattern, K., Patterson, B., Shaw, E., Kobrin, J., & Barbuti, S. (2008). *Differential validity and prediction of the SAT* (College Board Research Report 2008-4). New York: The College Board.

Brief Description:

This study examined the extent to which the revised SAT displayed differential validity and differential prediction for various subgroups.

Key Findings (including specifics related to statistical significance and magnitude of effect):

The results, in general, showed smaller, though still significant, correlations between SAT scores and FYGPA for African American and Hispanic students compared to those of white students. A similar pattern emerged from HSGPA, with higher correlations for white students compared to those of minority groups. The correlation between SAT scores and FYGPA was generally slightly higher for females than for males. There was a similar pattern for HSGPA, with a larger correlation for females compared to males. Finally, for best language, the correlation between SAT scores and FYGPA was highest for students whose best language was English, in the middle for students who spoke English and another language, and lowest for students whose best language was something other than English; again, a similar pattern emerged for HSGPA.

In terms of differential prediction, FYGPA tended to overpredict for males and African American, American Indian, and Hispanic students but underpredict for females and students whose best language was not English. Similar patterns of over- and underprediction were seen when using HSGPA to predict FYGPA. Using a combination of SAT and HSGPA tended to result in the least amount of over- and underprediction of FYGPA.

Research Design:

Nonexperimental design. Differential validity was assessed by computing the correlation between SAT scores and HSGPA with FYGPA by subgroup. Correlations were corrected for restriction of range. To assess the extent to which the SAT, as well as HSGPA, exhibited differential prediction, regression equations within each institution were calculated.

Sample Size and Characteristics:

This study included students entering 110 four-year colleges and universities in fall 2006. The sample was representative of the 2006 SAT College-Bound Seniors cohort ($n = 151,316$).

Validity:

Moderate internal validity: The study used a correlational design and corrected correlations for restriction of range.

Moderate to high external validity: The sample of colleges and students was very diverse.

SAT

Patterson, B., & Mattern, K. (2011). *Validity of the SAT for predicting first-year grades: 2008 validity sample* (College Board Statistical Report 2011-5). New York: The College Board.

Brief Description:

This study evaluated the validity of the SAT for predicting FYGPA. The study provided a more recent update to that published previously regarding the validity of the SAT for predicting FYGPA.

Key Findings (including specifics related to statistical significance and magnitude of effect):

Unadjusted correlations with the FYGPA ranged from 0.28 to 0.35 for the three SAT sections (moderate effect sizes). Corrected correlations with FYGPA ranged from 0.48 to 0.52 for the three SAT sections. Of the three sections of the SAT, the writing section had the highest correlation with FYGPA.

When controlling for HSGPA, a positive relationship remained between SAT scores and FYGPA. Similar to previous research, the increment in predictive validity attributable to SAT scores over HSGPA was 0.07. The best predictor of FYGPA was a combination of HSGPA and SAT scores.

Research Design:

This study used a correlational (nonexperimental) design. The study looked at SAT and HSGPA, and their correlations with FYGPA.

Sample Size and Characteristics:

The study sample included first-year students at 129 four-year colleges and universities. Students who had valid SAT scores, HSGPA, and FYGPA were included, resulting in a sample size of 173,963.

Validity:

Moderate internal validity: The study used a correlational design and corrected correlations for restriction of range.

Moderate to high external validity: The sample of colleges and students was very diverse.

SpringBoard®

As the foundation of the College Board’s College Readiness System™, SpringBoard® infuses rigor, sets high expectations, and expands access and opportunity for all students. SpringBoard provides culturally and personally relevant activities designed to engage students in problem solving, academic discourse, and critical analysis. This unique approach to individualized learning provides teachers with a road map for opening the doors to a bright future for all students.

Key Findings

Key Finding 1: SpringBoard students at all grade levels performed significantly higher than non-SpringBoard students on both the English language arts and mathematics sections of the FCAT.

Westat (2008).....42



SpringBoard

Westat (2008). *SpringBoard Longitudinal Evaluation: Report 2008 Executive Summary.* Rockville, MD: Author.

Brief Description:

This study evaluated the impact of the SpringBoard program on student achievement. The evaluation included a system-wide teacher survey comparing SpringBoard and non-SpringBoard teachers to assess teachers' attitudes and opinions regarding conditions at their school, as well as SpringBoard implementation patterns. Teachers who participated in SpringBoard training in 2005 or 2006 were recruited for this study. The evaluation also included case studies of selected SpringBoard districts and schools, and a preliminary analysis of student achievement related to SpringBoard participation in selected districts.

Key Findings (including specifics related to statistical significance and magnitude of effect):

SpringBoard students at all grade levels performed significantly higher than non-SpringBoard students on both the English language arts and mathematics sections of the FCAT ($p < .01$). In SpringBoard English Language Arts, the estimated annual effect was 25.5 to 37.3 (Florida developmental scale score) units, or 2.5 months to more than a year of additional growth per year. A student who stayed in SpringBoard for three years could be expected to grow about the same extra amount per year, which could add up to an additional three years of achievement, or a total of six years of growth in three years. In SpringBoard Mathematics, the estimated effect was between 4.4 to 19.4 scale score units, or 0.4 to 4.5 months of additional growth per year.

Research Design:

Quasi-experimental design (weak). The data were analyzed using a repeated-measures, multilevel modeling approach in which the growth in students' test score for any given year was predicted based on their gender, race, free/reduced-price lunch participation, participation in SpringBoard, a variable to measure trends over time, and two variables measuring school characteristics (percentage eligible for free/reduced-price lunch and the percentage of students who were minority). The major variable of interest was participation in SpringBoard and its ability to explain differences in student achievement after some other differences in the groups were accounted for.

Sample Size and Characteristics:

Four districts in the state of Florida submitted student-level achievement data from the state assessment (FCAT) from both SpringBoard students and non-SpringBoard students. The reading data from Florida included 419,709 students and 1,370,654 test scores over seven years. The reading test scores represented 134,426 SpringBoard observations and 1,236,228 non-SpringBoard comparison observations. The mathematics test scores represented 113,944 SpringBoard observations and 1,240,298 non-SpringBoard observations.

Validity:

Moderate internal validity: The study was lacking strong statistical controls.

Low external validity: The sample size was based on data from four districts in one state, not a national sample.

Glossary of Terms

Control Variable — A control variable is held constant in an analysis in order to assess or clarify the relationship between two other variables. For example, to better understand the relationship an intervention in high school predicts SAT assessments scores, a researcher would want to control for achievement prior to the intervention (such as through PSAT/NMSQT assessment scores). It is not to be confused with the creation of a matched sample, as is done in strong quasi-experimental research designs (see below).

Correlation — A correlation means that one variable is related to another. It can suggest a possible causal relationship, but a correlation does not mean causation. A negative correlation suggests that as one variable increases, the other decreases; for example, as a student’s achievement test scores increase, the likelihood that they will need remediation before beginning college courses decreases. A positive correlation suggests that as one variable increases, the other increases: As a student’s achievement test scores increase, his or her likelihood of being accepted and succeeding in college increases. In statistics, a correlation is often demonstrated with a Pearson correlation coefficient (r) with a scale that ranges from -1 (for a perfect negative linear relationship), 0 (for no relationship) to 1 (a perfect positive linear relationship). For example, in the study of Proctor and Kim (2010), students’ test scores on the PSAT/NMSQT critical reading section during the sophomore year in 2007 were positively correlated with their test scores on this same section of the PSAT/NMSQT during the junior year in 2008 (p. 28). The correlation coefficient is 0.85 in this example.

Selection Bias — Researchers must concern themselves with two types of selection bias in any study that is not an experiment.

1. Sampling bias: Is the sample they are studying representative of the larger population in which they are interested? For example, are high school senior SAT takers in a given state representative of the entire high school senior population of that state? If the sample is substantially different from the larger population in ways that would lead to differences in the outcomes of interest, then there is a problem with selection bias; in other words, the “selection” of the sample is biased.
2. Within-sample selection bias: Given an outcome of interest (such as college enrollment), are students who receive a certain “treatment” (e.g., take an AP course in high school) different than students in the “control” group (students who do not take an AP course) in ways that make the treatment group more likely to get the outcome of interest with or without the “treatment”? For example, are students who have high achievement scores at the start of high school more likely to enroll in AP courses and enroll in college?

Statistical Significance — Statistical significance refers to the probability that a result occurred by chance alone. A result is generally considered “statistically significant” if the probability it occurred by chance alone is 5% or less.

Effect Size/Magnitude of Effect — Sometimes statistically significant effects are found in a study simply because the sample size is very large and it is therefore difficult to know whether the significance has any practical meaning. Calculating the effect size helps researchers understand the magnitude of the effects seen because it takes into account the sample size and variation in the outcome measure across the population(s).

Research Design

Nonexperimental — Includes studies that are descriptive, comparative, or correlational. No causal conclusions about relationships can be drawn from these types of studies, although they can suggest relationships that warrant further research.

Descriptive: This includes studies that report the number and percentage of students taking different types of assessments. It includes studies that report descriptive outcomes of surveys (i.e., percentage of respondents who answered a question in a certain way).

Comparative or correlational: Studies that examine the relationship between two variables but do not account for other factors that may impact this relationship.

Quasi-Experimental Design (QED) — A research design in which subjects are assigned to “treatment” (that is, they receive the intervention being studied) and “comparison” groups through a process that is not random. QED studies may be classified as weak or strong, depending on the level of rigor with which the treatment and comparison groups are truly similar before the treatment occurs. Strong QED designs can address problems of within-sample selection bias that weak QED designs (and comparative and correlational designs) cannot.

Weak QED: The design controls for certain background characteristics of treatment and comparison groups.

Strong QED: The treatment and comparison groups must be similar in terms of the outcomes being studied before the treatment is applied (includes comparative interrupted time series studies and other analyses using matched comparison groups, such as propensity score matching).

Validity

External — Is the study generalizable? A study is considered to be externally valid if we could expect to see the same results with a different sample of students or schools; thus, the validity does not depend on the methods used or the way characteristics are measured and considered in the analysis. Also, the closer the sample of a study mirrors the population of interest, the more externally valid it is; therefore, larger samples tend to have higher validity (though this depends on the size of the population of interest).

Internal — Can we be confident of the accuracy of the relationships found? A study is considered to be internally valid if there is confidence that the outcomes are caused by the variable considered in the analysis, not by some other measured or unmeasured variable or variables. (Experimental designs and strong QED designs have high internal validity; descriptive and correlational studies do not.)

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

