College Board Research Report No. 2005-2

Everyone Gains: Extracurricular Activities in High School and Higher SAT® Scores

Howard T. Everson and Roger E. Millsap

College Entrance Examination Board, New York, 2005

Howard T. Everson is vice president of Academic Initiatives and chief research scientist at the College Board.

Roger E. Millsap is a professor in the psychology department at Arizona State University.

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 4,700 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three and a half million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com.

Additional copies of this report (item #040481375) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is $15. Please include $4 for postage and handling.

Printed in the United States of America.

Contents

Abstract

This report presents evidence that links participation in extracurricular activities (ECAs) in high school with higher SAT Reasoning Test™ (SAT®) scores. Using structural equation models (SEMs) with latent means, we analyzed data from a national sample of college-bound high school students. A series of structural equation models—isolating the influence of ECAs on SAT verbal and mathematics scores—were fit simultaneously to eight subgroups (disaggregated by both gender and ethnicity) of high school students. The SEMs analyses suggest: (1) that often observed group differences in SAT scores shrink, and (2) that students' levels of participation in ECAs in high school are related to meaningful gains in SAT scores, once the influences of socioeconomic background and academic achievement are controlled statistically. These analyses suggest that participation in ECAs benefits minority and socioeconomically disadvantaged students as much as, or more than, economically advantaged white students. These findings support the conclusion that supplementary education programs benefit minorities and disadvantaged high school students who are often ill-served by traditional academic curricula.

Introduction

There is little debate that public high schools in the United States need improvement. Student achievement suffers in many high schools, and innovation and change are required to address the challenge of improving learning across the curriculum, particularly when it comes to the achievement gap between minority and nonminority students. The contributors to this report offer a variety of alternative approaches, all under the heading of supplementary education programs. But do these programs and interventions actually lead to increases in student achievement? What is the quality of the evidence in support of these programs and activities? Indeed, if these activities are to be expanded, as some have argued elsewhere in this report, then rigorous scientific evidence will have to be developed and made available to policymakers, parents, students, and other stakeholders.

Marsh and Kleitman (2002), writing in the *Harvard Educational Review*, present a persuasive case for the efficacy of extracurricular activities. They conclude, for example, that

> Whereas most school activities exacerbate the already substantial gap in academic outcomes between socioeconomically advantaged and disadvantaged students, ESAs (extracurricular school activities) appear to actually reduce this inequality gap. Although the ESA benefits generalize widely, the benefits tend to be larger, certainly not smaller, for more disadvantaged students (p. 508).

Others (see, for example, Camp, 1990; Gerber, 1996; Holland and Andre, 1987; Holloway, 2000; Marsh, 1992) have reached similar conclusions. Despite these efforts, policymakers may be constrained in the current environment because many of these studies may not meet the standard of rigorous scientific evidence promulgated by the U.S. Department of Education.

Recently, the U.S. Department of Education's Institute of Education Sciences released a set of guidelines intended to help educators identify educational interventions that are backed by strong evidence of effectiveness (U.S. Department of Education, 2003). Comparison group studies using closely matched groups are cited, albeit somewhat tentatively, by the Institute of Education Sciences as providing reliable scientific evidence of effectiveness. The research in this report, which used a matched comparison group design, provides, we argue, strong preliminary evidence that extracurricular activities—a not so uncommon form of educational intervention—contribute to student performance on important, high-stakes tests such as the SAT. We reached this conclusion by analyzing the SAT verbal and mathematics scores of more than 480,000 high school students matched on a number of socioeconomic and academic characteristics.

Extracurricular Activities

High school kids call them 3:05ers. There are many of them in our nation's high schools. These are the kids who fly out the door and away from school as soon as their last class ends at 3 o'clock in the afternoon. No after-school or supplemental programs for them. Even if offered, no clubs, no band practices, no athletic teams or other extracurricular programs keep them past the last bell. Many have questioned just how strongly these students are engaged with school. But more and more, there is a large and growing number of high school students who enjoy staying after school. These students participate in any number of extracurricular activities (ECAs) including music, art and drama clubs, intercollegiate and intramural athletics, and other academic and vocational clubs. These activities are voluntary, and students do not receive grades or academic credit for them (Holloway, 2000).

The purpose of the study reported here was to examine the effects of participation in a range of extracurricular activities in high school on students' all-important SAT scores, while controlling for the effects of other important factors such as socioeconomic background, high school achievement, gender, and ethnicity. In this study, we extend previous research in this area and attempt to address some of the methodological issues that would increase policymakers' skepticism about the effects of extracurricular activities on academic achievement. We have organized the report to provide detail on our sample of college-bound students, to explain our model-based analytic approach, describe briefly the data and the model-fitting framework commonly referred to as structural equation modeling, and, finally, to present the results of our analysis. We conclude with a discussion of the implications of our research for educational practice and addressing the achievement gap.

Method

The analyses we present in this report examined the relationships among and influences of socioeconomic background, academic achievement, and extracurricular activity levels on high school students' SAT verbal and mathematics scores. We looked at these relationships across eight subgroups of students, based on ethnic group membership and gender. The explanatory models we developed were tested against the SAT verbal and mathematics scores of students in all eight subgroups. In the following sections we describe our sample and data sources, as well as our model-based approach.

Our Sample

Our sample comes from a subset of college-bound seniors who took the SAT during their junior or senior year of high school, and who graduated from high school in 1995. This cohort of students had mean SAT verbal and mathematics scores of 504 and 506, respectively. They represent about 41 percent of all the high school seniors in the United States in 1995. Girls make up about 54 percent of this group, and the cohort is largely white (69 percent), with 11 percent African American, 8 percent Asian American, 4 percent Mexican American, 4 percent other Latinos, 1 percent Native American, and 3 percent who marked "other" when noting their race or ethnicity. Since our analyses focus on subgroup differences in SAT scores, Table 1, at right, displays the mean SAT verbal and mathematics scores disaggregated by race/ethnicity and gender for this cohort of college-bound students.

The magnitude of group differences in SAT scores is clear. Males outperform females in mathematics, and white and Asian American students, in general, score higher on both the verbal and mathematics SAT tests than African American and Hispanic students.

Table 1

SAT Verbal and Mathematics Scores by Gender and Race/Ethnicity

	White	*Asian American*	*African American*	*Hispanic*
Males				
SAT-M	551	577	444	495
SAT-V	537	533	443	484
Females				
SAT-M	515	543	427	462
SAT-V	534	531	448	478

Table 2

Number of Students Responding to the Survey by Gender and Ethnicity

	Males	*Females*	*Total*
White	170,270	212,412	382,682
African American	18,411	27,644	46,055
Asian American	12,333	13,732	26,065
Hispanic	13,026	16,666	29,692
Total	214,040	270,454	484,494

Table 3

High School Achievement Variables

HSAVG	High School Grade Point Average
CRANK	High School Class Rank
ARTGR	GPA in Art and Music Courses
SOCGR	GPA in Social Science and History
ENGR	GPA in English Courses
LANGR	GPA in Foreign Language Courses
MATHGR	GPA in Mathematics Courses
SCIGR	GPA in Natural Science Courses

Table 4

Extracurricular Activities Variables

ACTCNT	Number of Extracurricular Activities (pursued for at least 3 years)
APCNT	Number of AP® Exams Intended
HNRCNT	Number of Honors Classes Taken
ENGCNT	Number of Literature Experiences
COMPCNT	Number of Computer Experiences
ARTCNT	Number of Art, Music, and Theater Experiences

Table 5

Family Socioeconomic Background Variables

FATHED	Father's Education Level
MOTHED	Mother's Education Level
FAMINC	Combined Parental Income

Data Source

When students register with the College Board to sit for the SAT they complete a lengthy questionnaire, answering 43 questions about their high school courses, participation in a sweep of extracurricular activities, academic achievement levels (i.e., grades), parental education, family income, and their race or ethnicity (see www.collegeboard.org for a copy of the Student Descriptive Questionnaire). Responses to these questions formed much of the data for this study. Table 2 shows the number of students enrolled in public high schools, disaggregated by race/ethnicity and gender, who responded to all the relevant questions in the College Board survey, thus comprising our sample. This subset of more than 480,000 students provided the data used in the subsequent analyses.

The College Board questionnaire, for example, asked students to indicate the total number of years they took high school courses in specific subject areas, and to report their grade point average (GPA) on a scale of A to F for each academic subject. The data elements we used to model academic achievement are presented in Table 3.

Similarly, students indicated their participation in a range of extracurricular activities. Table 4 provides the complete list of variables we used to model participation in academic and nonacademic extracurricular activities while in high school.

Students in our sample also reported their best estimates of annual family income in increments of $10,000, with reporting categories ranging from a low of $10,000 to a maximum of $100,000 or more per year. In addition, they reported the highest level of education attained by both parents. These three variables were used to model students' socioeconomic backgrounds. See Table 5.

In addition to these self-reported measures, each student's SAT verbal and mathematics scores (reported on a scale from 200 to 800) were used as the outcome measures in our analyses.

A Model-Based Approach

Our approach relied on structural equation modeling (SEM). The use of SEMs in educational and psychological research has steadily grown since the late 1970s. This approach is particularly well-suited for our study because of the large number, 19 in all, of observed variables in our model, and our interest in linking participation in a variety of extracurricular activities with performance on the SAT.

SEM analyses include three broad stages: specifying the model that relates the variables one to another, estimating the parameters of the model, and, finally, estimating how well the model fits the empirical data, i.e., how well the theoretical model replicates the empirical correlations between and among the variables included in the model. Specifying the model requires us to translate the theory we wish to test, in this case the relationship between ECAs and SAT scores, into a particular structural model that can be derived and tested given the empirical data on hand. Thus, the resulting models should be testable, i.e., supported or refuted by the data. During the parameter estimation stage we used the College Board data to obtain estimates of the model parameters—the coefficients calibrating the relationships among the variables—that are optimal according to any one of several statistical estimation methods. And finally, to evaluate the fit of our model, we used the derived parameter estimates to examine whether the hypothesized model can reproduce the covariation found in the empirical data.

Model Specification

Model specification in SEM begins with a theory about the relationships among the variables under study. For convenience, a distinction is commonly made between the measurement and structural portions of the model. The measurement model describes the relations between the measured variables and the latent variables (the underlying factor that accounts for the relationships among the observed measures) hypothesized to underlie these measured variables. The 19 measured variables in this study were believed to be represented well by three latent variables (i.e., socioeconomic background, high school achievement, and extracurricular activities), and two observed variables (i.e., the SAT verbal and mathematics test scores). This hypothetical measurement model of the three latent variables is depicted in Figure 1.

The boxes represent measured variables from the questionnaire completed by all the students taking the SAT, and the circles represent the common factors or latent variables we hypothesized were underlying 17 of the measured variables. The directed arrows linking the latent and measured variables indicate which measured variables are hypothesized as measures of each latent variable. In our model, each measured variable is linked to a single latent variable.

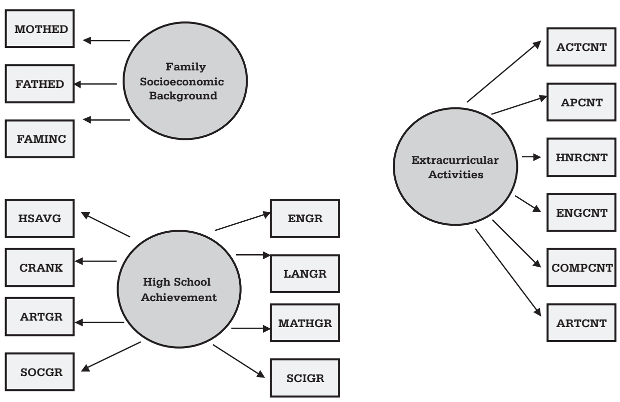

Figure 1. Measurement model of latent variables.

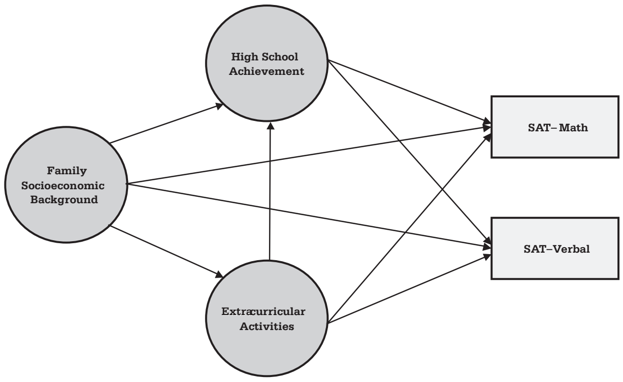

Figure 2. Structural model of relationship among measures of family background, high school achievement, and extra-curricular activities on SAT verbal and math scores.

The structural portion of the model specifies the directional relations among the latent variables or among the measured variables if no latent variables are included. The choice of which latent variables are linked directly by paths, and which are related indirectly, is based on theory. The hypothesized structural portion of our model is depicted in Figure 2.

Our first consideration when defining the structural model was the choice of which latent variables are exogenous, and considered causally prior to the other latent variables. Our model asserts that performance on the SAT is dependent upon high school achievement and participation in extracurricular activities, which, in turn, are dependent on socioeconomic background. We also hypothesized that socioeconomic background directly influences SAT scores, as well (Everson and Millsap, 2004; Everson, Millsap, and Diones, 1995; Everson and Michna, 2004).

Results

As noted earlier, the SEM approach centers on two steps: validating the measurement model and fitting the structural model. We divided our sample of N=484,494 into eight groups based on the factorial combination of gender with the four ethnic classifications: Asian American, African American, Hispanic, and white. In this initial stage, a nested series of structural equation models were fit to the data, permitting tests of ethnic and gender invariance of the structural and measurement models. Mean structures were used to allow for tests for group differences in mean SAT performance after adjusting for the influences of the other factors. The use of means permitted tests for group differences in mean SAT performance after adjustment for modeled influences—socioeconomic background, achievement, and extracurricular activities—on the SAT scores. We proceeded by fitting a series of measurement and structural models, starting with the most general—least constrained—model and moving on to more specific models that assume invariant relationships that may or may not hold across all the subgroups of students.

Our initial model, Model 1, fit a five-factor model to the 19 measured variables within each of the eight student groups. This is the most general model, and in some ways the least interesting, though it does suggest that the underlying factors do account for the relationships among the observed variables. Table 6 presents the fit information for Model 1, and for all of the other subsequent models we fit to these data. Three fit statistics are given for each model: the chi-square fit statistic and degrees of freedom, the root mean square error of approximation (RMSEA) (Steigler and Lind, 1980), and the comparative fit index (CFI) (McDonald and Marsh, 1990).

The “null” model in Table 6 is the independence model needed for the calculation of the various fit indices, but is of little intrinsic interest. Model 1 provides a good fit, as indicated by the RMSEA and CFI indices. Again, these fit indices support the claim that five factors are sufficient to represent the 19 measured variables in all groups.

Model 2 constrains Model 1 by requiring that each measured variable load on one and only one underlying factor—constraining the cross-loading of an observed variable to more than one underlying factor. In Model 2, however, two of the five factors are presumed to have nonzero loadings only for the SAT variables, with one factor representing SAT-V and the other SAT-M. This model, in short, specifies that the observed SAT verbal and mathematics scores represent the latent factors of verbal reasoning and mathematical reasoning. Table 6 gives the fit statistics for Model 2. The

Table 6

Fit Statistics for Competing Structural Models

Model	*Chi-Square*	*df*	*RMSEA*	*CFI*
Null	3,862,416	1,368		
(1) Measurement model, unconstrained loadings	87,831	704	.045	.98
(2) Measurement model, congeneric	286,713	1,152	.064	.93
(3) Measurement model, invariant loadings	307,133	1,250	.062	.92
(4) Structural model, invariant loadings and paths	309,145	1,306	.062	.92
(5) Structural model, invariant loadings, paths, and intercepts	374,822	1,334	.068	.90
(6) Structural model, invariant loadings, paths, partial invariant on intercepts	346,637	1,326	.065	.91

overall fit is acceptable, as indicated by the RMSEA and CFI indices. Further analyses suggest that the slight loss of fit in Model 2 relative to Model 1 results primarily from the sharp distinction between the extracurricular activities and high school achievement factors. We suspect, for example, that some of the observed variables in these factors may have nonzero loadings on both factors—academic achievement and ECAs, rather than only on one of the two. The variable HNRCNT, for example, which counts the number of honors courses taken, was constrained statistically to load only on the extracurricular activities factor in Model 2, but we expected, nevertheless, that it has nonzero loadings on both the high school achievement and extracurricular activities factors. The results, obviously, suggest that while the academic achievement–extracurricular activities distinction may not be as sharp as we had believed initially, the five-factor structure is, nevertheless, a good approximation of the relationships in our data.

Model 3 further restricts Model 2 by forcing the loadings to be invariant across all eight subgroups. Apart from these invariance constraints, all other parameter matrices and estimates were expected to have the same structure as in Model 2. Table 6 shows that the constraints introduced in Model 3 do not degrade the fit of the model relative to Model 2, suggesting that the factor loadings (or functional weights) can be presumed to be invariant without substantial loss of fit of the model. Clearly, the fit indices of these first three models provide confidence that a five-factor measurement model with invariant factor loadings fit the data reasonably well.

The next models we fit added restrictions on the relationships among the five underlying factors, creating a structural model that we then combined with the measurement model. Again, see Figure 1 and Figure 2 as representations of these hypothetical relationships.

Model 4, then, examines the invariance restrictions on the coefficients (the strength of relationship) of the paths among and between the underlying factors. The comparison between Models 3 and 4 is a test of these invariance restrictions, i.e., that the purported causal relationships among the latent factors are more or less the same across all the subgroups. Table 7 presents the estimates of the structural intercepts for the SAT-M and SAT-V in each of the eight groups resulting from Model 4.

Table 7

Intercept Estimates from Model 4

	White	*Hispanic*	*African American*	*Asian American*
Males				
SAT–M	205	200	154	231
SAT–V	240	241	198	240
Females				
SAT–M	152	155	120	177
SAT–V	218	222	187	214

The key comparison is between group differences on the intercept estimates in Table 7, and group differences in the corresponding SAT means in Table 1. To illustrate, Table 1 reveals a difference of more than 100 points on the SAT-M between white and African American males. Further analyses show that the differences on the SAT-M are only 50 points for the same groups. This reduction in score differences represents the impact of the statistical adjustment for socioeconomic background, high school achievement, and extracurricular activities. The remaining difference of 50 points represents a group difference in SAT-M performance that is unexplained after adjustment for these three explanatory factors.

To take another example, the SAT-M and SAT-V mean differences between Hispanic and white males is essentially eliminated by the adjustment for the explanatory factors. In contrast, the gender difference on the SAT-M in Table 1 within the white, Hispanic, and Asian American groups is smaller than the SAT-M gender difference within these groups. Here the adjustment for the explanatory factors served to widen the gender difference, rather than eliminate it. The basis for this finding lies in the complex pattern of gender differences on the socioeconomic background, high school achievement, and extracurricular activities variables in Table 1. Females score higher on the high school achievement variables, and score higher on most of the extracurricular activities variables. The results suggest that after adjusting for the females' higher scores on these academic variables, we expect to see an even larger gender difference on the SAT-M than is found in the unadjusted population of males and females. The higher academic performance of the females in the unadjusted population serves to reduce the average gender difference on the SAT-M. Once this higher academic performance is attenuated via statistical adjustment, the SAT-M score difference in favor of males is increased.

Model 5 is identical to Model 4 with the exception that now the structural intercepts (the latent means of the SAT-V and SAT-M scores) are constrained to be invariant across all groups. This restriction suggests that after adjusting for socioeconomic background, high school achievement, and extracurricular activities, there are no group differences in expected or mean SAT scores. If Model 5 fits well, we may be able to explain the observed group differences in SAT performance in terms of group differences on the three underlying factors in our theory-based structural model. Obviously, Model 5 is particularly important for this reason, and its fit must be carefully examined.

Table 6 shows that while there is some global loss of fit associated with the invariance restrictions on the structural intercepts noted above, the overall fit is still reasonably good. A more detailed look at the fit of Model 5 suggested that it does not fit perfectly, but the important question is whether the fit is below an acceptable threshold in any of the groups. Further inspection of the SAT-V and SAT-M means within each of the eight groups revealed that the African American SAT means are substantially lower than would be predicted by Model 5. The discrepancy is around a half of a standard deviation (50 points) for both males and females, on both the SAT-M and SAT-V score scales. This discrepancy suggests that while the invariance restrictions imposed in Model 5 may not degrade the fit of the model globally, the restrictions are too stringent for the African American group. The key conclusion is that after adjustment for the socioeconomic background, academic achievement, and extracurricular activities latent variables, the African American students—both males and females—continue to score lower on the SAT-V and SAT-M than would be expected by Model 5.

The final model, Model 6, relaxes the invariance restrictions on the structural intercepts for African American males and females in Model 5. The remaining six ethnic/gender groups are restricted to invariance on the structural intercepts, as in Model 5. All other parameter restrictions in Model 5 are retained in Model 6. As shown in Table 6, the global fit indices improve slightly with the loosening of the restrictions on the structural intercepts for the African American groups. Given this improvement, Model 6 is preferred to Model 5.

Parameter Estimates

The standardized path coefficient estimates are based on Model 6. The standardization uses a common metric across the eight groups, permitting the creation of a single set of standardized estimates. As revealed by these estimates, which are presented as the path coefficients in Figure 3, the direct impact of extracurricular activities on the SAT-V and SAT-M is larger than that of academic achievement levels.

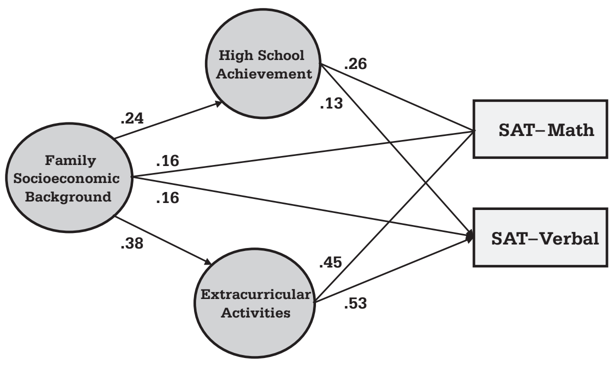

Figure 3. Path model of effects of family background, high school achievement, and extracurricular activities on SAT verbal and math scores.

As we can see, a unit change, a standard deviation difference, in the extracurricular activities latent variable produces a 45-point increase in SAT mathematics scores, and a 53-point change in SAT verbal scores. In contrast, a unit change in socioeconomic status (roughly equivalent to a $20,000 increment in family income, for example) only results in a 16-point increase in SAT verbal and mathematics scores. The direct impact of students' socioeconomic background variable is relatively small, but has indirect relations to the SAT in the path model also. The squared multiple correlations for the SAT-V ($R^2 = .49$) and SAT-M ($R^2 = .57$), an index of the explanatory power of the structural model, suggest that the structural model provides a reasonably good fit to the SAT score data. Though the squared multiple correlations are somewhat lower for African Americans, the structural model in Figure 3, in general, accounts for about half of the variance in the SAT scores.

Conclusion

The application of structural modeling techniques to these data from the College Board's SAT was revealing, and supports several important conclusions. First, the measurement model holds across all eight subgroups of college-bound students. Second, the structural model is useful and informative for representing the relationships among and between the socioeconomic and academic factors and SAT scores, irrespective of race and gender. Third, with the exception of the African Americans, subgroup differences on the SAT-M and SAT-V are explained by group differences on the socioeconomic background factor, the academic achievement factor, and the extracurricular activities factor. The somewhat poorer fit of the model for African American students suggests that the model may have to be expanded to include variables and indicators of the quality of the high schools that students attend. Given the historical patterns of racial segregation in housing and educational opportunities in the United States, it would be surprising indeed if one generic model would fit this particular group of American students.

The results of this study, we contend, are methodologically stronger than most other studies of the influences of extracurricular activities on high-stakes tests (see Marsh and Kleitman, 2002, for a more detailed discussion of the rigor of research on this topic). Like other investigators (Camp, 1990; Gerber, 1996; Holloway, 2000; Marsh and Kleitman, 2002), our study provides compelling evidence from the SAT, a national high-stakes test, that participation in extracurricular activities provides all students—including students from disadvantaged backgrounds, minorities, and those with otherwise less than distinguished academic achievements in high school—a measurable and meaningful gain in their college admissions test scores. The important reasoning abilities measured by tests like the SAT, evidently, are indeed developed both in and out of the classroom. To paraphrase Marsh and Kleitman (2002), participation in extracurricular activities in high school appears to be one of the few interventions that benefit low-status, disadvantaged students—those less well served by traditional educational programs—as much as or more than their more advantaged peers.

Although we cannot know definitively from our research whether the relationship between participation in extracurricular activities and the observed increases in SAT scores is causal, the reliance on strong measurement models and large, closely matched subgroups of students certainly suggests we may be detecting a meaningful effect in these data. Further research is needed, obviously, to address the limitations inherent in this study—self-selection effects, motivation, and the qualitative differences of schools, among them. Nevertheless, given the growing body of evidence suggesting the beneficial effects associated with participating in extracurricular activities at both the elementary and secondary education levels, we remain optimistic.

References

Camp, W. (1990). Participation in student activities and achievement: A covariance structural analysis. *Journal of Educational Research*, *83*, 272–278.

Everson, H.T., & Michna, G. (2004, January). *Is the SAT a wealth test? The influence of family income, ethnicity, and high school achievement on SAT scores.* Paper presented at a colloquium sponsored by the Institute for Urban and Minority Education, Teachers College, Columbia University, New York, NY.

Everson, H.T., & Millsap, R.E. (2004). *Beyond individual differences: Exploring school effects on SAT scores*. College Board Report No. 2004-3. New York: College Entrance Examination Board.

Everson, H.T., Millsap, R.E., & Diones, R. (1995, April). *Exploring group differences in SAT performance using structural equation modeling with latent means.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Gerber, S. (1996). Extracurricular activities and academic achievement. *Journal of Research and Development in Education*, *30(1)*, 42–50.

Holland, A., & Andre, T. (1987). Participation in extracurricular activities in secondary school: What is known, what needs to be known? *Review of Educational Research*, *57*, 437–466.

Holloway, J.H. (2000). Extracurricular activities: The path to academic success. *Educational Leadership*, December 1999/ January 2000. Association for Supervision and Curriculum Development. Arlington, VA.

Marsh, H.W. (1992). Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals? *Journal of Educational Psychology*, *84(4)*, 553–562.

Marsh, H.W. & Kleitman, S. (2002). Extracurricular school activities: The good, the bad, and the nonlinear. *Harvard Educational Review*, *72(4)*, 464–514.

McDonald, R.P., & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247–255.

Steigler, J.H., & Lind, A. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

U.S. Department of Education (2003). *Identifying and implementing educational practices supported by rigorous research: A user-friendly guide*. Washington, DC Institute for Education Sciences.