

Pitfalls of Using SAT Results to Compare Schools

Mark E. Fetler

Chancellor's Office, California Community Colleges

Public reports and comparisons of school SAT results are on the rise. This study uses a simple model of schooling to address two questions which are raised by school SAT comparisons. How are the demographic and academic characteristics of the student population in schools related to participation in the SAT? How are participation rates and the demographic characteristics of the sample of test-takers and of their schools related to SAT performance? School participation was associated with parent education, achievement, and English language fluency. The relationship between participation and parent education was curvilinear. School SAT scores were positively correlated with participation and parent education of the test-takers as well as with overall twelfth-grade school-average achievement and demographic characteristics. After adjustment for school demographic characteristics, participation rates were negatively associated with scores. School-average achievement was associated with SAT performance even after adjusting for demographic characteristics. These findings illustrate some of the complex relationships between the demographic characteristics of students and performance on the SAT which should be considered when making school comparisons.

MARK E. FETLER is the director of Planning and Performance Outcomes, Chancellor's Office, California Community Colleges, 1107 Ninth St., 4th Flr., Sacramento, CA 95814. He specializes in assessment, research, and evaluation.

Few educational statistics are more closely watched than the SAT (Scholastic Aptitude Test) results released annually by the College Board. Public reports and comparisons of national, state, and school scores are common even though the SAT was originally developed as an assessment for individual students (Angoff, 1986; Donlon, 1984). Various obstacles to the meaningful interpretation of aggregated scores include the selection bias which derives from the voluntary participation of students, the use of inappropriate statistical analyses, and the undue emphasis on the SAT as a measure of educational quality (Gohman, 1988; Haertel, 1986; Page & Feifs, 1985; Powell & Steelman, 1984; Taube & Linden, 1989; Wainer, 1986; Wainer, Holland, Swinton, & Wang, 1985). Two general findings illustrate the barriers to understanding direct state comparisons using SAT results. States with higher average income or less poverty tend to have higher average SAT scores. Moreover, higher participation rates are associated with lower average SAT scores, possibly because the greater levels of participation are associated with a lower overall aptitude of the sample of students taking the test. The continuing public commentary and dissemination of national, state, and school performance data, despite these known interpretive difficulties, attest to the need for more research and documentation of the issues associated with reports and comparisons of SAT results.

Holding schools accountable for SAT scores takes for granted a model of how those scores were produced. Wainer, Holland, Swinton, and Wang (1985) and Wainer (1986) developed a model for state results which can form a basis for thinking about schools. The elements distinguished in this model are student performance, educational programs, and those background factors associated with performance, for example, demographic characteristics. The purpose of the model is to investigate how various educational programs influence SAT performance. Direct comparisons of school or state averages are not useful because they are confounded with differences in demographic characteristics. If the appropriate statistical adjustments for the demographic characteristics can be made, then it might be possible to begin examining the influence of educational programs on performance.

This study seeks to widen the scope of research already done with states to the school level by addressing two main questions. How are demographic and academic characteristics of the student population in schools related to SAT participation? How are participation rates and the demographic characteristics of the sample of test-takers and of their schools related to SAT performance? Answers to these questions are discussed in connection with problems in using the SAT for making comparisons of school performance. The lack of information on educational programs limits the study in that the existence of such programs is probably correlated with the background factors. If the relevant information were available, it would probably attenuate the measured importance of the background factors.

A model of school SAT performance is adopted as a basis for examin-

ing both SAT participation and performance. Student participation is modeled in terms of twelfth-grade parent education, student English language fluency, and student achievement. Both parent education and student English language fluency are background variables which reflect demographic aspects of a school's population. A number of specific questions are posed in the context of this model. Is the relationship between parent education and participation in the examination linear, which is often assumed, or is it more complex? Is language fluency related to participation? Do limited English speaking students tend not to participate? If achievement is positively correlated with participation, to what extent is this relationship mediated by demographic factors?

SAT performance is modeled in terms of the parent education of the sample of test-takers, the participation rate, the percent of limited English speaking twelfth-graders, the parent education of all twelfth-graders, and overall twelfth-grade achievement. Participation and achievement may both be influenced by educational programs and demographics. The school achievement, English proficiency, and parent education indexes represent all twelfth-graders. The SAT average represents only a sample of students, mostly twelfth-graders. A number of specific questions can be addressed with this model. Are higher levels of participation associated with lower performance? Given the relatively select group of SAT participants, how strong is the relationship between parent education and performance? How strong is the relationship between achievement and SAT performance after adjusting for the other model components?

Method

Sample

Included in this study are most of the regular, comprehensive, public high schools in California. High school districts are governed by local, autonomous boards and are mostly comprised of schools with grades nine and above, but may infrequently include grades seven and above. Not included in the sample are private schools, continuation high schools for students at risk of dropping out, or schools maintained by county offices for special-needs students. States with low SAT participation rates typically have only their most talented students taking the SAT. In California, a relatively diverse subset of students takes the SAT with about one third of public school seniors participating.

Data Sources

Data used in this study were taken from the Admission Testing Program (ATP) of the College Board and the California Assessment Program (CAP) of California's Department of Education. CAP annually administers a standardized achievement test and student background survey to all twelfth-grade students attending public schools. Data from the December 1984 test-

ing were used in this study, with roughly 95 percent of high school seniors participating. Variables used in this study included student achievement, a parent education index, and percentage of limited English proficient students. Achievement is the average of the reading and mathematics percentages of correct scores. The parent education index is estimated from a student survey requesting information on the highest educational level attained by a parent or guardian. This index is similar to other commonly used measurements of socioeconomic status (e.g., Bonjean, Hill, & McLemore, 1967), and results obtained with it are consistent with those typically found elsewhere (Fetler & Carlson, 1985; White, 1982). The percentage of limited English proficient students is computed from data supplied by teachers who used state-mandated testing criteria to classify students as English only, fluent English, or limited English proficient.

The ATP includes the SAT and the Student Descriptive Questionnaire. The 769 public high schools providing data for this study represented 79,429 SAT takers from 235,338 seniors in 1984. ETS (Educational Testing Service) reported 83,422 California public high school students took the SAT in 1984. This discrepancy is most likely the result of inconsistent school codes used by the College Board and by the California Department of Education, or incorrect or missing information supplied by students. Information from the ATP included the number of students taking the SAT, the verbal and mathematics SAT scores, averaged to compute an overall score, and the parent education of the subset of participating students.

Analyses

The school is the unit of analysis in this study. School averages were weighted by the twelfth-grade enrollment. This procedure gives more emphasis to large schools with small standard errors of the mean. Preliminary analyses showed that the relationship between the twelfth-grade parent education index and student participation in the SAT was nonlinear and suggested the use of a quadratic term. In order to account for this nonlinearity without needlessly multiplying study variables, the twelfth-grade parent education index (PEI) was transformed to equal the squared deviation of the original index from its square, i.e., $(PEI - PEI^2)^2$. This transformation is justified in that parent education and its square are not really two different variables, and there is little reason to allot proportions of variance to each part separately. Using the transformed variable avoids collinearity problems. Table 1 displays descriptive statistics for the study variables. Regressions of the SAT participation rate and the average SAT score on school achievement and background factors are shown in Table 2.

Results

The variables in Table 1 to be examined with the participation rate were achievement, the percentage of limited English proficient students, and the

Table 1
**Correlations of Twelfth-Grade Achievement
 and SAT Results with Background Factors**

Variable	Twelfth-Grade			SAT		
	Achieve- ment score	Percent limited English	Parent education index	Percent partici- pation	Score	Parent education index
Mean	65.42	4.62	52.80	33.72	444.71	4.41
SD	5.86	6.81	43.93	15.08	40.37	0.77
Achieve- ment score		-.35	.77	.65	.85	.80
Percent limited English			-.31	-.17	-.35	-.45
Grade 12 parent education index				.81	.65	.82
Percent partici- pation					.44	.59
SAT score						.80

Note. The twelfth-grade parent education index (PEI) equals the squared deviation of the original index from its square, i.e., $(PEI - PEI^2)^2$. All correlations are significant ($p < .0001$).

transformed twelfth-grade parent education index. Although not unexpected, perhaps the main finding is the positive correlation of parent education with participation. A closer look at the relationship between the parent education index and the participation rate is provided by Figure 1. The scatter plot displays the joint distribution of the untransformed grade-twelve parent education index and the participation rate. The curve in the figure represents a fitted regression line. The regression model was developed using the transformed parent education index to predict the participation rate. An R-square value of .59 was obtained for this model, which is described by the following equation: $PREDICTED\ PARTICIPATION = 18.60 + 0.28(PEI)$. The parameter estimate for PEI is significant, ($p < .0001$), supporting the visual evidence of a curvilinear relationship. Of course, this curvilinear relationship is not necessarily quadratic. It appears that among relatively advantaged schools, greater participation is associated with higher

Table 2
**Regressions of SAT Participation Rates and
 Scores with Achievement and Background Factors**

Variable	Parameter Estimate	
	SAT participation	SAT score
Intercept	2.00 (0)	35.70 ^b (0)
Achievement score	0.25 ^a (0.10)	5.13 ^b (0.75)
Grade 12 parent education index	0.27 ^b (0.77)	- 1.56 ^b (- 0.14)
Percent limited English	.023 ^b (0.10)	0.19 ^b (0.03)
SAT parent education index	—	24.42 ^b (0.47)
Percent participation	—	- 0.59 ^b (- 0.22)
R-Square	.67 ^b	.81 ^b

Note. Standardized parameter estimates are shown in parentheses. ^aStatistically significant ($p < .01$); ^bStatistically significant ($p < .0001$)

parent education. Among less advantaged schools, the relationship between parent education and participation is relatively flat.

Those schools with either very high or very low participation rates can be seen as outliers in Figure 1. An inspection of the data revealed that schools with very high rates were mostly located in wealthier communities. Several other high participation schools have specialized programs and admit highly motivated and talented students by examination. One small school with a twelfth-grade enrollment of two students was found to have 100 percent participation. Schools with zero or very low participation rates were located in the state's poorest areas, or else were rural or quite small.

A regression model of the percentage of students taking the SAT, predicted by the transformed twelfth-grade parent education index, student achievement, and the percentage of limited English proficient students, is displayed on the left side of Table 2. An R-square value of .67 was obtained for this model. Of the three predictors, parent education was the most strongly associated with participation, followed by English proficiency and then by achievement. The importance of achievement as a predictor is smaller than might have been expected on the basis of the Pearson correlations, suggesting that the relationship between achievement and participation could be mediated by socioeconomic characteristics or school

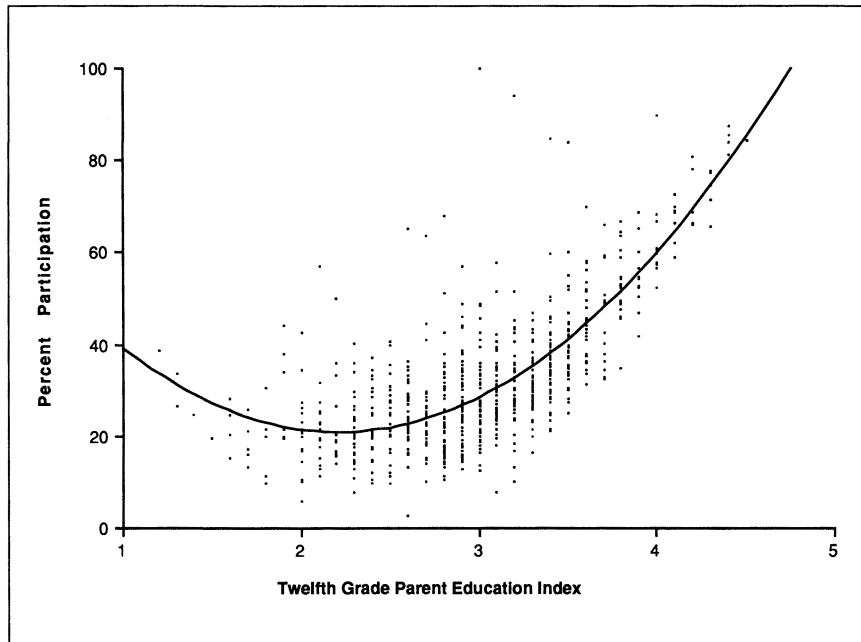


Figure 1. **School average SAT participation versus parent education.**

programs. The standardized parameter estimates for English proficiency and achievement are roughly the same magnitude. English proficiency, even after adjusting for the other predictors, remains significantly associated with participation.

Correlations of the SAT scores with the parent education index suggested that over half the variance in the SAT could be accounted for by the parent education of SAT participants. The selection of students for the examination might have attenuated this correlation, but apparently not greatly. The parent education of test-takers predicts SAT scores about as well as the parent education of twelfth-graders predicts achievement.

Contrary to expectation, participation was positively related to SAT performance. The extant analyses of state averages suggested that greater SAT participation is associated with lower scores. Here, the school level relationship between participation and performance was the opposite: greater participation was associated with higher scores. Given that the state and school level analyses are conducted at very different levels of aggregation, one should be wary of direct comparisons, i.e., the “ecological fallacy.” An explanation is hinted at by the positive correlation between the parent education index for twelfth-graders with both SAT participation and results. It is possible that socioeconomically advantaged schools tend to have both higher rates of SAT participation and higher average scores.

Results of the regression analysis of SAT scores are contained on the right side of Table 2. An R-square value of .81 was obtained for this model. Achievement was most strongly related to SAT performance, followed by parent education of SAT participants, the participation rate, the parent education of twelfth-graders, and percentage of limited English speakers.

Participation is negatively associated with performance, which is more in line with conventional wisdom, presumably the result of adjusting for achievement and parent education.

Discussion

A model of SAT performance was originally developed to examine state average scores. This study sought to extend that model in two ways. The first extension had to do with the school level focus of the two main study questions. How are the demographic and academic characteristics of the student population in schools related to SAT participation? How are participation rates and the demographic characteristics of the sample of test-takers and of their schools related to SAT performance? Although there is a body of research on the identification of effective schools (Austin & Holowenzak, 1985; Dyer, Linn, & Patton, 1969; Good & Weinstein, 1986; Marco, 1974), little detailed research has been done to examine the appropriateness of SAT scores as a measure of effectiveness. Of course, the intent of this study was to document the pitfalls arising from use of the SAT as a means of comparing schools and not to validate such use.

This study also went beyond previous work in its use of the participation rate as a criterion variable. The previous studies of state and national results focused primarily on test scores. More than just an indicator of potential bias, SAT participation can be regarded as one measure of access to the examination and is worthy of study in its own right. The study findings suggest that SAT participation is subject to the same kinds of demographic influences that are traditionally found with other measures of academic performance (White, 1982). What is perhaps more interesting is the curvilinearity of the relationship between participation and parent education. Below a certain "threshold" range of the parent education index, participation flattens out. The complexity of this relationship is one more reason for approaching school comparisons with caution.

Other factors which surfaced in the examination of the participation rates of particular outlying schools were enrollment, urban versus rural location, the existence of admission requirements in some cases, and extremes of socioeconomic conditions. It is tempting to dismiss or to ignore these schools as isolated instances. Yet if comparisons of schools are to be made, consideration should be given to the possibility of special underlying conditions, particularly where participation is extremely high or low.

Specific demographic conditions, e.g., the language proficiency of immigrant students, should also be taken into account when making com-

parisons. Although language proficiency may not be an important factor in all states, California is experiencing significant growth in the population of immigrant students (California Department of Education, 1988). The statistically significant relationship between English proficiency and participation, which persisted after adjustment for parent education and achievement, hints that there may be linguistic barriers to SAT participation for some immigrant students.

The twelfth-grade student achievement test score is a proxy for the academic preparation of the pool of students available to take the SAT. The results from the regression models imply that student achievement is another factor which should be taken into account. This relationship is difficult to interpret in that the various educational programs at schools which are intended to improve achievement might also influence college going aspirations and SAT participation. Pending the availability of relevant information, the model could be expanded to investigate the influence of educational programs on SAT participation or performance.

The conceptual model used to investigate the SAT scores included participation as a measure of selection and the parent education of SAT takers as well as school-average twelfth-grade achievement, parent education, and English proficiency. An example of the potential for confusion in comparing schools is illustrated by the positive Pearson correlation which was found between participation and performance. This result suggests a need to be sensitive to the level of aggregation of the analysis. Among states, greater participation apparently reflects the involvement of a more academically diverse group, not just an elite, which results in lower scores. By contrast, with smaller school level aggregates, more advantaged schools tend to have both higher rates of SAT participation and higher scores. Only after adjustment for parent education did it turn out that higher rates of participation were associated with lower scores.

The usual relationship between measures of socioeconomic status and academic performance was replicated for parent education and SAT scores. A more interesting finding is that high achievement is positively associated with SAT performance after adjusting for parent education, student participation, and school demographic characteristics. A more detailed study is needed to investigate the existence of particular factors or programs which are related to school achievement and SAT performance.

Conclusions

Taylor (1984) summarized appropriate and inappropriate interpretations of the SAT. The test was designed to measure developed verbal and mathematical skills of students, and can be used to assess their individual academic performance and needs. State and national summaries of SAT scores can help put individual performance into context. Score summaries, along with information provided by test-takers, can help colleges to plan curriculum

and student services. SAT scores are not an appropriate basis for comparing the educational quality of schools, districts, or states. The selection of students participating in the examination, the sensitivity of the SAT to social and economic factors outside of schools, and differences in student populations and educational programs in schools, districts, and states all combine to make such comparisons of educational quality inappropriate.

Test scores have become a widely accepted means for judging school effectiveness (Fetler & Carlson, 1985; Fetler, 1986; Fetler, 1991; Frechtling, 1989). However desirable it is to discourage school SAT comparisons, public attention to the SAT will probably continue to grow. If this is true, all audiences will be better served by a more thoughtful presentation of SAT results than is frequently the case. Defensive or strident denials of the appropriateness of school or state comparisons may draw public attention and generate more suspicion than understanding. More thoughtful and constructive approaches could involve a model including background characteristics of the relevant test-taking population and relevant educational programs along with the SAT results. This type of model might also diminish the temptation to rely on the SAT as a single measure of educational quality. The difficulty of the task should not discourage the attempt to improve public understanding. Coping effectively with the consequences of public access to SAT scores does not necessarily mean that one condones poor reporting practices.

Note

Views expressed in this study are not necessarily those of the Chancellor's Office of the California Community Colleges. I am grateful for the thoughtful and constructive remarks of several anonymous reviewers.

References

- Angoff, W. (1986). Some contributions of the College Board SAT to psychometric theory and practice. *Educational Measurement: Issues and Practice*, 5, 7–11.
- Austin, G., & Holowenzak, S. (1985). An examination of 10 years of research on exemplary schools. In G. Austin & H. Garber (Eds.), *Research on Exemplary Schools* (pp. 65–82). New York: Academic Press.
- Bonjean, C., Hill, R., & McLemore, S. (1967). *Sociological measurement: An inventory of scales and indices*. San Francisco: Chandler Publishing Company.
- California Department of Education. (1988). *Racial or ethnic distribution of staff and students in California public schools*. Sacramento: author.
- California Postsecondary Education Commission. (1986). *California college going rates*. Sacramento: author.
- Donlon T. (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.
- Dyer, H., Linn, R., & Patton, M. (1969). A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. *American Educational Research Journal*, 6, 591–605.

- Fetler, M., & Carlson, D. (1985). Identification of exemplary schools on a large scale. In G. Austin & H. Garber (Eds.), *Research on exemplary schools* (pp. 83–96). New York: Academic Press.
- Fetler, M. (1986). Accountability in California public schools. *Educational Evaluation and Policy Analysis*, 8, 31–44.
- Fetler, M. (1991). A method for the construction of differentiated school norms. *Applied Measurement in Education*, 4, 53–66.
- Frechtling, J. (1989). Administrative uses of school testing programs. In R. Linn (Ed.), *Educational Measurement* (pp. 475–484). New York: American Council of Education & Macmillan Publishing Company.
- Gohman, S. (1988). Comparing state SAT scores: Problems, biases, and corrections. *Journal of Educational Measurement*, 25, 137–148.
- Good, T., & Weinstein, R. (1986). Schools make a difference: Evidence criticisms and new directions. *American Psychologist*, 41, 1090–1097.
- Haertel, E. (1986). Measuring school performance to improve school practice. *Education and Urban Society*, 18, 312–325.
- Marco, G. (1974). A comparison of selected school effectiveness measures based on longitudinal data. *Journal of Educational Measurement*, 11, 225–234.
- Page, E., & Feifs, H. (1985). SAT scores and American states: Seeking for useful meaning. *Journal of Educational Measurement*, 22, 305–312.
- Powell, B., & Steelman, L. (1984). Variations in state SAT performance: Meaningful or misleading? *Harvard Educational Review*, 54, 389–412.
- Taylor, D. (1984, February 22). Comparing SAT scores won't reveal the state of American education. *Education Week*, p. 18.
- Taube, K., & Linden, K. (1989). State mean SAT score as a function of participation rate and other educational and demographic variables. *Applied Measurement in Education*, 2, 143–159.
- Wainer, H., Holland, P., Swinton, S., & Wang, M. (1985). On "State education statistics." *Journal of Educational Statistics*, 10, 293–325.
- Wainer, H. (1986). Five pitfalls encountered while trying to compare states on their SAT scores. *Journal of Educational Measurement*, 23, 69–81.
- White, K. (1982). The relation between socioeconomic status and achievement. *Psychological Bulletin*, 91, 461–481.