State-By-State SAT Comparisons: Pillars Of Salt Driving Educational Practices?

Bob Algozzine
Fred Spooner
Bob Audette

University of North Carolina at Charlotte

Each year, public and professional opinion is drawn to reports of state-by-state comparisons of the performance of high school students on the Scholastic Aptitude Test (SAT). States with high scores are viewed as "doing the right thing" and states at the other end of the distribution scramble to defend the quality of their educational system or to release plans to revamp, restore, or reform their practices. The question at the base of this research was: Are these reactions justified? Scores for different geographic areas were compared and variation in percent of students taking the test and teachers' salaries were used to predict overall performance scores. Strong relations were indicated between differences in scores reported for states and differences in percentages of students taking the test; teacher salary information was less important in these predictions. Reporting state-by-state performance scores appears to be a practice of limited professional or practical value.

Each year efforts to stay ahead in the educational foot race culminate with publication and analysis of state-by-state comparisons of Scholatic Aptitude Test (SAT) performance (Drescher & Eichel, 1991; Kelly, 1991; Morrell, 1989). Much like the arrival of the first report card of the year (SAT reports typically appear in September), proclamations from the College Entrance Exam Board (CEEB), which owns the test, and the Educational Testing Service, which administers it, sometimes bring misery and fear to even the most optimistic educators. Indeed, placement in the SAT hierarchy is viewed as evidence to support the need and progress of education reform. For example, when his state remained last in the national rankings, a spokesman for the South Carolina Business-Education Committee said: "It doesn't surprise me... To go the next step, we've got to get some new thing going" (Drescher & Eichel, 1991, p. 4A).

The SAT has gained value because individuals most likely to enter professional fields requiring strong verbal and quantitative skills take it and because the "SAT is the test most frequently used by scholars, politicians, and the lay public to assess school quality" (Powell & Steelman, 1984, p. 391). Explaining state and regional variation in SAT performance has been the topic of educational interest and research for some time.

Powell and Steelman (1984) used percentage of test-takers, sex, racial composition of the test-taking sample, median family income, years of academic coursework, public expenditures per student, and public/private status of the sample to critically evaluate state and regional variations in SAT performance. They noted that four fifths of the differences in SAT scores between states could be attributed to factors such as percentage of students taking the test, composition of the test-taking sample, and average family income. Demographic/compositional factors (sex composition, racial composition, median family income) and school/structural charateristics (years of academic coursework, public expenditures per student, educational status of the sample) accounted for significantly less variance than percentage of students taking the test. Average teacher salary is a variable noticeably absent in analyses of variation in SAT performance.

Although differences in teachers' salaries are often related to age and cost of living differences across states, salary differences are sometimes considered as indicators of

differences in support for public education—high salaries reflective of more support and better systems (Feiman-Nemser & Floden, 1986). And, while education officials and advisory commissions often recommend increased expenditures as a means of increasing scores on standardized tests (Powell & Steelman, 1984), relations between the extrinsic rewards of teaching (e.g., salaries) and student achievement have not been the topic of a large amount of research (Feiman-Nemser & Floden, 1986; Page & Feifs, 1985). In an effort to shed additional light on the nature and meaning of SAT differences across states testing different groups of high school students in a particular year, relations between percentage of students taking the test, average teacher salaries, and average SAT scores were evaluated in this research.

Method

Data compiled and reported by the College Entrance Exam Board for performance by high school students in 1991 on the SAT were available for analysis. The dataset was reprinted in the Charlotte Observer in an article headlined as "NC edges up to 48th; SC remains dead last" on August 27, 1991. States were ranked in the table, which also included previous year's scores and the percent of high school students taking the SAT. Data on average teachers' salaries were presented in an article entitled, "Teacher's pay increases 5.4%" which appeared in the May 8, 1991 issue of USA Today (Ordovensky, 1991). The 90-91 salaries, ranked from top to bottom by state, and percent increases were presented; the source of the information was listed as the National Education Association.

Data Analysis

The information was entered into a single dataset and subjected to a series of statistical analyses. First, distributions of available scores were described and compared by using states as the units of analysis. Simple descriptive statistics were used to illustrate levels of these demographic variables. Second, scores were compared for different geographic areas. One-way analyses of variance with appropriate followup tests (i.e., TUKEY LSD) were used in these comparisons. Finally, correlations between teachers' salaries, the percentage of students taking the SAT in a year, and the combined scores reported for that year were tested. Simple and multiple regression analyses were completed. The 0.05 level of significance was used in judging significance of overall and followup statistical tests.

Results

Approximately 35% of the projected U. S. high school graduates took the SAT in 1991. The minimum taking it in any state was 4% and the maximum was 81%; this distribution was slightly positively skewed. The average score obtained on the SAT was 946 (SD = 66);. Scores on the Verbal portion (M = 449, SD = 32) were somewhat lower than scores on the Quantitative portion (M = 497, SD = 35) of the test. The average teacher salary was $31,115 ($SD$ = $5,261), with the top salary ($43,861) in Alaska and the bottom in South Dakota ($22,363).

A set of comparison groups was formed by combining states from four geographic regions (i.e., Northeast, South, Midwest, West). Significant overall F ratios were indicated for SAT Performance, percentage of students taking the test,

Table 1
Relations Among SAT Performance And Teacher Salary[a]

	Verbal 1991	Quant 1991	Total 1991	Percent 1991	Average Salary
Quant	.97 (.88)				
Total	.99 (.97)	.99 (.97)			
Percent	-.87	-.86	-.87		
Salary	-.51 (.09)	-.45 (.22)	-.48 (.16)	.63	
Increase	-.06 (.11)	-.13 (-.05)	-.10 (.03)	.13	.04 (-.06)

[a]First order partial correlations controlling for percent taking the SAT are presented in parentheses. Zero order correlations between salary increase and other variables are non-sigificant; all other zero order relations are significant at 0.05 level. First order partial correlations between salary and salary increase and other variables are non-sigificant; all other first order relations are significant at 0.05 level

and teachers' salary information ($F(3,46) = 9.58, 17.78, 4.15$ respectively). SAT scores from the Midwest region were significantly higher than scores from all other clusters of states. No differences were indicated between scores for students in the South and West or South and Northeast regions. Percentages of students taking the SAT were also different. The percentage of students taking the SAT from the Midwest was significantly lower, and the percent taking the test from the Northeast was significantly higher than the other regions. No differences were indicated between the percents taking the SAT in the South and West. As expected, teachers' salaries were significantly higher in the Northeast. The largest significant difference (121 points) was evident between SAT combined scores for Northeastern and Midwestern states. These states also had the largest difference (57 points) in percentages of students taking the test. In general, large differences between states in combined scores were associated with large differences in percent of high school seniors taking the test in these states; and insignificant differences in combined scores were associated with insignificant differences in numbers of students taking the SAT in high school in these states.

Simple and partial correlations among SAT performance scores, percentage of students taking the SAT, and average teachers' salary information are presented in Table 1. High inverse zero order correlations were indicated between percentage of students taking the SAT and actual test performance scores. Moderate inverse zero order correlations were indicated between teachers' salary information and SAT performance scores. A moderately high zero relation was indicated between salary information and the percentage of students taking the SAT. Relations between salary information and SAT performance were nonsignificant when effects of percent taking the test were partialed out.

A stepwise regression analysis was completed to evaluate the relation between the percentage of students taking the SAT and teachers' salary information and the total combined performance recorded in 1991. The percentage of students taking the test in 1991 accounted for the most variance in the total SAT scores for that year. The obtained correlation coefficient was 0.87. The percentage of students taking the SAT accounted for 76% of variance in total scores reported for 1991. Salary information did not account for significant additional variance. The simple regression equation based on these data follows:

$$\text{SAT Combined}_{(1991)} = -2.21 * \text{percent taking}_{(1991)} + 1023.69$$

As percentage of students taking the SAT in 1991 dropped 1 point, the combined score increased by more than 2 points; 10-point differences in percent of students taking the SAT were associated with 20-point differences in overall SAT scores. Visual inspection of a scattergram for the 1991 relationship and an analysis of residuals suggested that a nonlinear relationship existed. A multiple regression analysis using the original and a transformation (score squared) of the percentage of students taking the test was completed. The obtained multiple regression coefficient was 0.93, and the regression equation accounted for 86% of the variance in total scores reported for that year.

Large differences were evident in SAT performance scores compiled across most geographic areas. Similarly large differences were evident in the percentage of students taking the SAT across these same regions. Small percentages of students taking the SAT were associated with higher performance scores, suggesting that better students may be included when samples of test-takers are smaller. Sizable relationships were indicated between percentages of students taking the SAT and the overall performance reported for each of the states. Regression analyses revealed that significant proportions of variance were accounted for by linear and nonlinear relations. Average teacher salary and salary increase indices were less important in all relational analyses.

Discussion

Professionals worry about SAT scores and use them as indices for describing all kinds of social ills and educational problems (Gould, 1981; Owen, 1985; Powell & Steelman, 1984). Similarly, officials in states with high overall scores are quick to praise their educational efforts and separate themselves from states with lower scores. According to Powell & Steelman, "One political spokesperson has suggested naming Iowa 'the state of minds,' in part because of its comparatively high performance on standardized tests" (1984, p. 390). Professionals in low-ranking states defend their educational system and use measures other than the SAT as more valid indications of the quality of what goes on in their schools. For example, opponents of Jesse Helms countered his attack on the educational system based on North Carolina's low SAT ranking by using improved performance on the CAT as an indicator of educational progress (Page & Feifs, 1985).

Indictments against using state-to-state variations in standardized test performance as indicators of quality and equality in education are not new. Womer (1983) noted hazards for those in the testing community when he pointed out that ". . . the SAT is not an agency of educational accountability; and the SAT is not designed to sample curricular attainments of high school seniors nationwide; and the SAT is not administered to random samples of high school seniors" (p. 4, emphasis in original). In 1984, Powell and Steelman found that "most state variation [in overall SAT performance scores] is a function of one factor, the percentage of eligible students taking the exam" (p. 408).

To "account for some of the variation in state means for the SATs," Page and Feifs (1985) used the percentage of minority students, pupil-teacher ratio, teacher salaries, state employment rates, and several other predictor variables in their regression analyses. They found that "the most power-

ful influence on the state means, of course, [was] the percentage of high school graduates who [took] the SAT exams" (p. 310). Recently, Wainer (1989) discussed problems related to drawing conclusions from samples in which selection criteria are not known (e.g., students who choose to take the SATs). Using a re-analysis of data provided by earlier investigators (i.e., Page & Feifs, 1985; Steelman & Powell, 1985), he found a moderate correlation (0.5) between state rankings but considerable variance in the relative standings of some individual states (e.g., New Mexico was first and 30th in separate analyses). A distinguished group of colleagues provided responses and generally supported Wainer's position (cf. Allen & Holland, 1989; Birnbaum & Mellers, 1989; Heckman, 1989; Rosenbaum, 1989; Rubin, 1989; Speed, 1989; Wachter, 1989).

Findings from the current analyses support previous work: SAT performance is highly related to sample characteristics and "the sun comes up in the east." Yet, the College Entrance Exam Board continues to report these scores; and professionals and lay people continue to think there is importance in them; and, in the absence of controlled studies of carefully selected samples of students, widespread dissemination of these data begs continued analysis. Contemporary educational practice justifies continued display of these relations.

Clearly, praising or worrying about the status of a state's educational system does not appear justified based on analyses of rankings of overall performance on the SAT, unless the influence of sample characteristics which implies more than just percentages is taken into account. This conclusion is not new, albeit supported here by new data. Yet the College Entrance Exam Board continues to publish unadjusted rankings, and educational professionals continue to be seduced to enjoy or dread them. One would hope this prestigious organization would engage in responsible reporting but, in the absence of evidence of such practices in the past, clearly "caveat emptor" has abundant and significant relevance to those who would use the SAT performance averages as measures of equity, quality, or the need for educational change.

References

Allen, N. L., & Holland, P. W. (1989). Exposing our ignorance: The only "solution" to selection bias. *Journal of Educational Statistics, 14*, 141-145.

Birnbaum, M. H., & Mellers, B. A. (1989). Mediated models for the analysis of confounded variables and self-selected samples. *Journal of Educational Statistics, 14*, 146-158.

Drescher, J., & Eichel, H. (1991, August 27). NC edges up to 48th; SC remains dead last. Charlotte Observer.

Feiman-Nemser, S., & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 505-526). New York, NY: Macmillan.

Gould, S. J. (1981). *The mismeasure of man.* New York: W. W. Norton.

Heckman, J. J. (1989). Causal inference and nonrandom samples. *Journal of Educational Statistics, 14*, 159-168.

Kelly, D. (1991, September 24). SAT Analysis: States score lower as more take test. *USA Today.*

Morell, R. (1989, September 12). NC scores on SATs hit bottom rung. *Charlotte Observer*, September 12.

Ordovensky, P. (1991, May 8). Teachers pay increases 5.4%. *USA Today.*

Owen, D. (1985). *None of the above: Behind the myth of scholastic aptitude*. Boston: Houghton Mifflin.

Page, E. B., & Feifs, H. (1985). SAT scores and American states: Seeking for useful meaning. *Journal of Educational Measurement, 22*, 305-312.

Powell, B., & Steelman, L. C. (1984). Variations in state SAT performance: Meaningful or misleading? *Harvard Educational Review, 54*, 389-412.

Rosenbaum, P. R. (1989). Safety in caution. *Journal of Educational Statistics, 14*, 169-173.

Rubin, D. B. (1989). Bugs, lacunae, and the Minnesota/DC effect: A discussion of Wainer's "Eelworms, bullet holes, and Geraldine Ferraro." *Journal of Educational Statistics, 14*, 175-178.

Speed, T. P. (1989). Contribution to discussion of "Eelworms, bullet holes, and Geraldine Ferraro." *Journal of Educational Statistics, 14*, 179-181.

Steelman, L. C., & Powell, B. (1985). Appraising the implications of the SAT for educational policy. *Phi Delta Kappan, 66*, 603-607.

Wachter, K. W. (1989). Statistical adjustment: Comments on H. Wainer's "Eelworms, bullet holes, and Geraldine Ferraro." *Journal of Educational Statistics, 14*, 183-186.

Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics, 14*, 121-140.

Womer, F. B. (1983). Congratulations, Mississippi! Shame on you, Rhode Island! (Editorial). *Educational Measurement*, 2, 4.

Bob Algozzine is Professor of Teacing Specialties in the College of Education and Allied Professions, University of North Carolina, Charlotte, NC.

Fred Spooner is Assistant Professor of Teaching Specialties in the College of Educatino and Allied Professions, University of North Carolina, Charlotte, NC.

Bob Audette is Assistant Professor of Curriculum and Instruction in the College of Education and Allied Professions, University of North Carolina, Charlotte, NC.