Briggs, D. C. (2002). Comment: Jack Kaplan's new study of SAT coaching. *Chance*, *15*(1), 7-8.

Comment: Jack Kaplan's "A New Study of SAT Coaching""

By Derek C. Briggs

The quality of SAT coaching matters! If there is a single message that Jack Kaplan would want a reader to take away from his article, I would guess that this might be it. Coaching that is above average in quality should be expected to produce associated score gains that are also above average. This is an intuitively plausible argument, and the results from Kaplan's study are suggestive along these lines. I would submit however, that beyond intuitive appeal, there is a dearth of evidence to make this argument empirically compelling.

Kaplan uses the word compelling to describe the evidence presented in six studies that supposedly indicate that some "commercial coaching programs raise average scores by more than the College Board is willing to admit." Kaplan has determined what the College Board is willing to admit by examining their web site, where they indicate that depending on program length, the effect of coaching on the SAT ranges from 15 to 30 points on the math section (SAT-M), and 10 to 20 points on the verbal section. Table 1 summarizes the largest statistically significant coaching effects estimated in each of the six studies Kaplan has cited.

Study	Coaching Effect	
	SAT-M	SAT-V
Pallone (1961)[a]		81
Marron (1965)[a]	59	35
Pike (1978)	33	
Zuman (1988)	58	52
Smyth (1990)[b]	32	6
Powers & Rock (1999)[c]	31	14

a. These are "effects" from Pallone and Marron's uncontrolled studies as estimated by Messick & Jungeblut (1981).
b. Smyth's estimates do not control for group differences. When this was done is a subsequent study, the SAT-M effect dropped to 18 points
c. This is taken from Powers & Rock's Heckman Model for "company B"

The studies by Pike, Smyth and Powers & Rock all report estimated effects within a few points of the range claimed by the College Board. The Pallone and Marron studies are well outside the range but were based solely upon samples of males enrolled in private college preparatory schools, not commercial coaching programs. The most compelling of the lot by far is the study by Zuman, but its generalizability is limited because the findings are based upon a very small sample. It seems skeptical readers should remain skeptical!

Kaplan gives a balanced summary of some of the notable academic reviews written on the topic of SAT coaching. I think it is important to consider the historical sequence of these reviews. The Messick & Jungeblut review was written at least in part as a direct response to a rather incendiary review by Slack & Porter. Slack & Porter claimed, among other things, that ETS was purposively cooking the books against coaching by ignoring the Pallone, Marron and Pike studies when estimating average effects. This provoked a slew of research and reviews on coaching in the early 1980s. The meta-analysis of DerSimonian and Laird should be placed within this context. As

independent researchers, not only did they find that coaching had a small effect on SAT scores, as Kaplan reports, but they found a systematic relationship between the size of the estimated coaching effect and the design of the study. Specifically, effects estimated in uncontrolled studies were found to be *four to five times larger* than those estimated in controlled or matched/randomized studies. So we should be cautious in interpreting Kaplan's finding of a coaching effect for SAT-M that is 2 to 3 times bigger than the largest effect conceded by the College Board.

Self-Selection and Small Samples

Kaplan calculates average score gains on the SAT-M of 60 and 87 points respectively for two different cohorts of nine students (excluding three students who had been previously coached in the second cohort). How much of this gain can be attributed to Kaplan's coaching? To estimate the coaching effect, Kaplan needs a comparable control group of high school students who take the test in the spring of their junior year and then again in the fall of their senior year . Not having such a group in his sample, he looks elsewhere, first at national score gains calculated from a 1997-98 College Board study, then at score gains calculated for 50 students from a local public high school in 2000. In each case Kaplan finds the average score gain on the SAT-M to be 13 points. If 13 points is used as a control baseline, the estimated effect of Kaplan's coaching for each cohort of students would be 47 and 74 points .

Because of the self-selected nature of Kaplan's students, the suitability of this baseline is highly questionable. We have little reason to expect Kaplan's cohorts to be comparable to the 1997-98 national population. To begin with , the average 13 point gain calculated by the College Board comes from all students taking the test twice, including those with prior SAT-M scores as low as 280. In Kaplan's cohorts the prior SAT-M scores are 460 and 480 respectively. Beyond this, students in the national population are likely to vary substantially along any number of variables correlated with SAT-M performance. The 2000 high school sample Kaplan uses is more comparable with respect to prior SAT-M scores and socioeconomic background, but other potentially confounding variables—for which we have no information—include race/ethnicity, academic achievement, personal motivation and other test preparation activities.

How much of an impact might self-selection bias have on estimated coaching effects? This is pure speculation. In the analysis I did using data from NELS:88, controlling just for demographic and academic achievement variables reduced the coaching effect for SAT-M by 26 percent. Extrapolating this reduction to Kaplan's study would reduce his coaching effect per cohort to 35 and 55 points.

It is difficult to attach much confidence to the precision of Kaplan's estimated coaching effects because they are based on extremely small samples of students. This makes it more likely that the effect is either over or underestimated due to chance variability. For example, if the underlying population standard deviation for Kaplan's coached students was about 60 points, then a single standard error around his coaching

effect estimate would be 20 points. Again, assuming for the moment that self-selection bias reduces Kaplan's effect estimates by about 26 percent, the 95% confidence interval for his coaching effect would be between –5 and 75 points for the first cohort, and between 15 and 95 points for the second cohort.

Coaching Quality

A strength of Kaplan's study is that it takes us a step further inside the coaching definition. Kaplan describes the components of his program: instruction, practice on test items, working in pairs, individual mentoring and homework. What is still missing is a theory that ties together his coaching to score gains. What is it about Kaplan's instruction that should be expected to increase student scores? Does Kaplan teach test-taking strategies that make it easier to solve certain SAT-M item formats? Or is it just a basic review of principles from algebra and geometry that suffices?

Kaplan notes that his 2nd cohort of students show steady progress in their performance on five practice tests. He writes "if a student's score increased steadily as the course progressed, one can be fairly confident that the course itself was the determining factor." I am less confident. One could just as easily conclude that a student's score is increasing steadily due to systematic practice in taking the test. Is it the coaching plus the practice or is it just the practice? Without an adequate control group the effect cannot be disentangled.

The quality of SAT coaching *should* matter. But just what constitutes high quality coaching is an open question. There is limited empirical evidence that certain coaching programs are more effective than others. The evidence, however, is weak and needs to be substantiated more rigorously in controlled or even randomized study designs. If and when this happens it would be justifiable to expect the College Board to widen its range of estimates for commercial coaching's effectiveness. Until then the burden of proof is where it belongs: on those selling the product.

Further Reading

DerSimonian, R. and N. M. Laird (1983). "Evaluating the effect of coaching on SAT scores: A meta-analysis." Harvard Educational Review **53**: 1-15.

Evans, F. and L. Pike (1973). "The effects of instruction for three mathematics item formats." Journal of Educational Measurement **10**(4): 257-272.

Marron, J. E. (1965). Preparatory school test preparation: Special test preparation, its effect on College Board scores and the relationship of affected scores to subsequent college performance. West Point, N.Y., Research Division, Office of the Director of Admissions and Registrar, United States Military Academy.

Messick, S. and A. Jungeblut (1981). "Time and method in coaching for the SAT." Psychological Bulletin **89**: 191-216.

Pallone, N. J. (1961). "Effects of short- and long-term developmental reading courses upon the S.A.T. verbal scores." Personnel and Guidance Journal **39**(654-657).

Pike, L. W. (1978). Short-term instruction, testwiseness, and the Scholastic Aptitude Test: A literature review with research recommendations. Princeton, NJ, Educational Testing Service.

Powers, D. E. (1986). "Relations of test item characteristics to test preparation/test practice effects: A quantitative summary." Psychological Bulletin **100**(1): 67-77.

Powers, D. E. and D. A. Rock (1999). "Effects of Coaching on SAT I: Reasoning Test Scores." Journal of Educational Measurement **36**(2): 93-118.

Slack, W. V. and D. Porter (1980). “The Scholastic Aptitude Test: A critical appraisal.” Harvard Education Review **50**: 154-175.

Smyth, F. L. (1990). “SAT Coaching: What really happens and how we are led to expect more.” The Journal of College Admissions **129**(Fall): 7-17.

Zuman, J. P. (1988). “The effectiveness of special preparation for the SAT: An evaluation of a commercial coaching school.” Dissertation Abstracts International **48**(University Microfilms No. DA 8722714): 1749A-1750A.