DOCUMENT RESUME

ED 385 593 TM 024 041

AUTHOR Powers, Donald E.
TITLE Coaching for the SAT: A Summary of the Summaries and an Update. [Reprint.]
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-93-32
PUB DATE 93
NOTE 9p.
PUB TYPE Information Analyses (070) -- Journal Articles (080)
JOURNAL CIT Educational Measurement: Issues and Practice; p24-30,39 Sum 1993

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Cost Effectiveness; Higher Education; High Schools; *Mathematics Tests; Meta Analysis; *Test Coaching; *Time Factors (Learning)
IDENTIFIERS *Scholastic Aptitude Test

ABSTRACT

Several available summaries of research on coaching for the Scholastic Aptitude Test (SAT) are summarized and their principal findings discussed. Some additional studies, that have been completed since these summaries were reported, are considered and linked to the summaries. The four major meta-analyses considered are those of: (1) Messick and Jungeblut, 1981; (2) DerSimonian and Laird, 1983; (3) Kulik, Bangert-Drowns, and Kulik, 1984; and (4) Becker, 1990. Taken together, these studies indicate that the effects of coaching, special test preparation, are somewhat greater for the more curriculum-related mathematics section of the SAT than the verbal section. Longer coaching programs tend to yield somewhat greater effects, but simply doubling the effort does not double the effect. It is also apparent that the estimation of coaching effects depends on the degree to which spurious effects are controlled (e.g., regression, self-selection, noncomparable scores, differential motivation). In general, recent studies are consistent with the meta-analytic summaries. Those who seek coaching for the SAT should consider not only expected benefits, but also the cost in terms of time and money. Two tables summarize study findings. (Contains 30 references.) (SLD)

ETS RR-93-32

Coaching for the SAT: A Summary of the Summaries and an Update

Donald E. Powers
Educational Testing Service

Why is it important to investigate the effectiveness of coaching for a test such as the SAT? How can you evaluate the effectiveness of coaching? What are some common misconceptions? What do we know and what is still unclear about the effects of coaching? What can we tell students and their families about the results of coaching studies?

Performance on tests such as the Scholastic Aptitude Test (SAT) and the American College Testing (ACT) Program Assessments assumes at least some degree of importance to the several million college-bound students who take these tests each year. It is perhaps not surprising that because of the high-stakes nature of these examinations, parents, schools, and students are interested in maximizing performance on these assessments. In response to this attention, numerous secondary schools, commercial firms, and private entrepreneurs have developed a wide variety of special preparations for these tests, primarily for the SAT. Nowadays, there is what can be best described as a bewildering array of test preparation resources to help students prepare for the SAT (e.g. Powers, 1988). Test preparation books, software, and coaching courses of various sorts abound.

This article has two major aims. The first is to describe briefly several available summaries of coaching for the SAT and to discuss the principal findings of these summaries. The second is to consider several additional studies that have been completed since these summaries were reported and to relate the new findings to the summaries. At issue is whether or not special preparation (in particular, that which can be provided over a relatively short term) can, beyond the effects resulting from regular schooling, have a significant impact on test scores. The topic is important for several reasons. First, if extra preparation is effective but not reasonably available to all test takers, then some test takers may have an unfair advantage over others. Second, if short-term preparation that is geared mainly toward test-taking tricks is effective, then the interpretation of test scores as indicators of general academic ability (instead of simply the ability to take tests) is called into question. Third, because test preparation can be both costly and time-consuming, it may detract from students' participation in other worthwhile academic activities.

Misconceptions

The rate at which new test preparation resources are developed and marketed appears far greater than the rate at which adequate information is generated about the effectiveness of these offerings. Indeed, any comprehensive evaluation of extant test preparation resources seems unlikely. If the claims of the coaching enterprises have any validity, the most effective of the lot are the courses given by commercial coaching companies, which often promise, or even "guarantee," score improvements of 100 points or more.[1] These claims, however, appear to be based at least partly on false premises, partial information, and selective reporting. More will be said about these shortcomings later.

Consumers apparently are more easily convinced about the effectiveness of special test preparation than are others who research the topic. The most common misconception is that simple test-score gains from one occasion to another are an adequate reflection of the effects of coaching. Threats to the validity of this interpretation are readily apparent to experiment designers and measurement specialists. Because of practice with taking tests, measurement error, and real growth in abilities, an individual's test scores will vary from one test administration to another, regardless of any intervening test preparation. Quite predictably, some examinees will register large score increases upon retesting, and others will exhibit large decreases. Oftentimes, however, the only evidence needed to sway test takers is a large score increase by a single fellow student—a simple proof of concept. The problem is exacerbated when these

Donald E. Powers is a Senior Research Scientist at the Educational Testing Service, Mail Stop 17-R, Princeton, NJ 08541. His specialization is educational measurement.

mislabeled "effects" are reported selectively by coaching enterprises or by their clients. Coaching schools are more likely to publicize score increases than decreases, and big score gainers are more inclined to proclaim their success than score losers or "no changers" are to declare their lack of it.

Components of Test Score Gains

Even if they are fully reported, however, simple test score gains do not constitute a valid measure of effectiveness. As stated above, gains may reflect any and all of the following: test practice (i.e., simply having taken the test before), growth in the abilities measured by the test, and measurement error. Although the effect of test practice, unconfounded with other factors, is difficult to assess, it appears that simply repeating the SAT may improve test scores, perhaps by about 15 points on the 200–800 point verbal scale of the test and about 12 points on the math 200–800 point portion. These estimates, which are based on all students who took the SAT as juniors in the spring of 1990 and again as seniors in the fall of the same year (College Board, 1991), undoubtedly reflect some growth also.

With respect to growth, the rate at which the SAT verbal and quantitative reasoning skills develop has not been studied systematically, and it is likely that students develop these abilities at different rates. Some information is available, however. For example, in a longitudinal study of young, highly able students, Burton (Wilder, Casserly, & Burton, 1988) demonstrated a yearly average improvement of about 50 points on the verbal and mathematical portions of the SAT for students tested repeatedly between the ages of 13 and 17 years. It seems likely that these relatively consistent improvements resulted mainly from growth in abilities, because test practice effects should decrease over time. Growth may be less dramatic for the typical SAT taker than for these extremely able students. It is, however, a significant component of SAT score gains for all test takers and one that must be accounted for when estimating the effects of special test preparation.

Although growth and practice do increase test scores, measurement error, on the other hand, may contribute to either increases or decreases upon retesting. Typically, about 1 in 25 SAT takers will gain 100 or more total points, and about 1 in 110 will lose 100 or more points on retesting. Predictably, these changes will depend on students' initial scores, with low-scoring examinees more likely to register the largest gains and high-scoring test takers the greatest losses. For instance, junior-year test takers who score about 500 on SAT-V will, as seniors, average about 507. Those juniors who score about 500 on SAT-M will also average about 507 on this scale when they retest as seniors. In contrast, juniors who score at the 300 level upon retesting the next year will average about 331 on SAT-V and 342 on SAT-M (College Board, 1991). Johnson, Asbury, Wallace, Robinson, and Vaughn (1985) provided a good discussion of the effects of measurement error (specifically, regression to the mean) in the context of an SAT coaching study sponsored by the NAACP. The gains made by coached students varied markedly according to their initial test scores, respectively. Students whose initial scores were below 300 gained 41 verbal score points and 75 math points. For those scoring between 300 and 400 initially, gains were 30 and 19 points on the verbal and math scales, respectively. Test takers starting above 400 gained 23 points on the verbal scale but lost 5 points on the math. The authors properly noted the role of this regression effect in their estimates of the effects of the coaching program.[2]

The Evidence

Although test takers and their parents may be inclined to rely on word-of-mouth reports from previously coached students, good science demands more than anecdotal evidence—at a minimum, some comparison of coached examinees with uncoached ones,[3] many of whom will, for the reasons noted earlier, also register test score gains despite the lack of any special test preparation. Over the years, many individual studies of the effects of special preparation or coaching for the SAT have been conducted. These studies have differed not only in their methods but also in their results. The availability of these apparently conflicting results has made it possible to cite individual studies in support of claims that coaching for the SAT can be very effective or, on the other hand, that it does not work at all.

Meta-Analysis as a Tool

Fortunately, during the early 1980s a powerful analytical technique called "meta-analysis" was developed (Glass, McGaw, Smith, 1981; Hedges & Olkin, 1985; Hunter, Schmidt, & Jackson, 1982; Light & Pillemer, 1984; Rosenthal, 1984). This procedure has been heralded as one of the most significant developments in social science research methodology in recent years. Its strength is that it enables the integration of individual, possibly conflicting, studies conducted by different researchers under varying conditions. When considered separately, such studies often are based on samples that are too small, too limited in scope, or flawed in ways that preclude unequivocal conclusions. Collectively, however, these individual studies can be informative.

Recently, meta-analysis has been applied to studying the effects of special preparation for standardized tests, including the SAT. In fact, several summaries have now been reported—by ETS researchers Messick and Jungeblut (1981) and by university-affiliated investigators at Harvard (DerSimonian & Laird, 1983), Michigan (Kulik, Bangert-Drowns, & Kulik, 1984), and Michigan State (Becker, 1990). Each of these efforts is reviewed in subsequent paragraphs.

A Definition of Coaching

Before undertaking this review, however, it will be useful to consider more precisely what is meant by special preparation or coaching, for as Messick (1982) has suggested, the controversy over coaching has been stoked at least partly by different definitions of these terms. No resolution of these differences will be attempted here. It may, however, be

useful to mention some of the salient dimensions on which special preparation may vary.

Anastasi (1981), Bond (1989), Cole (1982), and Messick (1982) have each discussed various meanings and implications of test preparation and coaching. Special test preparation can vary according to objectives, duration, and methods. It may be designed to affect scores indirectly by increasing confidence or decreasing anxiety, or to raise scores more directly by teaching specific skills, strategies, or even "tricks." It can entail short-term cramming or long-term instruction. It can involve orientation to general test taking, familiarization for a particular test, review of relevant subject matter, drill-and-practice on sample test questions, or development of academic skills and competencies. When evaluating the effects of coaching it is necessary, therefore, to describe in some detail the germane characteristics of the preparation that is being evaluated. Unfortunately, however, evaluations do not always describe program characteristics well.

Findings From Four Meta-Analyses

What then is the answer to the question, "What is the effect of coaching for the SAT?" The meta-analytic summaries are useful in considering this query. For the "typical" program, the effect is about 15–25 points each on the verbal and on the mathematical portions of the SAT. A more precise answer, however, is a more qualified one, and the various meta-analyses now make it possible to explain some of the factors on which the results of individual studies seem to depend.

Objectives. First, however, a word is in order about the nature and intent of the various meta-analyses. In the earliest available quantitative summary, Messick and Jungeblut (1981) reviewed all of the available studies of coaching for the SAT regardless of the way in which the term "coaching" was defined. Focusing on school-based and proprietary programs, these researchers asked, "How much student time devoted to what kinds of coaching experiences yield what level of score improvements in comparison with the level of experiential growth that would have occurred anyway without these coaching experiences?" (p. 192). With regard to the methods used by individual investigators, the authors noted that all of the studies they reviewed were "methodologically flawed in various and divergent ways."

Next, DerSimonian and Laird (1983) analyzed all of the studies considered by Messick and Jungeblut (1981) as well as those reviewed in a narrative summary by Slack and Porter (1980). One of their intentions was to determine the extent to which the individual study estimates included in earlier summaries represent real variation among situations.

Kulik, Bangert-Drowns, and Kulik (1984) considered 38 studies of the SAT or other aptitude tests. These authors included only studies that involved "a true test-coaching program, not a program of practice or tutoring" (p. 180). By this, the reviewers meant that students were explicitly instructed in test-taking strategies, not merely allowed to practice on tests, and thereby infer effective strategies on their own. Studies that focused mainly on improving specific academic skills (other than verbal and quantitative reasoning) were not reviewed. In their analyses, these authors considered the features of the coaching programs, the methodological characteristics of the studies, and the attributes of the students involved.

In the most recent and most comprehensive review, Becker (1990) analyzed a total of 48 studies either taken from earlier meta-analyses or completed after these summaries were reported. Becker used an alternative measure of the effect of coaching that allowed the inclusion of all studies employing pretest-posttest comparisons, regardless of whether the studies incorporated a comparison group. Becker considered a number of factors simultaneously and asked about the relative contribution to coaching effect estimates of student characteristics, coaching interventions, and study design. She also asked whether or not coaching effects were different for the verbal and math portions of the SAT.

Principal Findings. Briefly then, what are the major revelations of these summaries? First, the effects of coaching are somewhat greater for the more curriculum-related mathematics section of the SAT than for the verbal section (Becker, 1990; Messick & Jungeblut, 1981). Also, as one might expect, longer coaching programs yield somewhat greater effects than do shorter ones. However, simply doubling the effort, for example, does not double the effect. Diminishing returns set in rather quickly, and the time needed to achieve average score increases that are much larger than the relatively small increases observed in typical programs rapidly approaches that of full-time schooling (Messick & Jungeblut, 1981). Becker (1990) also documented the relationship between duration of coaching and effects on SAT scores, noting a weaker association after controlling for differences in the kind of coaching and the study design.

Another important conclusion from these summaries is that the estimation of coaching effects depends heavily on the degree to which spurious effects are controlled (e.g., regression, self-selection, noncomparable scores, differential motivation). Studies that merely compare test score gains of coached students with national norms yield "coaching effects" that are about 4 to 5 times greater (and much less consistent) than effects estimated from studies that employ more scientifically rigorous designs (DerSimonian & Laird, 1983). Thus, if greater confidence is placed on the more rigorous studies, then the typical effect is less than 15–25 points on each (verbal and mathematical) portion of the SAT.

Using somewhat different analytical procedures than those used by DerSimonian and Laird (1983), Becker (1990) noted severe confounding between the characteristics of coaching studies, thus thwarting her attempt to fit a model that would explain variation in study results across a wide array of investigations. However, a simple model did explain differences among published studies that employed comparison groups. Becker concluded that if these comparison-group studies can be taken as the most rigorous evaluations of the effect of coaching, then "we must expect only modest gains from *any* coaching intervention" (p. 405)—on average,

about 9 points for SAT-V and 19 points for SAT-M.

Finally, the review by J. A. Kulik, Bangert-Drowns, and C. C. Kulik (1984) helps by viewing the SAT within the context of other tests. The average effect of coaching for a variety of other aptitude tests was estimated to be nearly three times the average effect for the SAT. This is not surprising because, unlike questions used in some aptitude tests, SAT questions are selected partly on the likelihood that they will not be susceptible to short-term coaching. Because questions that use complex formats have been shown to be more coachable than those using simpler ones (Powers, 1986), complicated formats are not used in the SAT. However, because question formats cannot always be simplified completely, the College Board now also provides a substantial number of materials to ensure that all prospective test takers have ample opportunity to become familiar with each of the question types that is used in the SAT.

Unresolved Issues. There are still many things that are not, and may never be, known with any absolute certainty about the effects of coaching for the SAT. There are simply too many kinds of coaching, too many kinds of students, and too many difficult-to-control variables to make any comprehensive evaluation feasible. Nonetheless, the meta-analytic summaries have shed some light on several important aspects of the effectiveness of coaching for the SAT. They also provide a context for judging the credibility of new studies. As further studies are completed, the results should be evaluated, to the extent possible, in terms of what these summaries have revealed. New studies with findings that deviate dramatically from the existing summaries may require very careful examination of the coaching methods used and the research designs on which the results are based. In particular, programs that appear especially effective but that (a) are short-term in duration or (b) have not been studied with any scientific rigor should be viewed cautiously until they can be verified.

Recent Studies

Since the various meta-analytic summaries were published, several additional studies of coaching for the SAT have in fact been completed (or had been completed previously but were not included in the summaries). These additional studies have examined several coaching programs and have used a number of alternative methods to estimate effects.

A Common Problem. All of these studies have included uncoached groups for purposes of comparison, but, as in earlier studies, these comparisons have varied with regard to their degree of scientific rigor. Random assignment of students to coached and uncoached conditions has been attempted infrequently and with mixed success. When students have been permitted to self-select coaching conditions, as is typical, not one of the individual studies has completely controlled for the possibly numerous important differences between coached and uncoached students. Indeed, completely adequate controls may not be possible.

Critical between-group differences may involve such obviously meaningful factors as the extent to which coached and uncoached students also undertake other forms of test preparation that are available to them. Some evidence suggests that the same factors that lead students to seek formal coaching may also cause them to use other resources in their preparation for the SAT. For example, students who attend coaching programs appear more likely than their uncoached counterparts to undertake a review of subject matter, to read test preparation books, and to attend review sessions given by their schools (Powers, 1981). To the extent that any of these other concomitant preparation strategies is effective, their use will confound the estimates of the effects of a given coaching program.

New Evidence. Although a completely unblemished and fully generalizable coaching study is unlikely, there are, nonetheless, several recent efforts that are highly informative, especially when considered collectively. Only studies that have employed comparison groups are discussed here. Some other less rigorous efforts have been critiqued by Smyth (1990).

At Deerfield Academy (in Massachusetts), Fraker (1986–87) studied the effects of a program conducted by a New York City-based commercial coaching company known as the Princeton Review. The simple SAT score gains made from January 1986 to November 1986 by 19 coached students were compared with those made by 119 students from the same school. Both groups were above average with respect to initial SAT scores, and uncoached students had significantly higher scores initially on both measures. Gains made by coached students over the 10-month period were 16 points greater on the math scale but 16 points less on the verbal scale than the gains made by uncoached students.

Whitla (1988) compared student-reported SAT score increases made by coached and uncoached students who subsequently enrolled at Harvard University. These students were well above average with respect to SAT performance, with uncoached students somewhat higher than coached students initially. Coached students reported having attended various in-school and commercial coaching programs, including those offered by Stanley Kaplan, Inc., and the Princeton Review. The score gains of coached students were greater than those of uncoached students by 11 points on the verbal scale and 16 points on the math scale. There were no significant differences among test preparation coaching enterprises with regard to average score gains.

Zuman (1988) studied the effects of the Princeton Review program for two small groups of New York City eleventh graders, one consisting of low-income minority students. One group was well above average on both SAT score scales, and the other was below average on both. Zuman attempted to constitute equivalent comparisons for each of these groups by recruiting students and randomly assigning them to early and later coaching. The intention was to compare the scores obtained by one group before coaching with those obtained by an equivalent group after coaching. This attempt was only partially successful, in that the comparison samples within each group were quite similar at the start of the study with respect to SAT performance.[4] However, because of

significant attrition the randomized design was not completely maintained. Furthermore, the results suggest that students in the comparison groups may have lacked motivation to do their best on the facsimile SAT, because their scores decreased slightly upon retesting.

Instead of simply comparing gain scores, Zuman employed regression analysis, using a variety of information about students' background characteristics to provide more statistically precise estimates. Given the study limitations mentioned above, the estimated effects were 52 verbal points and 58 math points for the first group, and 0 verbal and 57 math points for the low-income group.

For students in 10 private schools in the Philadelphia, PA, area, Snedecor (1989) compared the score gains made by 271 uncoached students with those made by 264 students who had attended 1 of 10 or more commercial coaching programs. Both groups had above-average SAT scores, with coached students reporting scores that were slightly lower initially than those reported by uncoached students. Average gains made by coached students exceeded those made by uncoached students by 15 points on the math portion of the test. Each group exhibited equal average gains on the verbal scale. The author reported that although some programs performed better than others, none showed dramatic results.

In a similar effort, Smyth (1989) examined the scores of 200 coached and 238 uncoached students at eight private college-preparatory schools in suburban Baltimore, MD. Students had above-average PSAT scores; coached students had PSAT scores that were somewhat lower on average than those of students who were not coached. Score improvements were defined as the difference between PSAT scores, all of which were obtained prior to any coaching, and best score on any of three subsequent official SAT administrations. By these standards, coached students gained 6 points more on SAT-V and 32 more points on SAT-M than did uncoached students. Students were coached by five or more different commercial firms. Analyses did not reveal any significant differences among the effect estimates for the major coaching enterprises.

Table 1
Estimated SAT Coaching Effects From Recent Studies

Study	Estimated effect		Number of students	
	Verbal	Math	Coached	Uncoached
Fraker (1986–87)	−16	16	19	119
Smyth (1990)	6	18	501	631
Snedecor (1989)	0	15	264	271
Whitla (1988)	11	16	341	1217
Zuman (1988)				
Group 1	52	58	21	34
Minority students	0	57	16	17
Median	3	17		

In a subsequent academic year, Smyth (1990) repeated this effort with an additional 300 coached and nearly 400 uncoached students who attended 14 independent secondary schools in Maryland and New Jersey. (Five of the schools had participated in the earlier study also.) Again, students in this study had above-average scores before being coached, and coached students had slightly lower scores initially than did their uncoached counterparts. Combining data with those from the earlier study, Smyth (1990) found that coached students gained 9 more verbal and 24 more math points than did uncoached students. When these differences were adjusted via analysis of covariance (ANCOVA) for between-group differences in PSAT scores and number of times the SAT had been taken, the more precise effect estimates were 6 and 18 points on the verbal and math scales of the SAT.

The results of all of these most recent studies are summarized in Table 1. It should be reiterated that, instead of simple test-score gains, the numbers shown are estimates of the effects of coaching above and beyond any effects resulting from growth, practice, and other factors that affect the scores of both coached and uncoached students. With some exceptions, these additional studies are generally consistent with the meta-analytic summaries. Coaching programs, even the most highly publicized ones, appear to have on average a small effect on SAT-V scores and a somewhat larger, though still modest, effect on SAT-M scores. The median effects for the studies listed in Table 1 are 3 points for SAT-V scores and 17 points for SAT-M scores. These estimates correspond closely with those computed by Smyth (1990) in the most recent, the largest (in terms of the number of coached students), and arguably the best controlled study of those reported recently. In addition, these median values are also quite close to those given by Becker (1990) of 9 and 19 points, even though only one of these more recent studies (Zuman's) was included in her estimates.

Some of the recent studies have reported results separately by program, but have not revealed any dramatic differences among major coaching enterprises with regard to effectiveness. It seems likely, then, that the major differences among individual study estimates result primarily from dissimilarities in the design and execution of the studies, and specifically in how comparisons were established, rather than from differential program effectiveness.

Nonetheless, the issue of especially effective programs may deserve more study. Two nationally-franchised commercial firms—the New York City-based Princeton Review and Stanley Kaplan, Inc.—offer programs that are generally longer in duration (40 hours or more of classroom coaching) and more expensive (currently up to $700 for the Princeton Review) than most other programs. These firms also appear to enjoy the largest share of the coaching market. There may be special interest, therefore, in the effectiveness of these particular programs.

Six studies have provided effect estimates for the Stanley Kaplan program, and eight have done so for the Princeton Review. These estimates, summarized in Table 2, are generally consistent with the estimates based on all studies considered earlier in this article. The median effect estimates for the Kaplan and Princeton Review programs suggest that these programs are hardly (if at all) more effective in improving SAT scores than are coaching programs generally. Moreover, there is not much basis in these estimates to suggest that either of these programs is any more effective than the other (although the median estimate for SAT-V appears somewhat higher for Kaplan than for the Princeton Review). A final observation is that the estimates for the Kaplan program are less variable than those for the Princeton Review. Whether this is a function of the size or design of the studies, or of the degree to which these coaching programs are conducted consistently from site to site, cannot be readily ascertained from the data at hand.

The bottom-line result of the quantitative summaries and of the most recently completed research then is as follows: Reasonably good estimates are now available about the effects of coaching for the SAT, including that provided by for-profit firms. These estimates are most assuredly better than the claims currently being made by some commercial coaching enterprises and the word-of-mouth accounts of individual test takers.

Table 2
SAT Coaching Effect Estimates for Two Commercial Coaching Programs

Source	Estimated effect: Verbal	Estimated effect: Math	Number of students coached
	Stanley Kaplan		
Wing (1987) as reported by Smyth (1990)	7[a]	24[a]	72
Snedecor (1989)	12	26	22
Sesnowitz et al. (1982)	28(14[b])	24(10[b])	≈250[c]
Smyth (1990)	19	26	93
Whitla (1988)	24[e]	18[e]	≈75[d]
Median	19(14[f])	24(24[f])	≈75
	Princeton Review		
Fraker (1986–87)	−16	16	19
Wing (1987) as reported by Smyth (1990)	−16	31	61
Snedecor (1989)	−2	4	48
Zuman (1988) minority students	0	57	17
Whitla (1988)	8[e]	13[e]	≈75[d]
Smyth (1990)	12	26	66
Zuman (1988) Group 1	52	58	34
Median	0	26	48

[a]These estimates are based on Smyth's report of a study by Cliff Wing (1987) at Wake Forest University. [b]Sesnowitz, Bernhardt, & Knain (1982) also reported estimates of 14 points for SAT-V and 10 points for SAT-M. These smaller estimates were based only on adjustments for differences between coached and uncoached students on a number of demographic and personal characteristics (e.g., rank in high school) that were related to test performance, not on earlier test scores. This adjustment was used because students scored lower on earlier tests than was expected from their demographic and personal characteristics. [c]The total number of coached students involved in the study was 514, divided among two schools. How these students were apportioned between the two coaching schools could not be determined from the report. [d]The exact number from each of these schools could not be readily determined. [e]Based on data collected by Whitla, but not reported in Whitla (1988). [f]Median when the smaller estimate from Sesnowitz et al. (1982) is used.

Implications

The decision to seek coaching for the SAT is still one that students and their parents must make individually, but this deliberation ought to be based on the best evidence available. Potential consumers should consider not only the likely benefits but also the expected costs. Costs may entail significant financial outlays, but equally important, lost opportunities. A legitimate question for test takers to ask then is, "What could I do, in the 40 or more hours that I might spend at a coaching school, to improve not only my chances of being accepted at the college of my choice but also my chances of succeeding at that college once enrolled? Would I, for example, be better served by concentrating on developing my subject-matter knowledge than my test-taking skills?" This question probably has more than one correct answer, and will undoubtedly vary from student to student.

With respect to benefits, the effects of coaching may be less than many students suppose. Although coaching schools promise large *score increases*, there are no real guarantees in terms of the actual *effects* of coaching. Test takers should be made aware of this critical distinction. Assuming that coaching improves scores by about 10 points on the verbal and about 20 points on the math portion of the SAT, a coached student can expect to improve more than his or her uncoached counterpart about 6 times in 10 on SAT-V and about 6–7 times in 10 on SAT-M.[5] On the other hand, about 4 times in 10 for SAT-V and 3–4 times in 10 for SAT-M, the typical *uncoached* student can be expected to exhibit larger score increases than one who has been coached. This is a better statement of the real "guarantee," which seems more in line with students' actual experiences. When they have been asked to give their opinions, less than a majority of coached students have said they were satisfied with their score changes—for example, 24% of those polled by Snedecor (1989) and 43% of those surveyed by Whitla (1988).

Messick (1982) suggested that

improvements in percentile ranking might serve as a gauge of the practical impact of coaching. Using the most recent data for college-bound seniors (College Board, 1991) as a basis, it appears that improvements of 10 points on SAT-V and 20 points on SAT-M will do relatively little to improve a "typical" (SAT-V = 420, SAT-M = 470) test taker's standing. Improvements of 10 and 20 points will push this test taker's ranking ahead of only a small proportion of additional test takers—from the 48th to 53rd percentile rank on SAT-V and from the 48th to the 54th on SAT-M. At higher as well as lower score levels, these rankings would change even less. For instance, improving an SAT-V score from 600 to 610 would raise one's standing from the 93rd to the 94th percentile rank, and going from 680 to 700 on SAT-M corresponds to an increase from the 94th to the 96th percentile rank. At the lower score levels, increasing an SAT-V score from 250 to 260 is equivalent to going from the fifth to the sixth percentile rank; increasing a 290 SAT-M score to 310 will improve the percentile rank from 5 to 8. These figures are, of course, for all college-bound senior SAT takers. Because applicants to individual colleges may as a group have less variable scores than do test takers in general, score increases may have a somewhat greater impact on relative standing within applicant pools.

In conclusion, it is hoped that this discussion of summaries, enhanced by a description of recent, individual studies of coaching for the SAT, will in some small way be useful to those who counsel students about preparing for the SAT. Test takers who contemplate undertaking coaching should be helped to critique the claims made by major commercial companies and to ask for explanations of any discrepancies between the assertions made by these companies and the conclusions of the several scholarly summaries discussed here.

Notes

This article is a considerably expanded version of a "Brief Overview" paper that was supported by funding from the Joint Staff Research and Development Committee of ETS and the College Board. The views are the author's and do not necessarily reflect those of either ETS or the College Board. Special thanks go to Ann Jungeblut, Michael Zieky, three anonymous reviewers for thoughtful comments, and Ruth Yoder for her careful handling of the production of this manuscript.

[1]Smyth (1990) provides an informative critique of the advertising practices of several commercial companies. He gives an interesting account of the ways in which these companies use the available scientific evidence to make their claims.

[2]It is interesting to note that this study is one of the very few such studies that has attempted to determine the locus of any coaching effect. The investigators noted that after being coached, test takers were able to reach more questions on both the verbal and the math tests. Correctly answering a reasonable fraction of these items may have accounted for a substantial portion of the test score improvements that were observed in the study.

[3]Probably the most appropriate and informative comparison is between students who receive extensive coaching and those who use less costly and less time-consuming resources, such as the free test familiarization provided by the College Board to all SAT takers.

[4]Actually, the study was based on self-reported PSAT scores, scores from a special administration of a retired form of the SAT (for which scores were not reported to colleges), and students' actual SAT scores.

[5]These estimates are based on a method suggested by McGaw and Wong (1992). Estimates used for standard deviations of test score gains (45 for SAT-V and 52 for SAT-M) are those reported by Donlon (1984) for a number of SAT testing years, and they correspond to the standard errors of differences reported most recently by the College Board (1991).

References

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, 36*(10), 1086–1093.

Becker, B. J. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research, 60*, 373–417.

Bond, L. (1989). The effects of special preparation on measures of scholastic ability. In R. L. Linn (Ed.), *Educational measurement*. New York: American Council on Education and Macmillan Publishing Company.

Cole, N. (1982). The implications of coaching for ability testing. In A. K. Wigdor and W. R. Gardner (Eds.), *Ability testing: Uses, consequences, and controversies. Part II: Documentation section*. Washington, DC: National Academy Press.

College Board (1991). *APT guide 1991–92 for high schools and colleges*. New York: College Board Publications.

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

Donlon, T. F. (Ed.) (1984). *The College Board's technical handbook for the Scholastic Aptitude Test's and Achievement Tests*. New York: College Entrance Examination Board.

Fraker, G. A. (Winter 1986–87). The Princeton Review reviewed. *The Newsletter*. Deerfield, MA: Deerfield Academy.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press, Inc.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage Publications.

Johnson, S. T., Asbury, C. A., Wallace, M. B., Robinson, S., & Vaughn, J. (1985, April). *The effectiveness of a program to increase Scholastic Aptitude Test scores of Black students in three cities*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago.

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95*, 179–188.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge: Harvard University Press.

McGaw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361–365.

Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist, 17*, 67–91.

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

Powers, D. E. (1988). *Preparing for the SAT: A survey of programs and resources* (College Board Rep. No. 88–7 and ETS Res. Rep. No. 88–40). New York: College Entrance Examination Board.

Powers, D. E. (1986). Relations of test item characteristics to test preparation/test practice effects: A quantita-

Continued on page 39

Continued from page 30

tive summary. *Psychological Bulletin, 100*, 67–77.

Powers, D. E. (1981). Students' use of and reactions to alternative methods of preparing for the SAT. *Measurement and Evaluation in Guidance, 14*, 118–126.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage Publications.

Sesnowitz, M., Bernhardt, K. L., & Knain, D. M. (1982). An analysis of the impact of commercial test preparation courses on SAT scores. *American Educational Research Journal, 19*, 429–441.

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

Smyth, F. L. (1990). SAT coaching: What really happens to scores and how we are led to expect more. *The Journal of College Admissions, 129*, 7–16.

Smyth, F. L. (1989). Commercial coaching and SAT scores: The effects on college preparatory students in private schools. *The Journal of College Admissions, 123*, 2–7.

Snedecor, P. J. (1989). Coaching: Does it pay—revisited. *The Journal of College Admissions, 125*, 15–18.

Whitla, D. K. (1988). Coaching: Does it pay? Not for Harvard students. *The College Board Review, 148*, 32–35.

Wilder, G., Casserly, P. L., & Burton, N. W. (1988). *Young SAT-takers: Two surveys* (College Board Rep. No. 88–1). New York: College Entrance Examination Board.

Wing, C. W., Jr. (1987, April). *Some field observations of the impact of test preparatory programs on high school students' Scholastic Aptitude Test scores*. Winston-Salem, NC: A report to the Awards Committee for Education and Wake Forest University.

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