DOCUMENT RESUME

ED 395 025 TM 025 042

AUTHOR	Baydar, Nazli
TITLE	Effects of Coaching on the Validity of the SAT: A Simulation Study.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-90-4
PUB DATE	Apr 90
NOTE	70p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Change; College Entrance Examinations; *College Freshmen; Estimation (Mathematics); Higher Education; *Predictive Validity; *Scores; *Selective Admission; Simulation; *Test Coaching; Testing Problems; Test Results
IDENTIFIERS	*Scholastic Aptitude Test

ABSTRACT

The effects of student coaching in preparation for the College Board Scholastic Aptitude Test (SAT) on the predictive validity of this test for freshman year performance were studied using data on 1985 freshman year students from four colleges. After the validity of the SAT was estimated for each school, a given proportion of students was picked, and a given magnitude of score gains was added to the observed SAT verbal and mathematics scores. The validity of the new SAT scores was then estimated and compared to the observed validity measures. Each simulation provides an estimate of the change in validity resulting from a hypothetical coaching effect. Sample sizes from the 4 colleges were: (1) 996 students from a highly selective school; (2) 1,346 from a moderately selective school; (3) 386 from a small, less selective school; and (4) 203 from a small religiously affiliated college. Results indicated that the validity indicators of the SAT in highly selective colleges are more contaminated by coaching effects than in less selective colleges. Results also indicate that changes in the levels of coaching do not seem large enough to account for the observed decline in the validity of the SAT over the past 10 years. Appendix A provides a detailed discussion of the second college, and Appendix B describes decrements in multiple partial correlation coefficients. (Contains 5 figures, 14 tables, and 19 references.) (SLD)

RESEARCH REPORT

EFFECTS OF COACHING ON THE VALIDITY OF THE SAT: A SIMULATION STUDY

Nazli Baydar

Educational Testing Service
Princeton, New Jersey
April 1990

EFFECTS OF COACHING ON THE VALIDITY OF THE SAT: A SIMULATION STUDY

NAZLI BAYDAR

December 19, 1989

INTRODUCTION

This report presents a simulation study of the effects of student coaching in preparation for the College Board Scholastic Aptitude Test (SAT) on the predictive validity of this test for the freshman year performance. The SAT is designed to measure long-term developed abilities of verbal and quantitative reasoning and comprehension (Donlon, 1984; Messick, 1980). The predictive validity of the SAT for forecasting college freshman year performance depends on two distinct components: (1) The extent to which the SAT scores reflect the verbal and quantitative reasoning abilities of students; and, (2) the extent to which these abilities are reflected in college performance as indicated by freshman year grades.

A recent study (Morgan, 1989) has documented a downward trend in the predictive validity of the SAT over the past ten years. During this period the correlation of SAT-Mathematical with college freshman grades, SAT-Verbal with freshman grades, as well as the multiple correlation of SAT-Math and SAT-Verbal with freshman grades have declined by about .04 (8-9%).[1] There could be a number of alternative explanations for this decline. These include possible changes during the past decade in what the SAT measures, in what the college grades reflect, in the sample of students and institutions comprising the yearly validity data bases, and in the educational practices entailed in preparation for the SAT. The latter concerns the possibility that increases in the incidence and

---

[1] These figures are based on the correlation coefficient estimated by employing a correction for multivariate restriction of range. The ten-year decline in the uncorrected multiple correlation coefficient is 16%.

effectiveness of coaching for the SAT may have systematically degraded the SAT's predictive validity over time.

This report presents an estimation of the possible impact of increases in student coaching on the predictive validity of the SAT by means of simulations. These simulations cannot provide conclusive evidence on the existence and the magnitude of the effect of coaching on the validity of the SAT. However, they can provide plausible bounds on the magnitude of the changes in the SAT validity coefficients, given certain assumptions on modelling and assumptions on the magnitude of the incidence and effectiveness of coaching.

For a correct interpretation of the results of the simulations, it is important to bear in mind how we define coaching effects. One may distinguish three types of score gains due to coaching (Messick, 1982): (1) gains due to test taking familiarization and reduced anxiety; (2) gains due to genuine improvement in reasoning and comprehension skills; and, (3) gains due to learning test-specific tricks and strategies. The first type of score gain resulting from test familiarization improves the validity of the SAT. These gains lead to valid increases in the test scores of students who might have obtained scores below their ability levels, because of their unfamiliarity with the test procedures and their apprehension over being tested. The second type of score gains do not affect the validity of the SAT. Instructional test preparation programs that result in genuine improvement in skills should lead to an improved college performance as well as improved test scores. The third type of score gains (i.e., those due to learned tricks and strategies of selecting answers), lead to spurious increases in test scores that affects the validity of the SAT negatively.

When studying the impact of coaching on the predictive validity of the SAT it is important to distinguish between these three types of score gains. However, score gains over time could occur for other reasons such as academic growth, increased motivation, test practice and stochastic fluctuation of the scores. Therefore, it is important to distinguish between total score gains of students and coaching effects. Coaching effects refer to the average difference between the score gains of coached students and the score gains of comparable uncoached students over the same time period. Therefore, conceptually, coaching effects exclude gains due tc maturation, increased motivation, and stochastic fluctuations. Unfortunately, not all studies on coaching effects adequately control for these sources of score gains. Moreover, the results of some studies are further contaminated by the self-selection of coached and uncoached students. The claims of the coaching schools, on the other hand, refer to total score gains of students over time, and therefore they include gains due to maturation, increased motivation and stochastic fluctuations.

The best information on the effects of coaching are based on pre- and post test studies of coached and uncoached students, who are assigned to these groups randomly (Messick & Jungeblut, 1981). It is very difficult, if not impossible, to separate coaching effects from score increases due to (i) learning and maturation that occurs independently of coaching; (ii) test familiarization; (iii) instruction in skills that are relevant to college performance; (iv) stochastic fluctuations of the test scores; and (v) in non-randomized studies, self-selection of students who enroll in the coaching programs who differ from non-enrollers by observed background characteristics (e.g. parental income) and by unobserved personality attributes (e.g. motivation). It is clear that estimation of the magnitude of the score gains of Types 1, 2, and 3 presents a non-trivial problem.

In the simulations presented below, the assumptions on the magnitude of score gains due to coaching are based on various sources that do not distinguish between coaching effects and total score gains due to all other reasons. However, the score gains in these simulations are assumed to be only of Type 1 and Type 3 gains, that do not lead to a corresponding increase in the freshman year performance. Note that Type 1 gains, i.e. the score gains due to test familiarization could be achieved in a variety of ways other than coaching.[2] Therefore, these simulations are likely to over-represent the impact of coaching. Even if score increases based on test specific information on tricks and strategies were possible, the magnitude of such increases is not likely to be large, especially since the ETS makes an effort to render the applicability of such strategies difficult. The simulations presented here can be regarded as the "worst possible scenario," in that they represent possible changes in the predictive validity of the SAT if all score gains reported by studies on coaching and claimed in coaching schools' advertisements were score gains resulting from test

[2]Many high schools provide SAT familiarization courses (Powers, 1988). Additionally, textbooks and a variety of printed, computer based and audio-visual documents published by the College Board provide effective test familiarization (College Board, 1989).

familiarization and trick learning and not resulting from instruction in skills nor from academic growth.

The mechanics of the simulations are fairly simple. Data on 1985 freshman year students from four colleges are used to simulate the effects of coaching on the validity of SAT. After the validity of the SAT for each school is estimated, a given proportion of students are picked and a given magnitude of score gains are added to the observed SAT-Verbal and SAT-Math scores of each student. The validity of this new set of SAT scores is then estimated and compared to the observed validity measures. Each simulation provides an estimate of the change in the validity of the SAT resulting from a hypothetical coaching effect of a given size for a given proportion of the freshman students. Note that an association between coaching effects and first year average is not introduced in the simulations. Therefore, by the design of the simulations, the assumed coaching effects do not include Type 2 gains (i.e. gains due to genuine skills improvement).

The observed SAT scores that are used in the simulations are probably already contaminated by coaching effects. Although the probable presence of coaching effects in the data is a disadvantage to using real data for the simulations, using entirely simulated data would have other disadvantages. For example, observed data on student background characteristics, academic performance, and SAT scores already reflect unmeasured associations due to interactions between these variables and their common unobserved determinants (such as motivation). If simulations were based only on modelled data, introduction of such interactions and unobserved determinants would be very difficult, and the relevance of the results of such simulations would be limited. The presence of coaching effects in the observed data will affect the observed validity coefficient. This study involves the comparison of this validity coefficient to the coefficient estimated after a known level of coaching effects are simulated.

METHODS

A. Students who are likely to receive coaching

Using data collected by Powers (1988) from a stratified random sample of SAT takers, descriptive statistics were obtained comparing the students who have been coached for the SAT and those who have not been coached. Coaching status of each student is self reported. The comparative analysis of coached and uncoached students led to a simple logistic model of the probability of being coached for each student depending on a set of background characteristics. In the simulations presented below, this logistic model determines the selection of students to be "coached," i.e., those students whose SAT scores will be inflated in order to simulate coaching effects. A selection process on the basis of background characteristics will bring the simulations closer to reality than simple random selection.

The estimates of the probabilities of being coached are based on data from four random samples of high school students (juniors and seniors) who registered to take the SAT during 1986 and 1987. From 2,378 sample members, 54% returned the questionnaires that requested information on SAT preparation. Females, Asian-Americans, students aspiring for advanced degrees and students who ranked in the top 10% of their high school classes were over represented among the students who returned their completed questionnaires. Further details on the data can be found in Powers (1988). For the purposes of modelling probabilities of being coached, we define being "coached" as having attended an SAT preparation program either outside of one's high school or at one's high school but conducted by an outside organization.

From 1,287 students who responded to the survey, the coaching status of 1,243 students could be identified. Most (86%) of these students did not receive any coaching. Among 176 students who received coaching, 89% were coached for both SAT-Verbal and SAT-Math. Only 9% received coaching for SAT-Verbal and not SAT-Math, while 3% received coaching for SAT-Math and not SAT-Verbal. Slightly under half of the 1,243 students took the SAT once, while the remaining students took the SAT twice or more. Table 1 shows the characteristics of coached and uncoached students.

The similarity of the ethnic distribution of the coached and uncoached students is inconsistent with previous research on SAT coaching that found lower proportions of ethnic minorities among coached students than among uncoached students (FTC, 1979). There are almost equal proportions of Black or Hispanic students among coached and uncoached students (11% and 10% respectively) in the Powers sample. Similarly, there are equal proportions of coached students among Black/Hispanic students and students from other ethnic groups (13% and 12%, respectively). It is possible that this inconsistency with previous coaching studies results from the earlier studies having been based only on commercial coaching school data, whereas the Powers data also include students who have been coached in community organizations and other less costly programs. Additionally, the FTC study dates almost a decade earlier than the Powers study, and it is likely that test preparation and coaching are more available to minority students now than a decade ago.

Table 1 shows that the students whose fathers are reported to have higher levels of education and those who have higher family incomes are more likely to have received coaching. These findings are consistent with previous research (FTC, 1979). Students who received coaching have somewhat lower self-reported high school GPAs; however, this difference is not large ($p<.10$). Students who received coaching, on the average, took the SAT more times than those who were not coached. This suggests that students who are not satisfied with their scores during their junior year might seek coaching subsequently. Indeed the mean first SAT scores of the coached students are lower than the mean first SAT scores of the uncoached students by 17 points on the verbal section and 11 points on the math section. At the last administration of the SAT, students who received coaching are behind those who did not receive coaching by 8 points on the verbal section; however, they are ahead of those who did not receive coaching by 5 points on the math section. When the score gains of coached and uncoached students are computed for those students who took the SAT at least twice, it is found that the score gains of the coached students exceed the score gains of the uncoached students by 5 points on SAT-Verbal ($t=1.74$, $p<.10$) and 21 points on SAT-Math ($t=6.73$, $p<.001$).

An estimate of the probability of being coached on the basis of the Powers data might give us an overestimate of the true probability in the population due to the biases introduced by the subgroup differences in response rates to the survey. For example, socioeconomic status of the respondents is higher than the socioeconomic status of all SAT takers. The homogeneity of the background characteristics of the respondents of the Powers study might lead to a slight downward bias in the estimated coefficients of the logistic model that quantifies the association between the probability of being coached and these characteristics. Nevertheless, the background differences between coached and uncoached students are well documented and substantial, and it is desirable to retain these in the simulations. In order to represent these differentials, the simulations below are based on the following logistic model of probabilities of being coached, estimated from the Powers data:

$$\ln \frac{p}{1-p} = -2.1097 + (.2428\ \text{FATHER'S EDUCATION}) + (.1288\ \text{PARENTAL INCOME}) - (.2250\ \text{HIGH SCHOOL GPA}),$$

where p is the probability of being coached. All independent variables in this equation are standardized to have means equal to zero. The intercept term (-2.1097) is the log-odds of receiving coaching for an average student in the sample. The estimated intercept implies that an "average" student in Powers sample has probability of 0.11 of being coached.

Alderman and Powers (1980) found that about 3% of test takers attended coaching sessions outside of school for the December 1977 administration of SAT. Powers' 1986 and 1987 data show that 14% of the students report having received coaching either outside school or inside school but given by a separate organization. These figures indicate that there has been an increase in the proportion of students who seek and receive coaching for the SAT during the past ten years. Whitla (1988) reports that 22% of Harvard students had been coached. Since Harvard students come from a selective group who have background characteristics that are associated with higher levels of coaching, it is expected that the proportions of coached students at Harvard and other highly selective colleges are higher than the same proportion in a random sample of SAT takers. Nevertheless, it is important to note that in highly selective schools almost a quarter of the students might have received coaching.

B. Simulated increases in SAT scores due to coaching

Assessment of the effectiveness of coaching has been subject to dispute. There are few reliable studies on the effectiveness of SAT coaching. Meta-analyses by Messick and Jungeblut (1981) and DerSimonian and Laird (1983) summarize previous studies on the magnitude of coaching effects that have a wide variety of program and study designs. Messick and Jungeblut summarize their study with a regression analysis of coaching effects on (the logarithm of) hours spent in coaching programs. The mean coaching effects are about 23 points for Math and 14 points for Verbal in this study. DerSimonian and Laird provide estimates of coaching effects adjusted for inter-study correlations. Their coaching effect estimates are 18 points for Math and 19 points for Verbal. Both the Messick-Jungeblut and the DerSimonian-Laird estimates could be somewhat high since they include data from full-time post-high school preparatory schools such as those reported by Marron (1965). The estimates based on full-time programs are expected to reflect genuine skills improvement (Type 2) effects as well. When the data from such long term programs are excluded, Messick-Jungeblut estimates of coaching effects are 16 and 14 for SAT-Math and Verbal respectively. The magnitude of coaching effects estimated by Messick and Jungeblut and DerSimonian and Laird are somewhat lower than those claimed by the FTC (1979) study which were 20-33 points for SAT-Math and 27-34 points for SAT-Verbal, yielding an average of 25 and 32 points for Math and Verbal respectively (Stroud, 1980).[3] The Harvard study (Whitla, 1988) found smaller net effects of coaching: 16 and 11 points for Math and Verbal respectively. Smyth (1989) has also found increases (15-35 points) in Math scores due to coaching but almost no increases in Verbal scores. Note that these effect estimates include Type 1 (test familiarization), Type 2 (genuine improvement of skills) and Type 3 (trick learning) effects, but not score gains due to maturation.

The analysis of the FTC (1979) data showed that there was substantial variation in the score gains of coached students even within a single coaching program. In a sample of coached students, the sources of variation in coaching effects could be threefold: (i) due to differences between subjects in their level of motivation or ability to retain information given by the programs; (ii) due to differences between coaching programs; and, (iii) due to differences in the number of hours invested in the programs. No information about the first component is available, and in practice, differences due to intersubject variation would be impossible to distinguish from the stochastic component of the

[3]Note that except for the FTC study and DerSimonian-Laird study, all coaching effect studies estimate larger coaching effects for SAT-Math than SAT-Verbal. A reanalysis of the FTC data accounting for differential growth rates of coached and un-coached students estimated an SAT-Verbal coaching effect of 17 points (Rock, 1980), which is smaller than the SAT-Math effect.

SAT scores.

DerSimonian and Laird (1983), using techniques of meta-analysis, estimated the variance due to program differences in a group of programs that various studies reported. Accounting for inter-study correlations, they estimated the standard deviation of coaching effects on SAT-Math as 13.2 points and SAT-Verbal as 16.5 points, with means of 17.7 and 19.3, respectively. However, these estimates reflect a combination of variance due to program differences and variance due to hours spent in each program, since hours were not controlled for in this study. The Messick-Jungeblut meta-analysis introduces a control for hours, but it is not possible to partition the remaining variance in coaching effects into components due to differential study designs, due to differential methods of estimation, and due to program differences.

Better information exists on the variance of the coaching effects due to differences in the number of hours spent in coaching programs than due to program differences. Data from Powers' 1986-87 study show that there is substantial variation of hours spent in coaching among the coached students. The coached students reported having received means of 12.9 and 12.5 hours of coaching for SAT-Math and SAT-Verbal with standard deviations of 11.2 and 11.0 hours respectively. The variation in coaching effects due to the variation in hours could be estimated using the regression equations reported by Messick and Jungeblut (1981)[4] that relate the logarithm of hours spent in coaching programs to coaching effects. The estimated standard deviations of coaching effects due to

---

[4]Regression equations of coaching effects are:
SAT-Math = $-14.072 + 26.646 \log_{10}$ (hours spent for SAT-M coaching)
SAT-Verbal = $-6.587 + 15.155 \log_{10}$ (hours spent for SAT-V coaching).

differential hours spent in coaching programs are 10.7 math points and 5.7 verbal points with corresponding mean coaching effects of 11.3 and 7.8 math and verbal points respectively. The coefficients of variation (standard deviation divided by mean) are .954 and .734 for math and verbal components.

It is unlikely that coaching effects on SAT-Math and SAT-Verbal for a given student will be independent. From the Messick-Jungeblut study, it is known that the number of hours spent in coaching are highly correlated with coaching effects, and from the Powers data, it is known that the number of hours spent on Verbal and Math coaching are highly correlated. There is, however, no direct evidence on the magnitude of the correlation between Math and Verbal coaching effects. This correlation might stem from two sources: (1) the individual level coaching effects on both components of the SAT might reflect the same underlying factors such as motivation and ability to retain information; (2) students are likely to receive coaching for both components from the same program, which is likely to adopt a uniform approach to Math and Verbal coaching. In view of these factors, it appears reasonable to assume that SAT-Math and SAT-Verbal coaching effects are correlated.

C. Some simple formulas that describe the relationship between coaching and the predictive validity of the SAT

A simplified model of coaching allows one to see the probable impact of various parameters that describe coaching effects on the predictive validity of the SAT. Consider the following simplifications:

(1) Coaching effects (c_i, $i=1,\ldots,n_c$) are normally distributed with mean μ_c and variance σ_c^2.

(2) The proportion of coached students is p (n_c/N).

(3) Coaching effects and the underlying (uncoached) SAT scores are independent.

(4) Coaching effects and freshman GPAs are independent.

Under these assumptions, the correlation r^* between freshman GPA and the coached SAT-total score is:

$$r^* = \frac{\Sigma\ (F_i - \bar{F})(S_i^* - \bar{S}^*)}{N\ \sigma_F\ \sigma_{S*}},$$

where S_i^*, ($i=1,\ldots,N$) are the SAT scores when proportion p of N students are coached and F_i are the freshmen GPAs; $\bar{S}^*$ and $\bar{F}$ represent the means, and σ_{S*} and σ_F represent the standard deviations of these two quantities, respectively. Similarly, $\bar{S}$ and σ_S represent the mean and the standard deviation of the underlying uncoached SAT scores.

The expected value of σ_{S*} is:

$$E\ [\sigma_{S*}] = \sqrt{\sigma_S^2 + p\sigma_c^2 + p\ (1-p)\ \mu_c^2}$$

The expected covariance of F_i and S_i^* is:

$$E\ [(F_i - \bar{F})(S_i^* - \bar{S}^*)] = E\ [(F_i - \bar{F})(S_i - \bar{S})] + E\ [(F_i - \bar{F})(C_i - p\mu_c)],$$

where C_i are the coaching effects for all students ($C_i=0$ if a student is not coached, $C_i = c_i$ if coached).

Since the covariance of freshman GPA with coaching effects are assumed to be zero by definition, the covariance of coached SAT scores with freshman GPA will be equal to the covariance of uncoached SAT scores with freshman GPA. Hence, the expected decline in validity can be expressed as the ratio:

$$\frac{r^*}{r} = \frac{\sigma_S}{\sqrt{\sigma_S^2 + p\sigma_c^2 + p(1-p)\,\mu_c^2}},$$

where r is the correlation between uncoached SAT scores and freshman GPA. Under these simple assumptions, we see that decline in the validity gets larger as the variance and mean of the coaching effects increase and as p approaches 0.5.

In the simulations presented in this report, the coaching effects are added in two parts, for Math and Verbal separately. In this case, the mean and the variance of the coaching effects on each component and their correlation determine the effect of coaching on the predictive validity. Let $\bar{m}$ and $\bar{v}$ represent the mean coaching effects on SAT-Math and SAT-Verbal respectively, and let σ_m^2 and σ_v^2 be the variances of these effects among the students who are coached. Then, the ratio of the correlation of coached SAT scores with freshman GPA to the correlation of the uncoached SAT scores with freshman GPA can be written as:

$$\frac{r^*}{r} = \frac{\sigma_S}{\sqrt{\sigma_S^2 + p\,(\sigma_m^2 + \sigma_v^2 + 2\sigma_m\sigma_v r_c) + p\,(1-p)\,(\bar{m} + \bar{v})^2}}$$

where r_c is the correlation between Math and Verbal coaching effects. As r_c approaches unity, the ratio of r^* to r gets larger.

Figures 1.a to 1.e give the sensitivity of the ratio of r^* to r, to varying values of p, σ_m^2 and σ_v^2, r_c, $\bar{m}$ and $\bar{v}$, and σ_S^2. It is easily seen that p, $\bar{m}$ and $\bar{v}$ have the strongest effects on the validity of SAT when other parameter remain constant. Figure 2 shows the combined impact of p, $\bar{m}$ and $\bar{v}$ on the validity of the SAT, based on the above equation. The correlation between Math and Verbal coaching effects has the least impact on the validity of SAT.

Note that all the formulas and the Figures 1 and 2 are based on an important simplifying assumption, i.e., that the coaching effects are independent of the underlying (uncoached) SAT scores. This assumption is likely to be violated in reality due to: (i) the association between the choice of being coached and prior SAT scores; (ii) the association between the score gains due to coaching and the student characteristics, which, in turn, are associated with the SAT scores. Since the above equations do not allow for an association between the coaching effects and the freshman GPA (assumption (4) above), the coaching effects represented in this simplified model are only Type 1 and Type 3 gains. There are no studies that report on the magnitude of such gains; however, they are not expected to be large. The simulations presented in this report incorporate interdependencies of the SAT scores and the coaching effects due to common background determinants. Therefore, they are likely to provide a more realistic picture of the effects of coaching on the validity of the SAT than the simple analytic relations presented above.

D. Parameters of the simulations

Below, the parameter values that underlie the simulations are listed. Two sets of parameters determine an array of 12 variants of simulations for each of the four colleges for which the simulations were carried out. These parameters describe the assumed levels of the probability that a student will be coached and the effectiveness of coaching.

1. Probabilities of being coached

In the simulations, probabilities of being coached are determined by three background characteristics: father's education, income, and high school GPA. The logistic regression equation estimated using the Powers data generates the probability of being coached for each student (see Section A above). Different levels of coaching are simulated by manipulating the intercept term of the logistic regression. For each school the background variables are standardized to have means of zero so that the probability that an average student is coached remains approximately equal across colleges for a given variant of simulations. Three levels of probabilities of being coached were used in the simulations:

-- An estimate of the level of coaching ten years ago:	5%
-- The approximate current level of coaching:	15%
-- A high level of coaching, observed in some very selective colleges:	25%

Alderman and Powers (1980) found that about 3% of test takers attended coaching programs outside of school at the December 1977 administration of the SAT. A survey of SAT taking juniors in 1978 (Powers & Alderman, 1979; 1983) revealed that about 5% had been coached. Powers' 1986-87 data on juniors and seniors (Powers, 1988) indicate that 14% of the respondents were coached. Therefore the estimates that 5% of students were coached 10 years ago, and 15% receive coaching now, appear reasonable. The high variant of proportions of students coached reflect Whitla's (1988) finding that 22% of Harvard freshmen received coaching.

2. Effectiveness of coaching

On the basis of the findings of previous studies cited in section B above, four levels of coaching effects will be simulated. It is assumed that effectiveness is normally distributed with a given mean and variance. The variance for each level of effects is computed using the same coefficient of variation. The coefficients of variation represent differentials in the coaching effects due to differences in hours spent in coaching. The coefficients of variation are .954 and .734 for math and verbal coaching effects, respectively (see Section B, above). This approach ensures that the dispersion of the normal distribution of coaching effects remains the same between different variants of the simulations.

When the coefficient of variation is close to one, and random coaching effects are drawn from a normal distribution, it is possible that some score changes will be large and negative. Although a fair number of studies that document score losses following attendance at a coaching program (FTC, 1979), it was decided to put a restriction on the magnitude of the simulated negative "coaching effects." According to the Messick-Jungeblut study, a minimum of 1 hour of coaching is expected to result in losses of 14.1 points on SAT-Math and 6.6 points on Verbal. In the simulations all negative effects smaller than -14.1 and -6.6 for Math and Verbal respectively will be truncated at -14.1 and -6.6 points. This procedure ensures that large score losses assocaiated with coaching are impossible in the simulations. Due to this truncation, the simulations presented here will be likely to overrepresent the impact of coaching on the validity of the SAT.

The four levels of coaching effects that underlie the simulations are as follows:

1. One standard deviation unit below Messick-Jungeblut mean estimates:	12 points SAT Math 8 points SAT Verbal 20 points SAT Total
2. Messick-Jungeblut mean estimates:	23 points SAT Math 14 points SAT Verbal 37 points SAT Total
3. One standard deviation unit above Messick-Jungeblut mean estimates:	34 points SAT Math 20 points SAT Verbal 54 points SAT Total
4. Coaching school claims of 100 points in total:[5]	62 points SAT Math 38 points SAT Verbal 100 points SAT Total

The last variant of assumptions on coaching effects on the SAT total have been partitioned into SAT-Math and SAT-Verbal using the ratios implied by the Messick-Jungeblut estimates. The claims by the coaching schools should be considered as total score gains over time including the effects of growth and motivation, unlike the previous three variants of assumptions. If these gains were coaching effects, according to the Messick-Jungeblut regression equations, we would expect that the students received 791 hours of coaching for SAT-Math and 875 hours of coaching for SAT-Verbal. It is clear that, even if such gains were possible, they

[5]Claims by Princeton Review (Houston Post, 1985; The Record, 1985).

cannot be attributed to Type 1 or 3 effects (test familiarization or trick learning) only.

The first set of coaching effect assumptions comes very close to the coaching effects reported by Whitla (1988), Smyth (1989), and by Messick and Jungeblut (1981) in their meta-analysis excluding the long term preparatory schools. These coaching effects are also almost identical to the estimates obtained when the hours of coaching reported by the members of the Powers sample are used to predict coaching effects using Messick-Jungeblut regression equations that relate hours to coaching effects (11.3 Math and 7.8 Verbal points). The FTC study's coaching effect estimates are approximated by the third variant of coaching effect assumptions.

There is no available research that documents a trend in the coaching effects over the last ten years. The College Board continues its efforts to develop tests that are robust to test taking tricks, and the coaching programs are searching for more effective coaching methods. Additionally, most recent research on coaching effects (Smyth, 1989; Whitla, 1988; computations based on Powers 1987-88 data in Table 1) yields coaching effect estimates that are in parity with earlier studies summarized in the Messick-Jungeblut (1981) and DerSimonian-Laird (1983) meta-analyses. Changes related to coaching during the last decade are more likely to be in the proportions of students coached than in coaching effectiveness.

There is little quantitative evidence on the magnitude of the correlation between Math and Verbal coaching effects. The number of hours spent on SAT-Verbal and SAT-Math coaching are almost perfectly correlated (.97), according to Powers' 1986-1987 data. Since in Section C and especially in Figure 1.c it is established that the magnitude of the correlation of SAT-Math and Verbal coaching effects is of little consequence, the correlation coefficient of .97 is taken as a proxy to the correlation between math and verbal coaching effects. Appendix A documents that changing assumptions on this correlation have little impact on the estimated validity of the SAT.

RESULTS

Four colleges from the College Board's Validity Study Service database were chosen for simulations. Table 2 gives the characteristics of the 1985 freshmen at these colleges. College A represents a highly selective college with mean SAT scores in 1985 of 698 (S.D. = 66) for Math and 666 (S.D. = 74) for Verbal. College B represents colleges that are moderately selective with average scores of 566 (S.D. = 85) for SAT-Math and 520 (S.D. = 86) SAT-Verbal respectively. College C is a less selective college with respect to the SAT scores with 438 (S.D. = 85) and 397 (S.D. = 75) points on the average for SAT-Math and SAT-Verbal, respectively. In addition to these three colleges, simulations for College D were run. This is a religious affiliated college where SAT scores are not likely to be the primary criterion for admission Therefore, it is expected that few students at College D would have sought coaching before entry. The results of the simulations for this college are the least likely to be contaminated by coaching effects that are already present in the database. The mean SAT scores for College D are 467 (S.D. = 90) for SAT-Math and 426 (S.D. = 86) for SAT-Verbal.

Simulations for each combination of assumptions have been repeated 5 times in order to account for possible random fluctuations. The effect of random fluctuations are negligible, the larger the sample size. Hence, the standard deviation of the estimates between repetitions are fairly low especially for Colleges A and B, which have larger numbers of freshman students. Since College D is the smallest college in this study, and since the results of the simulations for this college fluctuated moderately, the simulations for College D have been repeated 10 times.

For each college, three indicators of the predictive validity of the SAT are presented for each variant of the simulations. These indicators are the multiple correlation coefficient, multiple partial correlation coefficient, and the increment in the r^2 due to the SAT scores. The multiple correlation coefficient measures the correlation of an optimal linear combination of SAT-Math and SAT-Verbal with freshman GPA. It is derived from a regression of freshman GPA on SAT-Math and SAT-Verbal. The multiple partial correlation coefficient measures the correlation of SAT scores with freshman GPA, controlling for the association of the SAT-scores and the freshman year GPA with high school GPA. Let $r_{SAT,FGPA.HGPA}$ be the partial multiple correlation coefficient of SAT scores and the freshman year GPA, given the high school GPA. This coefficient is defined as:

$$r_{SAT,FGPA.HGPA} = \frac{r_{SAT,FGPA} - r_{SAT,HGPA} \cdot r_{FGPA,HGPA}}{\sqrt{1-r^2_{SAT,HGPA}}\ \sqrt{1-r^2_{FGPA,HGPA}}}$$

where subscripts "SAT,HGPA" and "SAT,FGPA" denote multiple correlation coefficients of SAT scores with high school GPA and, SAT scores with freshman GPA, respectively; and the subscript "FGPA,HGPA" denote the simple correlation coefficient of high school GPA with freshman GPA.

The increment in r^2 due to the SAT scores gives the change in the proportion of the variance of freshman GPA accounted for by a multiple regression equation where the SAT scores and the high school GPA are the predictor variables, as compared to the proportion of the variance in freshman GPA accounted for by a simple regression equation including only the high school GPA.

A. College A

College A 1985 freshman data contain records of 996 students whose background characteristics, SAT scores, high school GPAs and college freshman GPAs are known. An average freshman at this college has a college graduate father and a family income that is slightly under $40,000 per year. The students' mean high school GPA is 3.85 and mean freshman GPA is 3.11. The observed multiple correlation coefficient of SAT-Math and SAT-Verbal with freshman GPA is .42. The results of the simulations are presented in Tables 3.a, 3.b, and 3.c.

Since College A is a highly selective college with students who have high socioeconomic status and high levels of academic achievement, as many as a quarter of the observed SAT scores in this college may include prior coaching effects. Hence, the addition of the simulated coaching effects for College A are expected to have the largest negative effects on the measures of validity of the SAT among the four colleges chosen for this study. The reasons for this are twofold. First, the simulated coaching effects in College A are probably added to already existing coaching effects, resulting in total coaching effects that are probably higher than in any other college in this study. The declines in the measures of validity of the SAT are expected to increase rapidly as the coaching effects increase (Figure 1.d). Second, the standard deviations of the SAT scores are smallest in College A and, everything else being equal, the declines in the SAT validity indicators are expected to be highest where standard deviations are lowest (Figure 1.e).

When the assumed effectiveness of coaching is low, the predictive validity of the SAT for College A is affected only slightly, even when the probabilities of being coached are as high as 25% (Table 3.a). Assuming that in College A, the probability of being coached for an average student is .25 (close to Whitla's estimate of .22 in a similar selective college), and the coaching effects total to 37 points on the average (Messick-Jungeblut mean estimates), the multiple correlation coefficient of SAT scores with first year GPA declines by 1.6% as compared to the case when no coaching effects are simulated. We have very little information about the proportions coached in selective colleges 10 years ago. Under the assumption that the growth of the proportions coached in selective colleges have been parallel to the growth in a random sample of SAT takers, we could assume that the proportions coached increased by about 1% a year during the last 10 years. Therefore, 10 years ago, we can estimate that about 15% of College A students were coached. If we also assume that mean coaching effects have remained constant during this period, at the level estimated by Messick and Jungeblut, and that these effects represent only Type 1 and Type 3 gains, then we could conclude that the multiple correlation coefficient of the SAT scores and the freshman year GPA for College A would decline by about 0.5% over the last 10 years due to the coaching effects (Table 3.a, second row, columns 2 and 3). The comparable percentage of decline in the multiple partial correlation would be 0.3% (same cells, Table 3.b).

Table 3.c provides the increment in the squared multiple correlation under various simulated conditions when the SAT scores are introduced into the validity equation that initially contained only high school GPA. The percentages of decline in this metric at higher levels of coaching effects are larger than the comparable entries in the previous two tables. For example, the decline in the increment in r^2 compared to the observed data (assumed to represent no coaching) is 4.2% when the coaching effects are assumed to be one standard deviation unit above the Messick-Jungeblut mean estimates, and when 25% of students are assumed to be coached. Under the same assumptions, the decline in the multiple partial correlation coefficient is 2.1%.

If the coaching school claims of 100 points of score gains in the SAT-total scores were accurate, and if all these score gains could be attributed to test familiarization and test taking strategy or trick learning (i.e., assuming no training or growth effects in these estimates), then in a highly selective college like College A, with as much as a quarter of the students having received coaching, we may expect to find the multiple correlation of the SAT with first year GPA to be lower by 7.6%, as compared to what the correlation would have been if no coaching effects were simulated. Under the same circumstances the partial correlation would have been reduced by about 6.1%.

Morgan (1989) reports a decline in the multiple correlation coefficient of the SAT scores with freshman GPA of 8.2% in selective colleges (coefficient corrected for restriction of range). If coaching effects were to account for this decline in College A, one would have to assume that the incidence of coaching has increased by at least five fold and the coaching effects have increased by 100 points over the past decade, excluding the increases in coaching effects due to genuine skills training. To the extent that these assumptions are unsupported by evidence, coaching effects could not have accounted for any more than a minor proportion of the decline in the validity of the SAT in highly selective colleges.

B. College B

Data from 1,346 freshman students of College B in 1985 are used. These students, on the average, have lower socioeconomic status indicators than College A students. Mean paternal education is some college education, although an average student is not expected to have a father who has completed a 4-year college. Mean parental income is close to that estimated for College A students and is slightly over $35,000 per year. Entering freshmen of College B have achieved a mean high school GPA of 3.44, and at the end of their freshman year, their mean GPA was 2.84. The observed multiple correlation between the SAT scores and freshman GPA is .42, equal to the correlation estimated for College A.

Tables 4.a, 4.b, and 4.c present the results of the simulations on College B data. There are no data documenting the proportion of students who received coaching in colleges of this type. The characteristics of the students of College B are similar to those of an average SAT taker. Thus, it is assumed that national averages of incidence of coaching would be roughly appropriate. If we assume that the probabilities of being coached for an average College B student is about .15 and the mean effectiveness of coaching is at mean Messick-Jungeblut estimates, we see that the multiple correlation between SAT scores and freshman GPA declines by about 1.2% compared to the (assumed) no coaching baseline. If 10 years ago about 5% of College B students were coached at the same levels of effectiveness, a 0.8% decline in the multiple correlation of the SAT scores and the freshman year GPA could be expected due to changes in proportions coached (Table 4.a, cells 2,1 and 2,2). If the validity of the SAT is measured by the multiple partial correlation coefficient, in the last 10 years, one would expect a 0.5% decline due to changes in the levels of coaching. It is probably unrealistic to assume that higher proportions of College B students would have been coached. If, however, our coaching effect estimates are low, and coaching programs lead to effects that are 1 standard deviation higher than Messick-Jungeblut mean estimates, then a decline of 1.4% in the multiple correlation (from .4196 to .4138) and 1.0% in multiple partial correlation (from .3432 to .3397) in the last 10 years can be expected due to the increases in proportions coached in College B.

According to Morgan (1989), in large colleges like College B (colleges with more than 750 freshmen) the (corrected) multiple correlation of the SAT scores and the freshman GPA declined by 12.2% during 1976-85. Even if one assumed that coaching effects increased from null to a total of 100 points, and even if the proportions coached increased from zero to 25%, there is no evidence that in colleges like College B, trends in coaching could have accounted for the observed decline in validity.

C. College C

College C has a relatively small number of students and data from 386 students could be used for simulations. The mean level of education of students' fathers is slightly under some college education, and their mean parental income is slightly over $25,000 per year. The mean high school GPA of the 1985 entering freshmen is 2.93, much below that of College A and College B freshmen. Mean freshman GPA of the students is 2.37. The multiple correlation between the SAT scores and freshman GPA is .32, lower than the corresponding figures for the previous two colleges.

Tables 5.a to 5.c show the sensitivity of different measures of the validity of SAT for College C to varying simulated levels of coaching effects and proportions coached. The changes in the estimated multiple correlation coefficients of the SAT scores are smaller than those found for College A and College B. There could be two reasons for smaller changes in the multiple correlations in College C: First, the standard deviation of the observed ("uncoached") SAT scores compared to their means are larger in College C than in Colleges A or B, and, everything else being equal, one would expect smaller declines in the correlation coefficient, the higher the standard deviation (see Methods Section C and Figure 1.e). Second, the data from College C are unlikely to be contaminated by already existing coaching effects since the characteristics of the students are typical of low levels of participation in coaching programs. Hence the simulations for this college probably do not result in the accumulation of actual and simulated coaching effects, unlike the case at College A.

A close inspection of the residuals of the observed validity regression equation shows that in College C, the freshman GPA of students who have highly educated fathers is higher than their expected freshman GPA given their high school and SAT performance. Since students who have higher levels of paternal education have higher probabilities of being coached, among the students with higher educated fathers the addition of coaching effects on the SAT scores results in an improved predictive power of the SAT scores for freshmen GPA (consistent with Type 1 gains). Although the same polarization of the residuals of the validity equation occurs, to a lesser extent, in College A and College B, the parental backgrounds of the students of these colleges are more homogenous, and the number of students with less educated parents is very small. Therefore, in Colleges A and B, the probabilities of being coached are more uniform, and the added coaching effects do not serve as a proxy to increased parental education.

It is extremely unlikely that high proportions of College C students would have received coaching. We do not have any data on the proportions of students entering less selective colleges who receive coaching. In the absence of other data, one might assume that the proportions coached in such colleges are close to or somewhat under national proportions. If 15% of College C students received coaching at effectiveness levels equal to Messick-Jungeblut mean estimates, the multiple correlation coefficient of SAT scores and freshman GPA would be lower by 0.9%, as compared to the observed baseline representing no coaching. The level of coaching that we assume present 10 years ago (5%) has almost no effect on the multiple correlation coefficient for the College C data at the same levels of effectiveness.

Table 5.b shows that the estimated multiple partial correlation coefficient increases when coaching effects are simulated. The multiple partial correlation coefficients has three components: the correlation of the SAT scores with freshman GPA; the correlation of the SAT scores with high school GPA; and the correlation of high school GPA with freshman GPA. Since the latter correlation is not altered by the simulations, the increase in the partial correlation that cannot be accounted for by a corresponding increase in the multiple correlation must be due to the changes in the correlation of the SAT scores with high school GPA (see Appendix B for a comprehensive explanation). A rule of thumb is, when the coaching effects are simulated, if the decrease in the multiple correlation of the SAT scores with high school GPA is 2-3 times higher than the decrease in the multiple correlation of the SAT scores with freshman GPA, one may expect an increase in the multiple partial correlation coefficient due to coaching effects (Appendix B). This will occur when the association of the background characteristics (that are related to coaching effects) with high school GPA is weaker than their association with freshman GPA. In other words, a change in the SAT

scores that is associated with socioeconomic status brings about a smaller decrement in the association of the SAT with freshman GPA than the decrement in the association of the SAT with high school GPA. The latter decrement may be high, when high school GPA is not associated with socioeconomic status.

The pattern of changes in the increment in r^2 due to the SAT scores (Table 5.c) is very similar to the pattern of changes in the multiple partial correlations (Table 5.b). These tables show that coaching could actually increase the validity of the SAT, when coaching effects result in an increase in the SAT scores of students who have higher levels of first year performance than expected, or who have received lower SAT scores than their aptitude levels (see Introduction, Type 1 score gains).

The observed 10 year decline in the validity of the SAT (multiple correlation coefficient corrected for restriction of range) in less selective colleges is estimated at 13.0% by Morgan (1989). The simulations indicate that such a decline could not be accounted for by trends in coaching in colleges like College C even if extreme changes in incidence and effectiveness of coaching were hypothesized.

D. College D

College D is a small religious-affiliated college, with under 300 students. Data from only 203 students could be used for the simulations. The average parental background and the academic performance of the students of this college are slightly higher than those observed in College C. Although there are somewhat higher proportions of students with more highly educated fathers and higher family incomes in College D than the proportions in College C, the mean levels of these indicators are approximately the same. The mean level of education of the fathers of the students is somewhat more than high school, and the mean family income is slightly over $25,000 per year. The high school GPA and the freshman GPA are 3.31 and 2.92, respectively. The observed multiple correlation between the SAT scores and freshman year GPA is .43, marginally higher than the coefficients for College A and College B, and much above the coefficient for College C.

Coaching effects are simulated for this college because it is reasonable to assume that a very small proportion of students in this college would have received coaching prior to taking the SAT. This assumption is based on the fact that, at College D, the mean SAT scores of the students are quite low with relatively high standard deviations (Table 2), which suggests that the admissions might have been based on other criteria than test performance. If this assumption is true, the observed SAT scores would be minimally contaminated by coaching. As with College C, the simulations for this college allow us to evaluate the impact of various levels of coaching on the predictive validity of uncoached SAT scores.

Since the number of students in College D is quite small, the fluctuations in the point estimates between simulations are quite large. This can be seen by comparing the magnitude of the standard deviations of the estimates in Tables 6.a, b, and c, to the standard deviations of the estimates in the previous tables (3 through 5). In order to obtain better point estimates of the measures of validity, the simulations for College D were repeated 10 times.

The magnitude of the declines in the estimated multiple correlation coefficients for College D are quite low, when the simulated coaching effects are at the level of Messick-Jungeblut mean estimates. If 15% of the students were coached with coaching effects estimated by Messick and Jungeblut, the decrement in the multiple correlation coefficient would be about 0.6% as compared to the assumed no coaching baseline. The multiple correlation coefficients at 5% and 15% probabilities of coaching are indistinguishable at this level of coaching effects. Only if coaching effects are assumed to be a standard deviation unit above mean Messick-Jungeblut estimates do the decrements in the multiple correlation coefficient exceed 1%.

The table of estimated multiple partial correlation coefficients (Table 6.b) reveal the phenomenon observed with College C data; i.e., the partial correlations of SAT scores with freshman GPA, controlled for high school GPA increase when moderately low coaching effects are assumed. When coaching effects are simulated, the decline in the correlation of high school GPA with the SAT scores is larger than the decline in the correlation of freshman GPA with the SAT scores. The association of the student background characteristics that determine the probabilities of being coached with the first year average accounts for this relative robustness of the correlation of freshman GPA with the SAT scores when coaching effects are simulated. Similar to College C, the students' background characteristics have substantial variance in College D. Therefore, the probabilities of being coached are well differentiated by these characteristics, and the simulated coaching points added to the SAT scores serve as a proxy to these characteristics. As the coaching effects increase, changes in the estimated partial correlation coefficients due to coaching are of the expected direction but lower than those in Colleges A and B. The increments in r^2 due to SAT scores display the same pattern of change at different levels of coaching, as does the multiple partial correlation coefficient.

DISCUSSION

The results of the research presented in this report provide a framework on which possible negative effects of coaching on the predictive validity of the SAT can be quantitatively evaluated in the context of different colleges. Not all types of score gains due to coaching affect the validity of the SAT negatively. In the simulations that are presented in this study, coaching effects are defined as score gains due to test familiarization and trick learning. No score gains that lead to genuine improvement of math or verbal skills were simulated. These latter score gains would be expected to lead to an increase in the freshman year performance as well.

Two factors related to coaching are very important in determining the magnitude of the decrement in the validity of the SAT as compared to its validity if no students were coached. These are the proportions of students who are coached and, more importantly, the effectiveness of coaching. Although the measurement of the proportion of students who receive coaching is relatively straight-forward, the coaching effects are very difficult to measure. Often, what is reported as the coaching effect is the total score gain that includes test familiarization effects (Type 1 gains) and training effects (Type 2 gains) as well as the effects of strategy and trick learning (Type 3 gains), and in some cases, maturation and self-selection effects. Most of the public opinion on the effectiveness of coaching is based on the advertising claims of the commercial coaching schools, and the news media articles that are based on these claims. According to what we know from the research on coaching effects, the coaching school claims represent highly overstated coaching effects. The FTC study is now a decade old, and most of the other studies that attempt to estimate coaching effects date even further back. For a better understanding of this issue, a well designed experimental study on the effectiveness of coaching programs under realistic conditions of motivation for taking the SAT, is sorely needed.

Besides the magnitude of the coaching effects and the proportions coached, several factors contribute to the magnitude of the effect of coaching on the validity of the SAT. One factor is the variance of the "uncoached," or underlying SAT scores. The higher the variance (or dispersion) of the SAT, everything else being equal, the less the impact of coaching on the measures of validity.

The mean level of the "uncoached" SAT scores and their variance are often associated with the background characteristics of the students. Highly selective colleges that have high mean SAT scores with low variances (low coefficients of variation) often have students who come from uniformly high socioeconomic status backgrounds. Therefore, these students' coaching behaviors are not differentiated by their socioeconomic status, but rather by factors that are hard to measure (e.g., motivation and accessibility of coaching schools). When the associations between coaching behavior and freshman GPA with background characteristics are low, the decrements in the validity of the SAT are expected to be larger than the case when the coaching effects proximate the effects of higher socioeconomic status.

The validity indicators of the SAT in highly selective colleges could be strongly affected by coaching because of (1) the small dispersion of the underlying SAT scores; (2) the high proportions who are likely to have been coached; and, (3) the homogeneity of the student characteristics that are associated both with coaching behavior and with freshman GPA, hence the lower likelihood that these associations will dampen the coaching related decrement in the validity measures of the SAT.

The validity of the SAT in less selective colleges is not likely to be strongly affected by coaching because of (1) the large dispersion of the underlying SAT scores; (2) the low proportions who are likely to have been coached; and, (3) the possibility that increases in SAT scores due to coaching would act as a proxy for background characteristics (e.g. father's education and income) that are also associated with freshman year performance.

These conclusions show that the validity indicators of the SAT in highly selective colleges are more contaminated by coaching effects than in less selective colleges. These conclusions, however, do not provide us with the means to assess the impact of coaching on the <u>changes</u> in the validity measures of the SAT in the last 10 years. It is not reasonable to assume that coaching was non-existent 10 years ago and the validity of the SAT in highly selective colleges was probably more contaminated with coaching effects than in less selective schools even a decade ago. In order to be able to infer the changes in the predictive validity of the SAT that can be attributed to the changes in the coaching effects in the last 10 years, one needs further information on the trends in the proportions of students coached in different types of colleges and on the trends in the magnitude of the coaching effects. Such an analysis was presented in this report.

Morgan (1989) presents evidence that in the past decade, there were smaller declines in the predictive validity of the SAT in selective private colleges than in less selective, public and large colleges. We could infer the last 10 years' changes in the SAT validity that are attributable to coaching, under the assumption that the coaching effects have remained constant, and that proportions coached have been increasing by about 1% a year. Research on coaching effects during the past 15 years indicates that coaching effects probably have not increased. Under these assumptions, the multiple correlation coefficient of the SAT scores with freshman GPA could be expected to decrease by less than 1% as a result of changes in students' coaching behavior during the last 10 years. Therefore, the changes in the levels of coaching does not seem to account for the observed decline in the validity of the SAT over the past 10 years (7.8%), as reported by Morgan (1989). According to the results of the simulations, trends in coaching might have resulted in a decline in the multiple correlation coefficient of the SAT scores with freshman GPA by 0.6% in highly selective colleges (College A)[6]; and by 0.8% in

[6]Assuming the Messick-Jungeblut mean estimates of coaching effects and an increase in the incidence of coaching from 15% to 25%.

moderately and less selective colleges (Colleges B and C)[7], during the past decade.

In the Morgan (1989) study, estimates including the colleges of the Validity Study Service database that reported high school GPAs of their students showed that the 10-year decline in the multiple correlation of the SAT scores with freshman GPA corrected for restriction of range was 7.8% (from .51 to .47). The simulations presented here indicate clearly that only unrealistically strong increases in the effectiveness and incidence of coaching could have accounted for this trend. Only if coaching incidence had increased from 5% to 25% and coaching effects had increased from the Messick-Jungeblut mean estimates to 100 points in total over the past decade, would the effects of the trends in coaching approach the magnitude of the decline in the validity of the SAT reported by Morgan. However, research on coaching effects indicate that 100 point score increases could only be obtained by long-term instructional programs that improve cognitive skills which, in turn, improve college performance. Score gains that affect college performance have not been modelled in the simulations presented in this report and they are not expected to affect the validity of the SAT.

---

[7]Assuming the Messick-Jungeblut mean estimates of coaching effects and an increase in the incidence of coaching from 5% to 15%.

REFERENCES

Alderman, D.L. & D.E. Powers. (1980). The Effects of Special Preparation on SAT-Verbal Scores. American Educational Research Journal, 17, 239-251.

College Board. (1989). 1989-1990 Featured Publications and Software from the College Board. New York: College Entrance Examination Board.

DerSimonian, R. & N.M. Laird. (1983). Evaluating the Effect of Coaching on SAT Scores: A Meta-Analysis. Harvard Educational Review, 53, 1-15.

Donlon, T. F. (1984). The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board.

Federal Trade Commission, Bureau of Consumer Protection. (1979). Effects of Coaching on Standardized Admission Examinations: Revised Statistical Analysis of Data Gathered by Boston Regional Office of the Federal Trade Commission. Washington, D.C. : Federal Trade Commission.

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, NY: Research Division Office of the Director of Admissions and Registrar, U. S. Military Academy.

Messick, S. (1980). The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC. Princeton, N.J. : ETS.

Messick, s. (1982). Issues of Efectiveness and Equity in the Coaching Contoversy: 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.

Morgan, R. (1989). Analyses of the Predictive Validity of the SAT and High School Grades From 1976 to 1985, Research Report RR-89-7. Princeton, N.J.: ETS.

Powers, D.E. (1988). Preparing for the SAT: A Survey of Programs and Resources. RR-88-7. Princeton, N.J.: ETS.

Powers, D.E. & D.L. Alderman. (1979). The Use, Acceptance, and Impact of "Taking the SAT: A Test Familiarization Booklet". Research Report RR-79-3. Princeton, NJ: ETS.

Powers, D.E. & D.L. Alderman. (1983). Effects of Test Familiarization on SAT Performance. Journal of Educational Measurement, 20, 71-79.

Rock, D.A. (1980). Disentangling Coaching Effects and Differential Growth in the FTC Commercial Coaching Study. In: Messick, S. The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC. Princeton, N.J.: ETS, 123-135.

Rowe, J. (December 27, 1985). The SAT Cult: Getting Help to Score Highest Points is Name of the Game. The Houston Post.

Smyth, F.L. (1989). Commercial Coaching and SAT Scores. The Journal of College Admission, Spring, 2-7.

Stroud, T.W.F. (1980). Reanalysis of the Federal Trade Commission Study of Commercial Coaching for the SAT. ETS RR-80-10. Princeton, N.J.: ETS.

Verdon, J. (1985, August 15). SAT Service, Instructor Argue Copyright Case. The Record.

Whitla, D.K. (1988). Coaching: Does It Pay? Not for Harvard Students. The College Board Review, 148, 32-35.

Appendix A.

In order to evaluate the impact of the assumptions about the correlation of Math with Verbal coaching effects on the validity of SAT, simulations were run for College B. For these simulations, it was assumed that (i) the probability of being coached for an average student is .15, and (ii) the mean effectiveness of coaching is equal to the Messick-Jungeblut mean estimates. Only the correlation coefficient between the SAT-Math and SAT-Verbal coaching effects was varied from .5 to 1.0. The results are given in Table A.1.

Comparison of the point estimates of the validity indicators along the columns of Table A.1 clearly show that different assumptions on the magnitude of the correlation of SAT-Math with SAT-Verbal coaching effects have a negligible effect on the validity indicators. For example, the point estimate of the multiple correlation when the Math and Verbal effects are correlated by 0.5 is within a standard deviation unit of the same point estimate when the Math and Verbal effects are correlated by 1.0. The point estimates of the increment in r^2 remain approximately constant when the correlation between the Math and Verbal coaching effects is varied between 0.5 and 1.0.

Table A.1

Comparison of the Changes in Estimated Validity Statistics Under Varying Assumptions of Correlation Between SAT-M and SAT-V Coaching Effectiveness College B[1]

Correlation between V and M Coaching IncreasEffectiveness	Multiple Correlation Coefficient[2]	Mult/Partial Correlation Coefficient	in r^2
Correlation = 0.5			
Mean	.4180	.3424	.0971
Standard Deviation	(.0022)	(.0028)	(.0016)
% Decline	1.1	0.7	1.4
Correlation = 0.6			
Mean	.4179	.3424	.0971
Standard Deviation	(.0021)	(.0027)	(.0015)
% Decline	1.1	0.7	1.5
Correlation = 0.7			
Mean	.4179	.3423	.0971
Standard Deviation	(.0020)	(.0026)	(.0015)
% Decline	1.2	0.7	1.5
Correlation = 0.8			
Mean	.4178	.3423	.0971
Standard Deviation	(.0019)	(.0024)	(.0014)
% Decline	1.2	0.8	1.5
Correlation = 0.9			
Mean	.4178	.3423	.0971
Standard Deviation	(.0018)	(.0022)	(.0012)
% Decline	1.2	0.8	1.5
Correlation = 1.0			
Mean	.4179	.3424	.0971
Standard Deviation	(.0017)	(.0017)	(.0010)
% Decline	1.1	0.7	1.4

[1]Assumptions underlying these simulations are that coaching effectiveness is equal to the mean Messick-Jungeblut estimates, with minimum gains at -6.6 and -14.1 in SAT-V and SAT-M respectively; the probability of being coached for an average student is 0.15.

[2]Observed multiple correlation coefficient is .4227, multiple partial correlation coefficient is .3449, increase in r^2 is .0985.

Appendix B.

The amount of decrement in the multiple partial correlation coefficients due to coaching effects depend on not only the decrement in the multiple correlation coefficient of the SAT scores and the freshman year GPA, but also on the decline in the multiple correlation coefficient of the SAT scores and the high school GPA. Let $r_{SAT,FGPA}$ be the multiple correlation coefficient of the SAT scores and the freshman GPA; $r_{SAT,HGPA}$ be the multiple correlation coefficient of the SAT scores and the high school GPA; and $r_{FGPA,HGPA}$ be the correlation coefficient of freshman GPA and high school GPA. Let r^* denote the correlation coefficients estimated after coaching effects have been simulated. The ratio of the multiple partial correlation coefficient with simulated coaching effects to the same coefficient with no simulated coaching effects can be written as:

$$\frac{r^*_{SAT,FGPA.HGPA}}{r_{SAT,FGPA.HGPA}} = \frac{(r^*_{SAT,FGPA} - r^*_{SAT,HGPA} \cdot r_{FGPA,HGPA})\,(\sqrt{1 - r^2_{SAT,HGPA}})}{(r_{SAT,FGPA} - r_{SAT,HGPA} \cdot r_{FGPA,HGPA})\,(\sqrt{1 - r^{*2}_{SAT,HGPA}})}.$$

Since the last terms in the numerator and the denominator are typically very close to unity, the above ratio is approximately equal to:

$$\frac{(r^*_{SAT,FGPA} - r^*_{SAT,HGPA} \cdot r_{FGPA,HGPA})}{(r_{SAT,FGPA} - r_{SAT,HGPA} \cdot r_{FGPA,HGPA})}.$$

Therefore, if the term $r_{FGPA,HGPA}(r_{SAT.HGPA} - r^*_{SAT.HGPA})$ is greater than the term $(r_{SAT,FGPA} - r^*_{SAT,FGPA})$, the multiple partial correlation coefficient of SAT scores and high school GPA will be higher when the coaching effects are present than when they are absent. Since $r_{FGPA,HGPA}$ typically range between .3 and .4, when coaching effects are simulated, if the decrement in the multiple correlation of the SAT scores and the high school GPA are approximately 2-3 times higher than the decrement in the multiple correlation of the SAT scores and the freshman GPA, the multiple partial correlation will increase even when the coaching effects lead to a decrement in all the multiple correlations concerned. A relatively small decrease in the term $r^*_{SAT,FGPA}$ could occur, if coaching effects are associated with background characteristics that are more closely related to freshman year GPA than the high school GPA. For example, if father's education (that is associated with being coached) is more highly correlated with freshman year GPA than with high school GPA, the decrease in the correlation between the coached SAT scores and the freshman GPA associated with coaching will be smaller than the decrease in the correlation between the coached SAT scores and the high school GPA.

Table 1

Comparison of Coached and UnCoached Students of the Powers Study

Characteristic	UnCoached	Coached
Percentage of female students	54	60
Percentage of Black or Hispanic students	10	11
Percentage of students with fathers who have at least college education	45	62*
Mean income of the student's families	$32,515	$38,590*
Mean self-reported GPA	3.25	3.14+
Mean self-reported grade in English	3.25	3.21
Mean self-reported grade in math	3.07	2.97
Mean number of times taking SAT	1.59	1.97*
Mean SAT-V score at the 1st administration	443	426*
Mean SAT-M score at the 1st administration	499	488
Mean SAT-V score at the last administration	456	448
Mean SAT-M score at the last administration	511	516
Difference between first and last SAT-V scores[1]	27	32
Difference between first and last SAT-M scores[1]	25	46
Total number of students	1067	176

*Difference between coached and uncoached students is significant at $p < .05$.

+Difference between coached and uncoached students is significant at $p < .10$.

[1]Computed only for 51% of 1,243 students of the sample, who took the SAT at least two times.

Table 2
Characteristics of the Students in the four Schools:
Means and Standard Deviations

Characteristics	School A	School B	School C	School D
High School GPA	3.85 (.25)	3.44 (.43)	2.93 (.50)	3.31 (.49)
SAT-Math	698 (66)	566 (85)	438 (85)	467 (90)
SAT-Verbal	666 (74)	520 (86)	397 (75)	426 (86)
Father's Education[1]	4.4 (1.1)	3.8 (1.3)	2.9 (1.2)	2.8 (1.3)
Family Income[2]	7.7 (2.8)	7.2 (2.9)	5.2 (2.5)	5.1 (2.9)
Freshman Year GPA	3.11 (.50)	2.84 (.61)	2.37 (.68)	2.92 (.63)
Number of Students	996	1346	386	203

[1] Scale is defined as 1: Under high school; 2: High school graduate; 3: Some college education; 4: College completed (BA or BS); 5: Some graduate education.

[2] Scale is defined as 1: <$10,000; 2: $10-15,000; 3: $15-20,000; 4: $20-25,000; 5: $25-30,000; 6: $30-35,000; 7: $35-40,000; 8: $40-50,000; 9: $50-60,000; 10: $60-70,000; 11: >$70,000.

Table 3.a

Multiple Correlation Coefficients -- College A[1]

	Probability of Being Coached for an Average Student[2]		
Coaching Effects	5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.4222	.4215	.4208
Standard Deviation	(.0016)	(.0025)	(.0023)
% Decline	0.2	0.3	0.5
Messick-Jungeblut Estimates			
Mean	.4208	.4184	.4162
Standard Deviation	(.0029)	(.0045)	(.0036)
% Decline	0.5	1.1	1.6
1 SD Above Messick-Jungeblut Estimates			
Mean	.4189	.4144	.4101
Standard Deviation	(.0041)	(.0062)	(.0046)
% Decline	0.9	2.0	3.0
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.4133	.4016	.3909
Standard Deviation	(.0062)	(.0097)	(.0064)
% Decline	2.3	5.0	7.6

[1] Observed multiple correlation coefficient is .4228.

[2] Average proportions simulated for this college are 5.5%, 16.5% and 27.9%.

Table 3.b

Multiple Partial Correlations -- College A[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.3801	.3803	.3800
Standard Deviation	(.0015)	(.0024)	(.0025)
% Decline	0.1	0.1	0.2
Messick-Jungeblut Estimates			
Mean	.3790	.3782	.3769
Standard Deviation	(.0026)	(.0042)	(.0040)
% Decline	0.4	0.6	1.0
1 SD Above Messsick-Jungeblut Estimates			
Mean	.3774	.3753	.3725
Standard Deviation	(.0037)	(.0058)	(.0052)
% Decline	0.8	1.4	2.1
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.3725	.3655	.3574
Standard Deviation	(.0055)	(.0092)	(.0071)
% Decline	2.1	4.0	6.1

[1] Observed multiple partial correlation coefficient is .3806.

[2] Average proportions simulated for this college are 5.5%, 16.5% and 27.9%.

Table 3.c

Increment in r^2 -- College A[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.1311	.1312	.1310
Standard Deviation	(.0010)	(.0017)	(.0017)
% Decline	0.2	0.2	0.3
Messick-Jungeblut Estimates			
Mean	.1303	.1297	.1289
Standard Deviation	(.0018)	(.0029)	(.0027)
% Decline	0.8	1.3	1.9
1 SD Above Messick-Jungeblut Estimates			
Mean	.1292	.1278	.1259
Standard Deviation	(.0025)	(.0039)	(.0035)
% Decline	1.7	2.8	4.2
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.1259	.1213	.1159
Standard Deviation	(.0037)	(.0061)	(.0046)
% Decline	4.2	7.7	11.8

[1] Observed increment in r^2 is .1314.

[2] Average proportions simulated for this college are 5.5%, 16.5% and 27.9%.

Table 4.a

Multiple Correlation Coefficients -- College B[1]

	Probability of Being Coached for an Average Student[2]		
Coaching Effects	5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.4221	.4208	.4212
Standard Deviation	(.0007)	(.0010)	(.0026)
% Decline	0.1	0.5	0.4
Messick-Jungeblut Estimates			
Mean	.4211	.4178	.4180
Standard Deviation	(.0014)	(.0018)	(.0047)
% Decline	0.4	1.2	1.1
1 SD Above Messick-Jungeblut Estimates			
Mean	.4196	.4138	.4133
Standard Deviation	(.0020)	(.0025)	(.0068)
% Decline	0.7	2.1	2.2
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.4133	.3989	.3952
Standard Deviation	(.0037)	(.0045)	(.0121)
% Decline	2.2	5.6	6.5

[1] Observed multiple correlation coefficient is .4227.

[2] Average proportions simulated for this college are 5.9%, 17.0% and 26.9%.

Table 4.b

Multiple Partial Correlations -- College B[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.3447	.3441	.3451
Standard Deviation	(.0005)	(.0011)	(.0025)
% Decline	0.0	0.2	-0.1
Messick-Jungeblut Estimates			
Mean	.3442	.3423	.3436
Standard Deviation	(.0010)	(.0020)	(.0046)
% Decline	0.2	0.7	0.4
1 SD Above Messick-Jungeblut Estimates			
Mean	.3432	.3397	.3408
Standard Deviation	(.0015)	(.0028)	(.0066)
% Decline	0.5	1.5	1.2
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.3386	.3291	.3284
Standard Deviation	(.0029)	(.0049)	(.0115)
% Decline	1.8	4.6	4.8

[1] Observed multiple partial correlation coefficient is .3449.

[2] Average proportions simulated for this college are 5.9%, 17.0% and 26.9%.

Table 4.c

Increment in r^2 -- College B[1]

Coaching Effects	Probability of Being Coached for an Average Student[2]		
	5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.0985	.0981	.0987
Standard Deviation	(.0003)	(.0006)	(.0014)
% Decline	0.1	0.5	-0.1
Messick-Jungeblut Estimates			
Mean	.0981	.0971	.0978
Standard Deviation	(.0006)	(.0011)	(.0026)
% Decline	0.4	1.5	0.7
1 SD Above Messick-Jungeblut Estimates			
Mean	.0976	.0956	.0963
Standard Deviation	(.0008)	(.0016)	(.0037)
% Decline	1.0	3.0	2.3
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.0950	.0897	.0894
Standard Deviation	(.0016)	(.0027)	(.0062)
% Decline	3.6	9.0	9.2

[1] Observed increment in r^2 is .0985.

[2] Average proportions simulated for this college are 5.9%, 17.0% and 26.9%.

Table 5.a

Multiple Correlation Coefficients -- College C[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.3249	.3237	.3249
Standard Deviation	(.0007)	(.0041)	(.0038)
% Decline	-0.0	0.3	-0.0
Messick-Jungeblut Estimates			
Mean	.3245	.3218	.3223
Standard Deviation	(.0014)	(.0075)	(.0068)
% Decline	0.1	0.9	0.8
1 SD Above Messick-Jungeblut Estimates			
Mean	.3236	.3188	.3178
Standard Deviation	(.0023)	(.0109)	(.0097)
% Decline	0.3	1.8	2.2
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.3196	.3061	.2993
Standard Deviation	(.0049)	(.0194)	(.0176)
% Decline	1.6	5.7	7.8

[1] Observed multiple correlation coefficient is .3247.

[2] Average proportions simulated for this college are 5.6%, 15.4% and 26.0%.

Table 5.b

Multiple Partial Correlations -- College C[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.2129	.2135	.2142
Standard Deviation	(.0015)	(.0046)	(.0053)
% Decline	-0.5	-0.7	-1.1
Messick-Jungeblut Estimates			
Mean	.2134	.2142	.2141
Standard Deviation	(.0028)	(.0084)	(.0095)
% Decline	-0.7	-1.1	-1.0
1 SD Above Messick-Jungeblut Estimates			
Mean	.2136	.2141	.2126
Standard Deviation	(.0042)	(.0121)	(.0136)
% Decline	-0.8	-1.0	-0.3
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.2127	.2104	.2041
Standard Deviation	(.0080)	(.0212)	(.0237)
% Decline	-0.4	0.7	3.7

[1] Observed multiple partial correlation coefficient is .2119.

[2] Average proportions simulated for this college are 5.6%, 15.4% and 26.0%.

Table 5.c

Increment in r^2 -- College C[1]

Coaching Effects	Probability of Being Coached for an Average Student[2]		
	5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.0368	.0371	.0373
Standard Deviation	(.0005)	(.0016)	(.0018)
% Decline	-0.9	-1.5	-2.2
Messick-Jungeblut Estimates			
Mean	.0370	.0373	.0373
Standard Deviation	(.0010)	(.0029)	(.0033)
% Decline	-1.4	-2.3	2.2
1 SD Above Messick-Jungeblut Estimates			
Mean	.0371	.0374	.0369
Standard Deviation	(.0014)	(.0041)	(.0046)
% Decline	-1.6	-2.4	-1.1
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.0368	.0363	.0343
Standard Deviation	(.0027)	(.0069)	(.0078)
% Decline	-1.0	0.5	6.0

[1] Observed increment in r^2 is .0365.

[2]Average proportions simulated for this college are 5.6%, 15.4% and 26.0%.

Table 6.a

Multiple Correlation Coefficients -- College D[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.4324	.4330	.4306
Standard Deviation	(.0032)	(.0048)	(.0041)
% Decline	0.3	0.1	0.7
Messick-Jungeblut Estimates			
Mean	.4307	.4309	.4257
Standard Deviation	(.0061)	(.0089)	(.0074)
% Decline	0.7	0.6	1.8
1 SD Above Messick-Jungeblut Estimates			
Mean	.4284	.4273	.4187
Standard Deviation	(.0090)	(.0129)	(.0107)
% Decline	1.2	1.5	3.4
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.4202	.4117	.3933
Standard Deviation	(.0167)	(.0226)	(.0185)
% Decline	3.1	5.1	9.3

[1] Observed multiple correlation coefficient is .4335.

[2] Average proportions simulated for this college are 6.1%, 16.9% and 27.0%.

Table 6.b

Multiple Partial Correlations -- College D[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.3853	.3880	.3878
Standard Deviation	(.0043)	(.0048)	(.0038)
% Decline	0.2	-0.5	-0.5
Messick-Jungeblut Estimates			
Mean	.3843	.3883	.3873
Standard Deviation	(.0081)	(.0090)	(.0069)
% Decline	0.4	-0.6	-0.4
1 SD Above Messick-Jungeblut Estimates			
Mean	.3827	.3873	.3851
Standard Deviation	(.0117)	(.0129)	(.0099)
% Decline	0.8	-0.4	0.2
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.3767	.3793	.3727
Standard Deviation	(.0211)	(.0224)	(.0167)
% Decline	2.4	1.7	3.4

[1] Observed multiple partial correlation coefficient is .3859.

[2] Average proportions simulated for this college are 6.1%, 16.9% and 27.0%.

Table 6.c

Increment in r^2 -- College D[1]

Coaching Effects	Probability of Being Coached for an Average Student[2] 5%	15%	25%
1 SD Below Messick-Jungeblut Estimates			
Mean	.1193	.1209	.1208
Standard Deviation	(.0027)	(.0030)	(.0024)
% Decline	0.3	-1.1	-1.0
Messick-Jungeblut Estimates			
Mean	.1187	.1212	.1205
Standard Deviation	(.0049)	(.0056)	(.0043)
% Decline	0.8	-1.3	-0.8
1 SD Above Messick-Jungeblut Estimates			
Mean	.1178	.1207	.1192
Standard Deviation	(.0071)	(.0081)	(.0061)
% Decline	1.6	-0.8	0.4
Score Gains			
Coaching School Claims of 100 Points Total			
Mean	.1144	.1159	.1118
Standard Deviation	(.0123)	(.0137)	(.0101)
% Decline	4.4	3.1	6.6

[1] Observed increment in r^2 is .1196.

[2] Average proportions simulated for this college are 6.1, 16.9% and 27.0%.

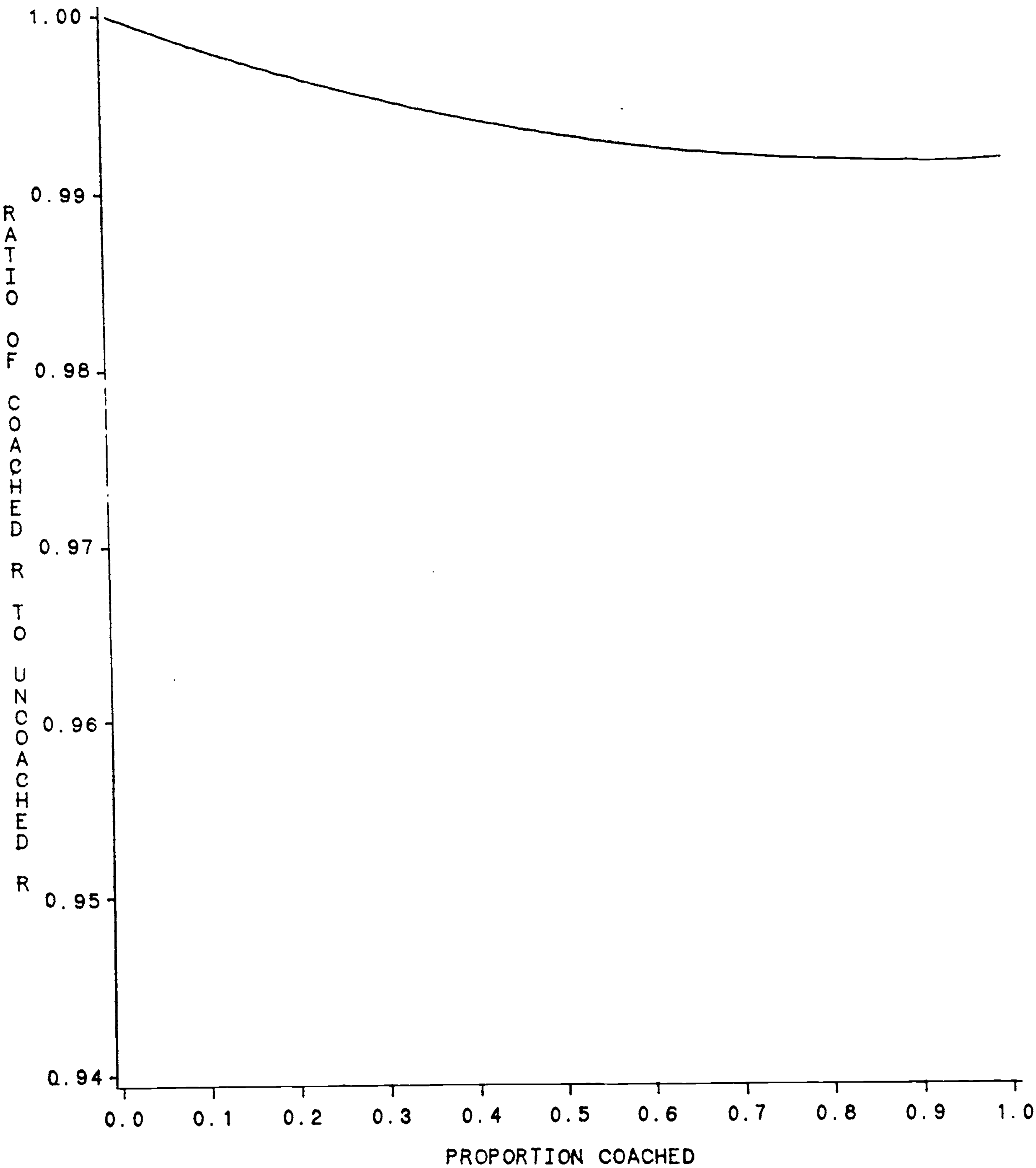
Figures 1.a to 1.e

Effects of Various Parameters Related to

Coaching and SAT-Scores on

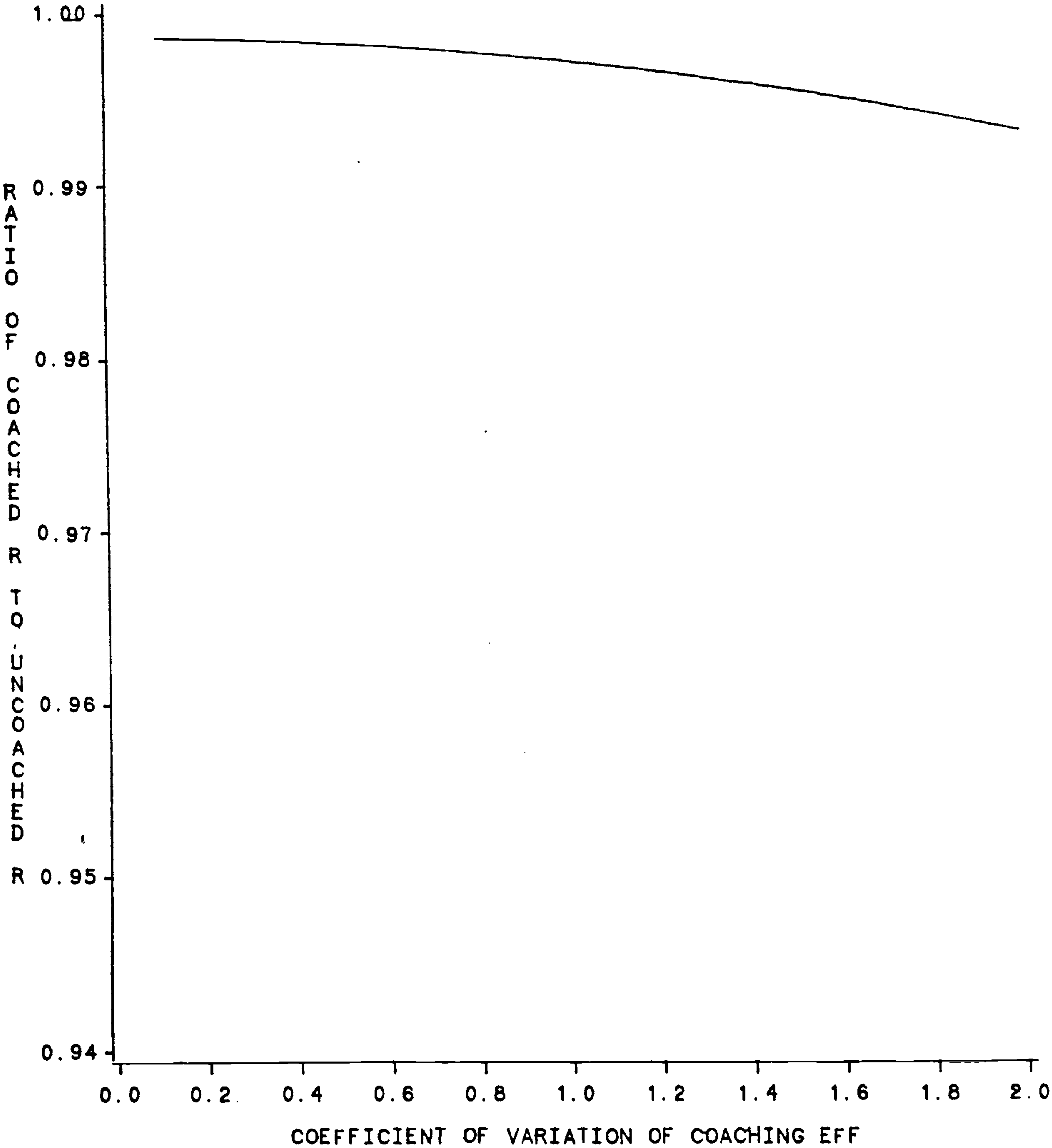
the Ratio of the Coached to Uncoached r

A. PROPORTION OF STUDENTS WHO RECEIVED COACHING



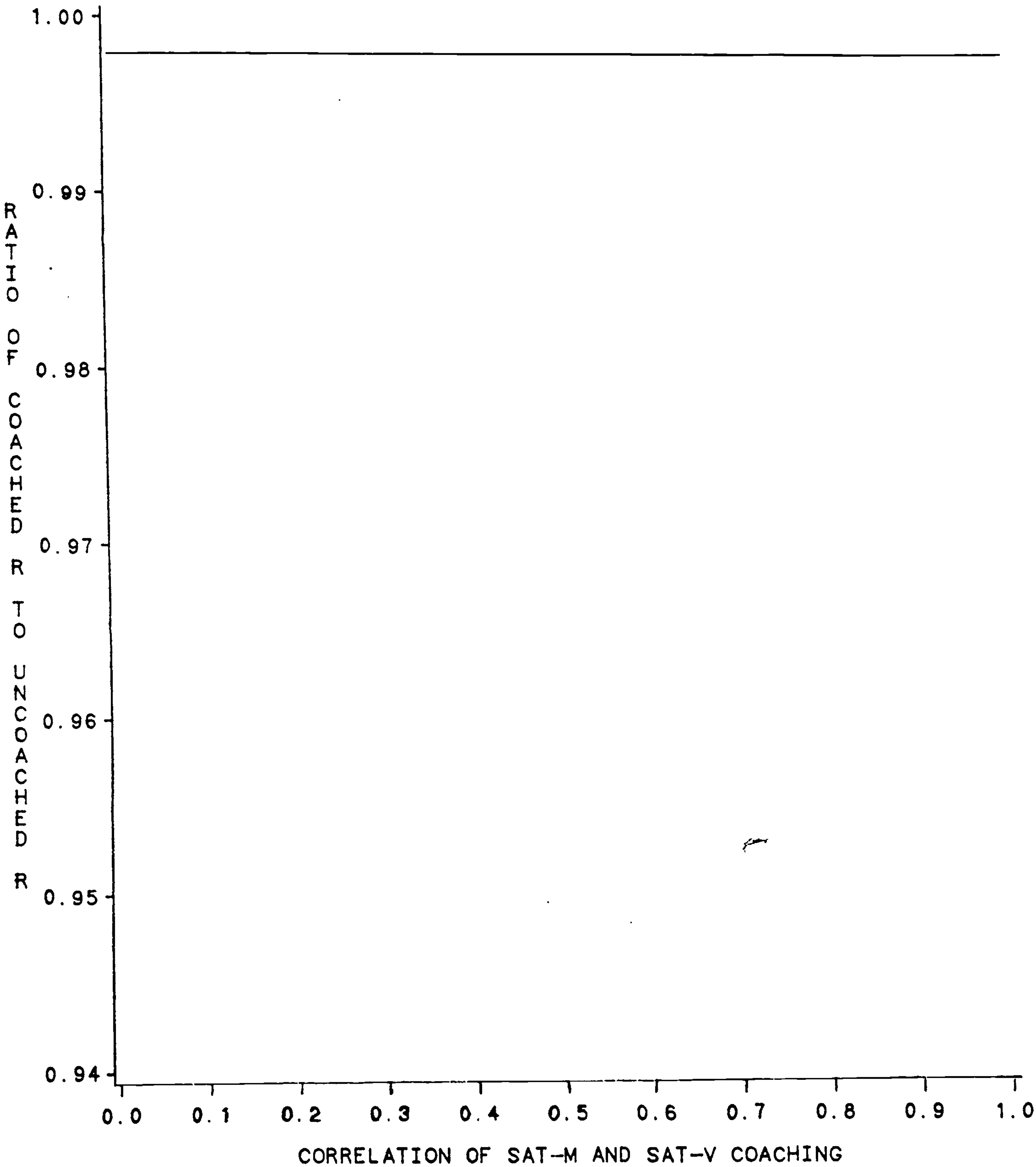
COACHING EFFECTS ARE 23 FOR SAT–M AND 14 FOR SAT–V
COEFFICIENT OF VARIATION OF COACHING EFFECTS ARE .954 FOR SAT–M AND .734 FOR SAT–V
CORRELATION OF SAT–M COACHING EFFECTS WITH SAT–V COACHING EFFECTS IS .97
STANDARD DEVIATION OF UNDERLYING (UNCOACHED) SAT–TOTAL IS 250

B. COEFFICIENT OF VARIATION OF COACHING EFFECTS



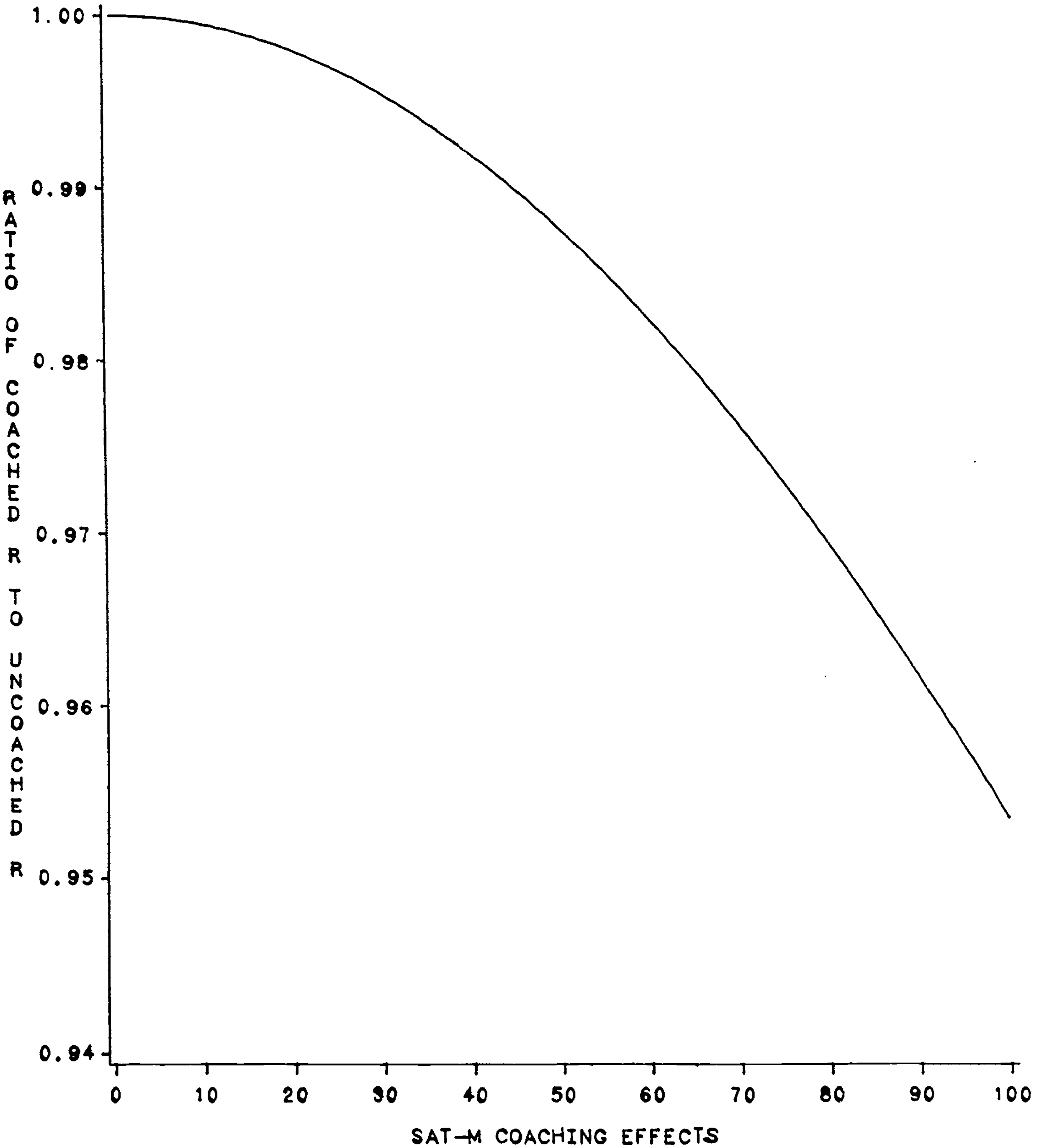
COACHING EFFECTS ARE 23 FOR SAT–M AND 14 FOR SAT–V
COEFFICIENT OF VARIATION OF SAT–V COACHING EFFECTS IS
(.734/.954) * COEFFICIENT OF VARIATION OF SAT–M COACHING EFFECTS
CORRELATION OF SAT–M COACHING EFFECTS WITH SAT–V COACHING EFFECTS IS .97
STANDARD DEVIATION OF UNDERLYING (UNCOACHED) SAT–TOTAL IS 250
PROPORTION OF COACHED STUDENTS IS 0.15

C. CORRELATION OF COACHING EFFECTS ON SAT–M AND SAT–V



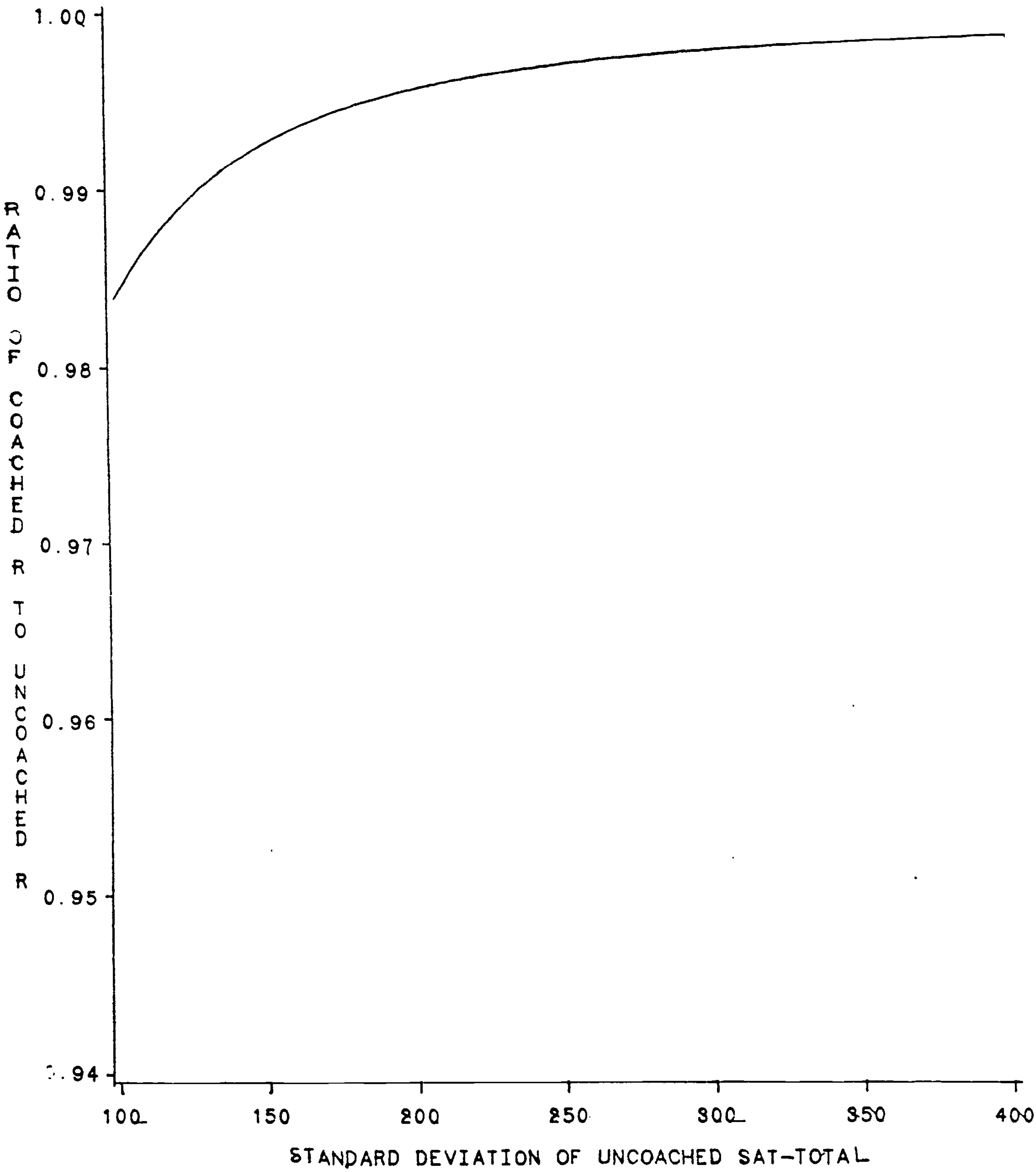
COACHING EFFECTS ARE 23 FOR SAT–M AND 14 FOR SAT–V
COEFFICIENTS OF VARIATION OF COACHING EFFECTS ARE .954 FOR SAT–M AND .734 FOR SAT–V
STANDARD DEVIATION OF UNDERLYING (UNCOACHED) SAT–TOTAL IS 250
PROPORTION OF COACHED STUDENTS IS 0.15

D. MEAN SAT—MATH AND SAT—VERBAL COACHING EFFECTS



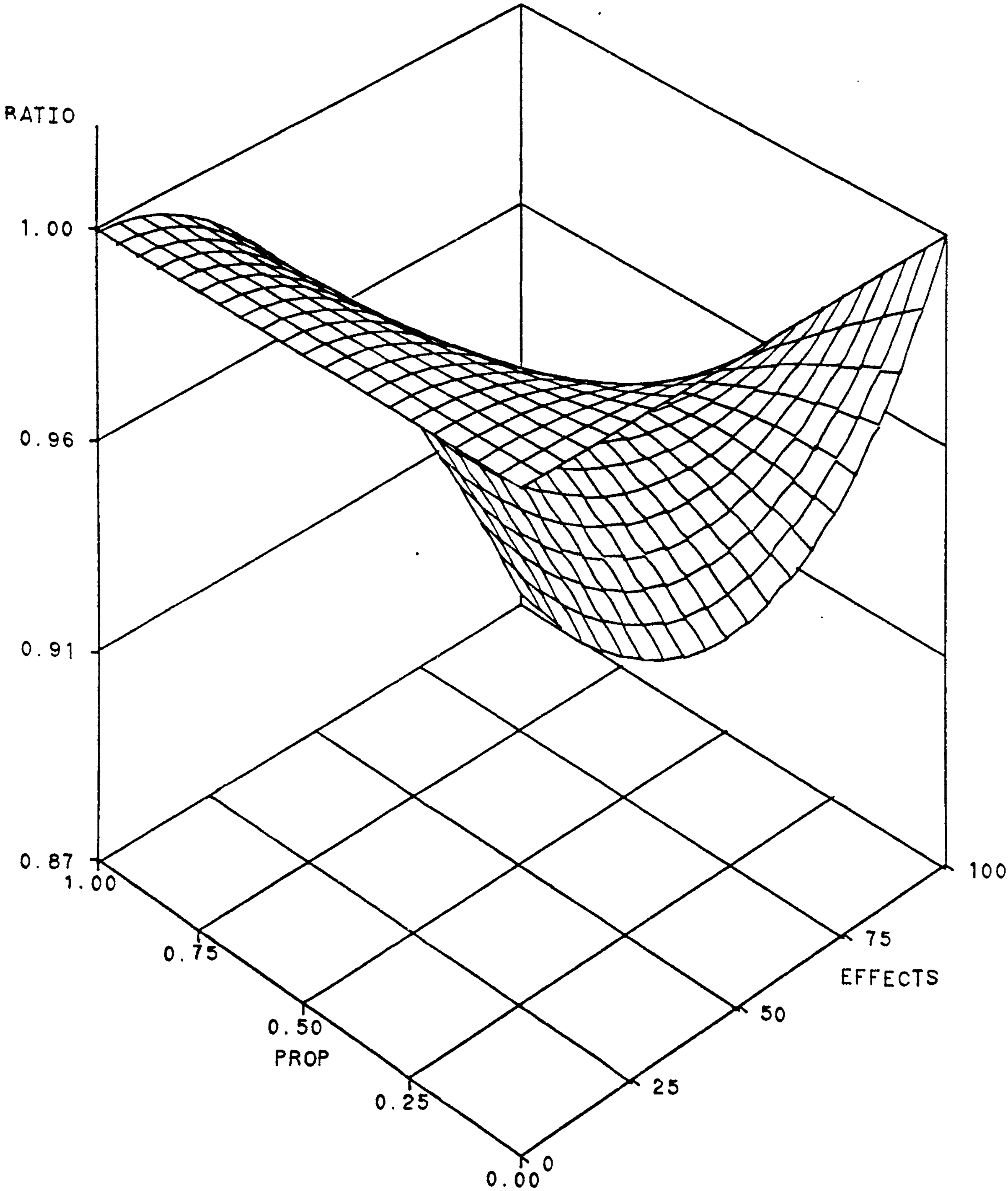
THE RATIO OF SAT—M TO SAT—V COACHING EFFECTS IS 62:38
COEFFICIENTS OF VARIATION OF COACHING EFFECTS ARE .954 FOR SAT—M AND .734 FOR SAT—V
STANDARD DEVIATION OF UNDERLYING (UNCOACHED) SAT—TOTAL IS 250
PROPORTION OF COACHED STUDENTS IS 0.15
CORRELATION OF SAT—M COACHING EFFECTS WITH SAT—V COACHING EFFECTS IS .97

E. STANDARD DEVIATION OF THE UNDERLYING (UNCOACHED) SAT-TOTAL



COACHING EFFECTS ARE 23 FOR SAT-M AND 14 FOR SAT-V
COEFFICIENTS OF VARIATION OF COACHING EFFECTS ARE .954 FOR SAT-M AND .734 FOR SAT-V
PROPORTION OF COACHED STUDENTS IS 0.15
CORRELATION OF SAT-M COACHING EFFECTS WITH SAT-V COACHING EFFECTS IS .97

Figure 2. Effects of the Proportion of Students Coached and the Magnitude of Coaching Effects on the Ratio of the Coached to Uncoached r



EFFECTS: SAT–MATH COACHING EFFECTS (SAT–V = SAT–M * .61)
PROP: PROPORTION OF COACHED STUDENTS
RATIO: RATIO OF THE COACHED TO UN–COACHED CORRELATION