

# Synthesis of Research on the Effects of Coaching for Aptitude and Admissions Tests

ROBERT L. BANGERT-DROWNS, JAMES A. KULIK, AND CHEN-LIN C. KULIK

Students are sometimes very critical of aptitude and admissions tests, which they say focus on trivia, ask tricky questions, and are not relevant to what is taught in schools. But such criticism is easy to dismiss. Who would expect students to be objective about tests that can limit their educational options?

Recently, however, several educational researchers have joined students in criticizing aptitude and admissions tests. Slack and Porter (1980) charged that scores on the Scholastic Aptitude Test, the most widely used of all such tests, can be influenced substantially by coaching programs and that the test is not very good at predicting college grades. Jencks and Crouse (1982) concluded that tests such as the SAT do not measure what we normally mean by "aptitude" any better than do conventional achievement tests. SAT scores appear to be just as dependent on home environment and school quality as are scores on conventional achievement tests. And SAT scores are not better than achievement test scores in predicting success in college and adult life.

These critics of aptitude testing have taken on one of the important rites in American education. Every year millions of students take such tests as the Scholastic Aptitude Test, Graduate Record Examination, and the Law School Admissions Test. And more than 50,000 students each year enroll in commercial coaching schools to prepare themselves for those tests. Many more students take coaching courses offered in their high schools. One survey estimated that one-third of the private and public schools in the Northeast offered

some sort of SAT preparation course (Alderman and Powers, 1980).

Teachers and school administrators are naturally interested in knowing more about the best ways of preparing students for tests. Questions loom up on all sides:

- Can performance on aptitude tests be influenced by coaching programs? Are aptitude tests less coachable than achievement tests?
- How much coaching is optimal?
- What sorts of programs offer the best preparation for the tests?
- In order to ensure equal access to coaching, should schools develop their own coaching programs? Or should they encourage enrollment in commercial courses?

Reviews of research and evaluation findings often investigate these questions, but the reviewers have not reached a consensus on coaching programs. Vernon (1954) summarizing the findings of a number of British experts in one of the first reviews of this area, reported that the average effect of coaching and practice was to increase aptitude scores by nearly .6 standard deviations, equivalent to nine points on an IQ scale. Vernon pointed out that such an effect could be achieved in a remarkably short time, usually between three and nine hours, and recommended that all students be coached. Later reviews, on

the other hand, stated that the SAT and similar tests are largely resistant to the effects of drill and practice. The trustees of the College Board, for example, reported that the average increase to be expected from intensive coaching on the SAT amounts to less than .1 standard deviations, or less than ten points on the SAT score-scale of 200 to 800 (College Entrance Examination Board, 1968, p. 4). They therefore viewed coaching programs as a waste of time and money.

There are at least two reasons for these inconsistencies in conclusions. One is that different reviewers have looked at different parts of the literature on coaching. None of the available reviews has come close to including the entire range of relevant studies. Another problem is that reviewers have not all analyzed study results with objective, statistical tools. The use of such objective methods forces reviewers to put their notions to a quantitative test. Reviewers who use less formal methods often see what they expect to see in a collection of results.

Recent research syntheses that we completed recently at the University of Michigan were designed to overcome these major problems (Bangert-Drowns, Kulik, and Kulik, in press; Kulik, Bangert-Drowns, and Kulik, in press; Kulik, Kulik, and Bangert-Drowns, in press). The primary goal of our project was to evaluate the effectiveness of coaching for aptitude tests. But to provide a context for interpreting our major results, we also included in our analysis studies of practice on aptitude tests and studies of coaching and practice on achievement tests. The pool of studies included in our analyses was thus unusually

---

*Robert L. Bangert-Drowns is Research Assistant, James A. Kulik is Research Scientist, and Chen-Lin C. Kulik is Assistant Research Scientist, Center for Research on Learning and Teaching, The University of Michigan, Ann Arbor.*

broad. In analyzing our results, we used Glass's (1976) meta-analytic techniques, which he developed to improve the reliability and objectivity of research reviews.

### Method

We located the studies for our analyses through computer searches of the educational literature. Forty studies contained results on practice or retest effects; 38 studies contained results on coaching for aptitude tests; and 30 studies contained results on coaching for achievement tests. In the practice studies, investigators gave students practice tests—either parallel or identical to a criterion test—and then measured gains in scores from the practice test to the criterion test. In the coaching studies, investigators examined programs designed to raise test scores by teaching students relevant information or test-taking skills. The coaching programs covered such things as familiarization with test instructions, training in "testwiseness," relaxation or anxiety-reduction exercises, and review of the content covered in the tests.

The studies included in the analyses differed from one another in experimental design and in other key features. We created categorical variables, similar to those used in previous meta-analyses (Kulik, Kulik, and Cohen, 1980), to classify the studies according to such features. We used Glass's index of effect size *E*S to transform effects measured on different tests to a common scale (Glass, McGaw, and Smith, 1981). Glass's index gives the number of standard-deviation units that separate the group averages being compared.

### Effects of Practice

The size of practice effects turned out to be a function of three factors. The first of these was the similarity of the practice and criterion tests. When the two were parallel forms of the same test, the effect of one practice trial was to raise criterion scores by .23 standard deviations. When practice and criterion tests were identical, the effect of one practice trial was to raise criterion scores by .42 standard deviations.

It is possible to express these effects in more familiar terms. The standard deviation on most IQ tests is 15 points, and the standard deviation on most achievement measures is approximately ten months on a grade-equivalent scale. The gain from one practice trial on a parallel test is, therefore, approximately

three IQ points or two months in grade-level achievement; the gain from a practice trial on a test identical to the criterion is approximately six IQ points or four months in grade-level achievement.

The second factor that influenced the size of gains on the criterion was the number of practice tests taken. Effects increased regularly with increasing amounts of practice—for both parallel and identical forms of practice tests and for both aptitude and achievement tests. With practice on parallel forms of an IQ test, for example, the gain in measured IQ would be approximately three IQ points after one practice test, approximately seven IQ points after four practice tests, and approximately 11 IQ points after seven practice tests.

A third factor influencing results of practice was the ability level of the students tested. Higher ability students gained more from practice than did other students. The relationship between ability and gains from practice was especially clear when the practice was on a test identical to the criterion, but it was also noticeable when practice was given on parallel forms of a test. Higher ability students are apparently able to grasp the lesson from a simple practice test more easily than are lower ability students; low ability students may need more explicit coaching.

### Coaching for Aptitude Tests

Coaching affected student performance on aptitude tests, but the size of the effect was a function of two major factors. The first of these was the test on which students were being coached. Coaching programs had different effects on the SAT than they did on other aptitude tests.

The second factor that affected study results was the experimental design used in the study. Some coaching studies used posttest-only designs, whereas others used pretest-posttest designs. In the posttest-only studies, no pretesting was done, and the difference in posttest scores of the coached experimental group and uncoached control group was used as the effect of coaching. Pretest-posttest studies were more elaborate in design. In these studies, a pretest was given to both groups before the beginning of the coaching program, and then the posttest was given to both groups at the completion of the coaching program. The difference in gains for experimental and control groups was taken as the effect of coaching alone. Pretest-posttest studies sometimes yield larger

estimates of effect size than do posttest-only studies because the pretest may sensitize members of the experimental group to the treatment.

*Coaching for the SAT.* All 14 studies of SAT coaching used the pretest-posttest design. This design usually yields the maximum estimate of the size of an effect because the treatment effect may be accentuated by sensitization from the pretest. We found, however, that coaching effects were small on the SAT. Improvement from pretest to posttest averaged .36 standard deviations for the experimental groups and .21 standard deviations for the control groups. The difference between these two gains, .15 standard deviations, is the effect attributable to an SAT coaching program. The effect is equivalent to 15 points on the SAT scale from 200 to 800.

Although most SAT studies found small effects from coaching, a few studies produced impressive results. A well-designed study by Evans and Pike (1973), for example, yielded sizeable coaching effects. The study was carried out by ETS researchers who were thoroughly familiar with the SAT items pool and who developed special coaching materials for specific item types included in the pool. The Federal Trade Commission's (1979) coaching study also found reasonably strong gains for students enrolled in the coaching program offered by the Stanley H. Kaplan Educational Center.

*Coaching for other aptitude tests.* Seventeen of the studies of coaching for aptitude tests other than the SAT used a pretest-posttest design. Improvement from pretest to posttest averaged .76 standard deviations for the experimental groups and .25 standard deviations for the control groups. The difference between these two gains, .51 standard deviations, is the effect attributable to coaching alone. The seven studies that did not use pretests yielded a significantly lower estimate of the size of coaching effects: scores of coached and uncoached groups in these studies differed on the average by .27 standard deviations. On the basis of all 24 studies, we estimated the effect of coaching to be .43 standard deviations. This effect is equivalent to a gain in IQ of approximately 6 points.

### Coaching for Achievement Tests

Coaching programs for achievement tests were evaluated using both pretest-posttest and posttest-only designs. The average effect estimated from pretest-

posttest studies (based on the difference in gains of experimental and control groups) was .32 standard deviations. The average effect estimated from the posttest-only studies was .18. On the basis of all 30 studies, the average effect of a coaching program was to raise performance by .25 standard deviations, or approximately two or three months on the grade-equivalent scale. This effect was produced by a coaching program of average length. Longer coaching programs produced greater gains on achievement test scores; shorter programs produced smaller gains. No other study feature influenced the outcome of these coaching studies.

### Summary and Conclusions

Effects from simple practice were surprisingly strong. The effect from a single practice trial on a test parallel to the criterion was a gain of approximately .23 standard deviations, three IQ points, or two months on a grade-equivalent scale. Gains were greater with additional practice trials on parallel forms of a test. And gains were much greater when practice was provided on a test identical to the criterion. In addition, the size of the practice effect was a function of student ability, with high-ability students gaining more from practice than did low-ability students.

Coaching programs also produced fairly good results on most aptitude tests, even when the coached students did not take a practice test or pretest. Coaching without practice raised aptitude test

scores by .27 standard deviations, roughly equivalent to about four points on an IQ scale. Coaching with practice raised aptitude test scores by .76 standard deviations, equivalent to approximately 11 IQ points. About one-third of this gain could be attributed to practice provided by the pretest and two-thirds to the coaching program.

Coaching programs for the SAT had to be considered separately from other coaching programs. These programs had their own characteristics, and they produced their own outcomes. For most SAT coaching programs these outcomes were not impressive. The average coaching effect was weaker, in fact, than the effect of a single practice test given under standard conditions. And yet it is not clear that the SAT is coach-proof. ETS researchers with extensive knowledge of item format were able to produce dramatic changes in scores on specific item types, and Stanley Kaplan's well-known coaching program also seems to produce fairly impressive changes in SAT scores.

Finally, achievement tests and aptitude tests seem to be equally susceptible to influence from programs of special preparation. Practice effects were approximately equal on the two kinds of tests, and the average coaching effect on an achievement test fell between the average effect for the SAT and the average for all other aptitude tests. These results may be important to those who have argued that schools should use achievement tests rather than aptitude

tests in their school admissions decisions. Achievement tests can no longer be dismissed as unacceptably open to influence from programs of special preparation. □

### References

- Alderman, D. L., and Powers, D. E. "The Effects of Special Preparation on SAT-Verbal Scores." *American Educational Research Journal* 17 (1980): 239-253.
- Bangert-Drowns, R. L.; Kulik, J. A.; and Kulik, C.-L. C. "Effects of Coaching Programs on Achievement Test Performance." *Review of Educational Research*, in press.
- College Entrance Examination Board. *Effects of Coaching on Scholastic Aptitude Test Scores*. New York, NY: College Entrance Examination Board, 1968. (ERIC Document Reproduction Service No. ED 169 130.)
- Evans, F. R., and Pike, L. W. "The Effects of Instruction for Three Mathematics Item Formats." *Journal of Educational Measurement* 10 (1973): 257-272.
- Federal Trade Commission, Bureau of Consumer Protection. *Effects of Coaching on Standardized Admission Examinations: Revised Statistical Analyses of Data Gathered by Boston Regional Office of the Federal Trade Commission*. Washington, DC: Federal Trade Commission, Bureau of Consumer Protection, 1979. (NTIS No. PB-296 196.)
- Glass, G. V. "Primary, Secondary, and Meta-analysis of Research." *Educational Researcher* 5 (1976): 3-8.
- Glass, G. V.; McGaw, B.; and Smith, M. L. *Meta-analysis in Social Research*. Beverly Hills, Calif.: Sage Publications, 1981.
- Jencks, C.; and Crouse, J. "Should We Relabel the SAT—Or Replace It?" In *New Directions for Testing and Measurement: Measurement, Guidance, and Program Improvement*, No. 13. Edited by W. Shrader. San Francisco: Jossey-Bass, 1982.
- Kulik, J. A.; Bangert-Drowns, R. L.; and Kulik, C.-L. C. "Effectiveness of Coaching for Aptitude Tests." *Psychological Bulletin*, in press.
- Kulik, J. A.; Kulik, C.-L. C.; and Bangert-Drowns, R. L. "Effects of Practice on Aptitude and Achievement Test Scores." *American Educational Research Journal*, in press.
- Kulik, J. A.; Kulik, C.-L. C.; and Cohen, P. A. "Effectiveness of Computer-based College Teaching: A Meta-Analysis of Findings." *Review of Educational Research* 50 (1980): 525-544.
- Slack, W. V., and Porter, D. "The Scholastic Aptitude Test: A Critical Appraisal." *Harvard Educational Review* 50 (1980): 154-175.
- Vernon, P. E. "Practice and Coaching Effects in Intelligence Tests." *Educational Forum* 18 (1954): 269-280.



## Highlights from Research on the Effects of Coaching for Tests

In general, coaching improves student scores on achievement and aptitude tests—including the SAT—but not very much. While the SAT is not invulnerable to coaching programs, only a few students have found impressive results.

Coaching produces the following average gains:

- On the SAT, .15 standard deviations, equivalent to 15 points on the SAT scale of 200 to 800 points.
- On other aptitude tests, .43 standard deviations, equivalent to a gain in IQ of approximately six points.
- On achievement tests, .25 standard deviations, or approximately two to three months on the grade-equivalent scale.

Regardless of which of the three types of tests students are coached for, gains are greater when students (1) are given a pretest prior to coaching, (2) practice on tests identical to the criterion tests, and (3) practice more, on a regular schedule, for a longer period of time.

In addition, high-ability students gain more from coaching programs than do lower ability students, who may need more explicit instruction.

Copyright © 1983 by the Association for Supervision and Curriculum Development. All rights reserved.