

The Encyclopedia of Education

Second Edition

Derek C. Briggs

Test Preparation Programs: Impact

July, 2001

The impact of test preparation programs is a very controversial topic. The controversy hinges in large part upon how program impacts are quantified. Students taking commercial programs are usually given some sort of pre-test, a period of preparatory training, and then a post-test. The companies supplying these services will typically quantify impact in terms of average score gains for students from pre to post. Conversely, most research on the subject of test preparation will quantify impact as the average gain of students in the program relative to the average gain of a comparable group of students not in the program. Here the impact of a test preparation program is equal to its estimated causal effect. The latter definition of impact is the more valid one when the aim is to evaluate the costs and benefits of a test preparation program. Program benefits should always be expressed relative to the outcome a person could expect had they chosen not to participate in the program. This review takes the approach that program impact can only be assessed by estimating program effects.

Test preparation programs share two common features. First, students are prepared to take one specific test. Second, the preparation students receive is systematic. Test preparation programs have been found to have differing degrees of the following characteristics:

1. Instruction that develops the skills and abilities that are to be tested.
2. Practice on problems that resemble those on the test.
3. Instruction and practice with test-taking strategies.

Short-term programs that include primarily this third characteristic are often classified as "coaching." There are strong shades of gray in this definition. First, the difference between "short-term" and "long-term" programs is difficult to quantify. Some researchers have used roughly 40 to 45 hours of student contact time as a threshold, but this is not set in stone (Jackson 1980). Second, within those programs classified as short-term or long-term, the intensity of preparation may differ. One might speculate that a program with 20 hours of student contact time spaced over one week is quite different from one with 20 hours spaced over one month. Finally, test preparation programs may include a mix of the three characteristics listed above. It is unclear what proportion of student contact time must be spent on test-taking strategies before a program can be classified as coaching. For more detailed analyses of the components of test coaching see Cole (1982) and Bond (1989).

There is little research base from which to draw conclusions about the effectiveness of preparatory programs for achievement tests given to students during their primary and secondary education. This may change in the early 21st century as these tests are used for increasingly high stakes purposes, particularly in the United States. There is a much more substantial research base with regard to the effectiveness of test preparation programs on achievement tests taken for the purpose of post-secondary admissions. The remainder of this review will focus on the effectiveness of this class of programs.

Program Effects on College Admissions Tests

Research studies on the effects of admissions test preparation programs have been published periodically since the early 1950s. Most of this research has been concerned with the effect of test preparation on the SAT, the most widely taken test for the purpose of college admission in the United States. A far smaller number of studies have considered the preparation effect on the ACT, another test often required for US college admission. On the main issues, there is a strong consensus in the literature:

- Test preparation programs have a statistically significant effect on the changes of SAT and ACT scores for students taking the test at least twice.
- The magnitude of this effect is quite small.

How small? The SAT consists of two sections, one verbal and one quantitative, each with a score range from 200 to 800 points. While the section averages of all students taking the SAT each year varies slightly, the standard deviation around these averages is pretty consistently about 100 points. The effect of test preparation programs on the verbal section of the SAT is probably between .05 and .15 of a standard deviation; the effect on the quantitative section is probably between .15 and .25 of a standard deviation (Powers 1993; Powers and Rock 1999; Briggs 2001).

The ACT consists of four sections: science, math, English and reading. Scores on each section range from 1 to 36 points and test-takers are also given a composite score for

all four sections on the same scale. The standard deviation on the test is usually about 5 points. A smaller body of research, most of it unpublished, has found a test preparation effect (expressed as a percentage of the 5 point standard deviation) of .02 on the composite ACT score, an effect of about .04 to .06 on the math section, an effect of about .08 to .12 on the English section, and a negative effect of .12 to .14 on the reading section (Briggs 2001).

The importance of how test preparation program effects are estimated cannot be overstated. In most studies, researchers are presented with a group of students who participate in a preparatory program in order to improve their scores on a test they have already taken once. Estimating the effect of the program is a question of causal inference: how much does exposure to systematic test preparation cause a student's test score to increase above the amount it would have increased without exposure to the preparation? Estimating this causal effect is quite different from calculating the average score gain of all students exposed to the program. A number of commercial test preparation companies have advertised—even guaranteed—score gains based on those calculated from students previously participating in their program. This confuses gains with effects. To determine if the program itself has a real effect on test scores one must contrast the gains of the treatment group (students who have taken the program) to the gains of a comparable control group (students who have not taken the program). This would be an example of a controlled study, and this is the principle underlying the estimation of effects for test preparation programs.

One vexing problem is how to interpret the score gains of students participating in preparatory programs in the absence of a control group. Some researchers have attempted to do this by subtracting the average gain of students participating in a program from the expected gain of the full test-taking population over a given time period (Slack and Porter 1980). This has been criticized primarily on the grounds that the former group is never a randomly drawn sample from the latter. Students participating in test preparation are in fact often systematically different from the full test-taking population along important characteristics that are correlated with test performance. This is an example of self-selection bias. Due in part to this bias, uncontrolled studies have been found to consistently arrive at estimates for the effect of test preparation programs that are as much as four to five times greater than those found in controlled studies. (DerSimonian and Laird 1983).

Is Test Preparation more effective under certain programs for certain people?

In a comprehensive review of studies written between 1950 and 1980 that estimate an effect for test preparation, Samuel Messick and Ann Jungblut (1981) found evidence of a positive relationship between time spent in a program and the estimated effect on SAT scores. But this relationship was not linear; there were diminishing returns to SAT score changes for time spent in a program beyond 45 hours. Messick and Jungblut concluded that "the student contact time required to achieve average score increases much greater than 20 to 30 points for both the SAT-V and SAT-M rapidly approaches that of full-time schooling" (p. 215). Since the Messick and Jungblut

review, several reviews of test preparation studies have been written by researchers using the statistical technique known as meta-analysis. Use of this technique allows for the synthesis of effect estimates from a wide range of studies conducted at different points in time (DerSimonian and Laird 1983; Kulik, Bangert-Drowns et al. 1984; Becker 1990). The findings from these reviews suggest that there is little systematic relationship between the observable characteristics of test preparation programs and the estimated effect on test scores. In particular, once a study's quality was taken into consideration, there was at best a very weak association between program duration and test score improvement.

There is also mixed evidence as to whether test preparation programs are more effective for particular subgroups of test-takers. Many of the studies that demonstrate interactions between the racial/ethnic and socioeconomic characteristics of test-takers and the effects of test preparation suffer from very small and self-selected samples. Because commercial test preparation programs charge a fee, sometimes a substantial one, most students participating in such programs tend to be socioeconomically advantaged. In one of the few studies with a nationally representative sample of test-takers, test preparation for the SAT was found to be most effective for students coming from high socioeconomic backgrounds. A similar association was not found among students who took a preparatory program for the ACT (Briggs 2001).

Discussion

The differential impact of test preparation programs on student subgroups is an area that merits further research. In addition, a theory describing why the pedagogical practices used within commercial test preparatory programs in the 21st century would be expected to increase test scores has, with few exceptions (Evans and Pike 1973; Powers 1986), not been adequately explicated or studied in a controlled setting at the item level. In any case, after more than five decades of research on the issue, there is little doubt that commercial preparatory programs, by advertising average score gains without reference to a control group, are misleading prospective test-takers about the benefits of their product. The costs of such programs are high, both in terms of money and in terms of opportunity. For consumers of test preparation programs, these benefits and costs should be weighed carefully.

Bibliography

Becker, B. J. (1990). "Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal." Review of Educational Research 60(3): 373-417.

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

Briggs, D. C. (2001). "The Effect of Admissions Test Preparation: Evidence from NELS:88." Chance 14(1): 10-18.

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

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

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

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

Jackson, R. (1980). "The Scholastic Aptitude Test: A response to Slack and Porter's "Critical Appraisal"." Harvard Educational Review 50(3): 382-391.

Kulik, J. A., R. L. Bangert-Drowns, et al. (1984). "Effectiveness of coaching for aptitude tests." Psychological Bulletin 95: 179-188.

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

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

Powers, D. E. (1993). "Coaching for the SAT: A summary of the summaries and an update." Educational Measurement: Issues and Practice(Summer): 24-39.

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

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