DOCUMENT RESUME

ED 242 757 TM 840 168

AUTHOR Halpin, Gerald; And Others
TITLE The Effects of a Program to Increase CAT Scores.
PUB DATE Apr 84
NOTE 25p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 24-26, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Gains; *Achievement Tests; Cost Estimates; Elementary Education; High Achievement; Intentional Learning; Multivariate Analysis; *Program Effectiveness; Scores; *Test Coaching; Test Wiseness
IDENTIFIERS *California Achievement Tests; *Scoring High on the California Achievement Tests

ABSTRACT

The effects of a test coaching program, "Scoring High on the California Achievement Test," were investigated with a sample of 876 students in grades 1, 2, 4, and 5. Multivariate analyses of variance were used to determine the effects of the program, grade level, sex, and race. Significant differences in favor of the Scoring High program were found on some of the subtest scores of the California Achievement Tests (CAT). Significant interaction effects prevented a straightforward interpretation of the program's impact. Some of the students in some grades increased some CAT subtest scores as a result of participating in the program. The authors state that since the program did not produce clear evidence that it is a valuable tool to increase student performance on the CAT, school administrators and others involved in the decision to use this particular program should weigh the results of this study as well as the financial requirements (approximately $2.00 per student) and the use of valuable classroom instruction time as they consider whether some small increases in some test scores for some students warrants the purchase and use of the program. (Author/PN)

THE EFFECTS OF A PROGRAM TO INCREASE CAT SCORES

Gerald Halpin, Auburn University
William L. Deaton, Auburn University
Tim Alford, Enterprise, AL

Presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, April, 1984

The Effects of a Program to Increase Scores on the CAT

Gerald Halpin, Auburn University
William L. Deaton, Auburn University
Tim Alford, Enterprise, AL School System

The effects of the Scoring High on the CAT program were investigated with a sample of 876 students in grades 1, 2, 4 and 5. Multivariate analyses of variance were used to determine the effects of the program, grade level, sex and race. Significant differences ($p<.01$) in favor of the Scoring High program were found on some of the subtest scores of the CAT. Significant interaction effects prevented a straightforward interpretation of the program's impact. Some of the students in some grades increased some CAT subtest scores as a result of participating in the program.

THE EFFECTS OF A PROGRAM TO INCREASE CAT SCORES

Considerable research and controversy have centered around the effects of programs designed to develop test-taking skills for aptitude and entrance tests (DerSimonian & Laird, 1983; Jackson, 1980; Messick & Jungeblut, 1981; Slack & Porter, 1980). Although these effects are dependent upon the amount and kind of coaching, it has been concluded that scores on aptitude and entrance measures can be influenced by coaching programs (Messick & Jungeblut, 1981; Slack & Porter, 1980). But what about achievement tests? Are achievement test results also affected by programs designed to improve examinee performance?

In 1954, Vernon hypothesized that coaching effects would be as great for achievement tests as they were for aptitude measures, but no empirical support for this statement was offered. Since that time, some studies regarding the effect of coaching on achievement tests have been done, but they have not received the attention of researchers as have aptitude tests chiefly because achievement test scores have tended not to be as crucial as aptitude test scores in many important scholastic decisions.

To summarize the research conducted in this area, Bangert-Drowns, Kulik and Kulik (1983) performed a meta-analysis of studies of the effects of coaching programs on achievement test scores.

They considered interventions at three basic levels: (a) short practice sessions including a test-taking orientation, (b) longer interventions that included drill or practice on sample test items, and (c) work on broader cognitive skills (cf. Anastasi, 1981; Mehrens & Lehmann, 1984). While the individual investigations showed variations in effectiveness, the authors concluded that the typical coaching program did improve achievement test scores. They further found that the effects were smallest for short test-taking orientation sessions, larger in longer programs actually involving intensive drill and practice, and largest in lengthy, broad skill training programs (cf. McPhail, 1977).

Studies reviewed by Bangert-Drowns et al. (1983) were limited and results were applicable only to the achievement tests and intervention programs included in their analyses. Much more research is needed before educators can use with confidence -- or not use at all -- the various intervention strategies that are available including the extensive commercially prepared programs. One such program in need of research is Scoring High on the California Achievement Test (Bates, 1981). The publishers of advertise it as "America's best selling test preparation series with over 5,000,000 copies in print." Although this program is presented as if it will have a substantial impact upon student achievement on the California Achievement Tests, no evidence is given to support this claim. This experimental investigation was undertaken to investigate the publisher's assertion.

Specifically, two questions were considered:

1. What are the effects of the Scoring High program (Bates, 1981) on scores obtained on the California Achievement Tests (CAT) for students in grades 1, 2, 4, and 5?

2. Are these effects moderated by grade level, sex, and race of the students?

Procedures

Participants in this investigation were 925 students in 40 classes at five elementary schools within one city system in the Southeast. At grades 1, 2, 4, and 5 within each of the five elementary schools, two intact classes were randomly selected and randomly assigned to either an experimental group or to a control group. Altogether there were 460 students in the 20 classes comprising the experimental group and 465 in the 20 classes assigned to the control group. The experimental group received the Scoring High on the California Achievement Test program. The control group received no formal instruction in test-taking skills.

The policy of the school system in which this study was conducted was to stratify students by race (black/white) and sex at each grade level within each school and then assign students to classes alphabetically within each race and sex combination. While this process was not one of true randomization, it is reasonable to assume that the students in different classes within schools at each grade level did not differ systematically.

Scoring High on the California Achievement Test is described by the publishers (Random House, 1983) as a program that enables students to master test-taking skills and strategies needed to maximize their performance on the California Achievement Tests. There are three books of the Scoring High program at three levels: Book A is for students in grades 1 - 4, Book B is for students in grades 3 - 6, and Book C is designed for students in grades 5 - 8. In this study, Book A was used for students in grades 1 and 2 and Book B was used with students in grades 4 and 5. Each of the 30 plus lessons in each book contained sample directions similar to those on the CAT and one or more sample items along with tips on test-taking strategies. Also included in each daily lesson was a practice section which included test items presented in formats similar to those on the CAT.

The Scoring High program was completed by students in the experimental group in March and early April of 1983. Toward the end of April, both the experimental and control group students received the California Achievement Test battery as a part of the state testing program.

Criterion measures in the study were the scale scores attained by the students on the Mathematics, Reading, and Language subtests of the CAT. Since the CAT subtest scores are not comparable measures, no attempt was made to make comparisons across subtest scores.

Results and Conclusions

Separate multivariate analyses of variance (MANOVA) were performed for grades 1 and 2 and for grades 4 and 5 since Book A of the Scoring High program was used in the first two grades and Book B was used in grades 4 and 5. Within each of these two grade level groups, MANOVAs were conducted on each of the three sections of the CAT: Reading (vocabulary, comprehension, and total reading scores), Mathematics (computation, concepts and applications, and total mathematics scores), and Language (mechanics, expression, and total language scores). A 2 X 2 X 2 X 2 sampling design (group by grade by sex by race) was used to determine if any statistically significant differences existed between the groups or the interactions of the group effect with the other effets.

The observed means and standard deviations for levels of main effects for each of the subtests of the CAT are shown in Table 1 (students in grade 1 did not receive the CAT language subtests). The students receiving the Scoring High program scored higher than the control group students on all CAT subtests except reading comprehension. Differences in favor of the experimental group varied from 5 to 14 scale score points which would correspond to about 2 or 3 months at most in grade equivalent scores. Table 1 also shows differences on subtest scores between boys and girls, whites and blacks, and grade level.

Tables 2, 3, and 4 contain the results of the multivariate analyses of variance for students in grades 4 and 5.

Statistically significant differences ($p < .01$) were found for three of the main effects (group, race, and grade) on all sets of CAT scores. A significant sex effect ($p < .01$) was noted for the mathematics and language scores. Three statistically significant ($p < .05$) interaction effects were found but only one of these contained the group effect. The mathematics scores showed a significant multivariate F-statistic for the group X grade interaction effect.

Since this study was designed to investigate the effects of the Scoring High program, only those significant multivariate effects involving the experimental and control groups were followed with univariate analyses of variance (ANOVA). Table 5 is a summary of the ANOVAs conducted on the CAT subtest scores which had significant multivariate group effects.

The group X grade effect on the mathematics section of the CAT was attributed to control group students in grade 4 showing higher scores on all three mathematics scores than students in the experimental group. However, at grade 5, students receiving the Scoring High program outperformed the control group students on all mathematics subtests. Table 6 contains the means and standard deviations of the CAT mathematics scores for the four combinations of the group by grade interaction effect.

The significant MANOVA F statistic for the group effect on the three reading subtest scores was followed with three univariate analyses of variance. From Table 5, though none of the F-ratios

are significant at the .05 level, reading comprehension and reading vocabulary showed greatest differences between the experimental and control groups. The experimental group students had higher scores on vocabulary and slightly lower scores on comprehension than students in the control group.

Language subtest scores showed differences between the groups on language mechanics and total language scores. The experimental group scored higher than the control group on both of these subtests ($p < .01$).

For students in grades 4 and 5, Scoring High on the CAT appeared to be somewhat effective in raising scores on a few subtests in reading and language of the CAT. Mathematics scores for fifth grade students showed an increase but lower math subtest scores were observed for fourth grade students receiving the Scoring High program.

Summaries of the MANOVAs performed on reading and mathematics subtest scores for students in grades 1 and 2 are given in Tables 7 and 8. Since students in the first grade did not receive the language portion of the CAT, Table 9 contains the MANOVA summary on these subtest for students in the second grade.

A significant ($p < .05$) four-way interaction was found on the reading subtest scores. Separate ANOVAs conducted on the three reading subtest scores suggested that only the reading vocabulary scores were involved in this interaction effect (see Table 10). Cell means and standard deviations for reading vocabulary scores

corresponding to this interaction effect are shown in Table 11. Though not conclusive, Table 11 reflects the very slight group differences at grade 1 as well as consistent sex differences for the white students at grade 1. First grade black males showed higher scores than first grade black females in the experimental group but the reverse was true in the control group.

Grade two results demonstrated little differences between the control and experimental groups for white students yet black females seemed to benefit by the Scoring High program even though black males did not. Also contained in Table 11 is the differential grade effect; white males and females and black males tended to show increases of about 40 or 60 scale score points from grade 1 to grade 2 regardless of the group to which they had been assigned. Black females in the control group showed similar differences from grade 1 to grade 2 but black females who had received the Scoring High program in grade 2 were over 80 scale score units above the first grade black females in the experimental group.

The MANOVA results on the mathematics subtest scores are presented in Table 8. A significant three-way interaction effect ($p < .05$) was found and followed with ANOVAs, reported in Table 10, determined that the group X sex X race interaction was only for the math computation subtest scores. Table 12 contains the means and standard deviation for the combinations of this interaction effect. White male and female students showed little differences in mean

scores between the control and experimental groups. However, black students in the Scoring High program had higher scores than black students in the control group; this difference was most pronounced for black males. Although differences appear to be small, the values in Table 12 are mean differences across both grade levels.

Students in grade 2 did not show significant multivariate effects involving the group variable on language subtest scores. Table 9 presents these results.

The effectiveness of the Scoring High program on student achievement at grades 1 and 2 was difficult to interpret because of the confounding influence of sex, race, and grade level on the group variable. The significant interaction effects may have been due, in part, to the level of the materials used. That is, Book A of the Scoring High program may not have been appropriate for these students. The differential effects observed may also have been related to a lack of test taking experiences of the students.

In summary, Scoring High on the CAT did not produce clear evidence that it is a valuable tool to increase student performance on the CAT. School administrators and others involved in the decision to use this particular program should weigh the results of this study as well as the financial requirements (approximately $2.00 per student) and the use of valuable classroom instruction time as they consider whether some small increases in some test scores for some students warrants the purchase and use of the program.

Table 1
Means and Standard Deviations on CAT Subtest Scores
for Levels of the Main Effects

		Reading			Language			Mathematics		
		Vocab	Compr	Total	Mech	Exp	Total	Compu	Conce	Total
Group										
E	X	433.5	446.0	431.9	542.6	503.6	490.2	403.6	438.2	424.4
	S	75.9	77.4	80.0	70.3	70.2	91.9	94.7	75.2	78.8
	N	450	447	445	337	337	337	450	450	449
C	X	428.8	446.0	429.2	528.5	498.2	474.2	397.2	423.1	413.2
	S	73.1	84.1	85.7	57.9	68.2	86.5	91.6	66.0	71.3
	N	465	447	462	346	345	345	465	464	464
Sex										
M	X	428.5	442.6	427.4	529.9	498.8	475.7	394.9	429.6	416.3
	S	74.8	83.7	84.6	67.2	69.6	90.8	91.8	72.8	75.7
	N	476	472	471	344	343	343	472	472	471
F	X	433.9	449.6	434.0	541.0	502.9	489.0	406.2	431.5	421.4
	S	74.1	77.6	81.0	61.4	68.8	87.4	94.2	69.3	74.8
	N	439	438	436	339	339	339	443	442	442
Race										
B	X	395.8	405.2	386.6	504.1	455.1	432.3	371.5	393.4	387.0
	S	68.6	68.9	74.2	62.1	55.4	80.3	91.0	64.6	69.0
	N	212	212	210	148	148	148	214	215	214
W	X	442.2	458.9	444.3	544.3	513.9	497.6	409.7	442.4	428.9
	S	72.8	80.0	80.6	62.5	66.8	86.2	91.8	68.8	74.2
	N	700	695	694	534	533	533	698	696	696
Grade										
1	X	345.0	359.8	335.6				292.9	351.8	334.1
	S	31.2	39.2	38.4				29.6	36.8	27.9
	N	230	229	227				231	232	231
2	X	402.8	415.0	404.7	478.5	437.3	455.0	349.9	405.5	382.6
	S	27.8	38.4	43.4	36.9	39.4	49.0	23.8	36.3	27.1
	N	230	226	226	226	226	226	227	226	226
4	X	465.5	475.8	463.3	547.2	513.3	525.4	448.7	461.0	452.7
	S	46.2	55.6	53.1	56.1	57.4	62.4	46.4	43.8	41.3
	N	219	219	218	219	218	218	218	217	217
5	X	510.6	531.6	516.3	578.6	549.4	561.9	508.0	502.9	503.9
	S	49.3	56.2	54.3	52.3	54.5	57.2	50.4	48.3	46.5
	N	236	236	236	238	238	238	239	239	239

Table 2
MANOVA Summary: Reading Vocabulary, Comprehension, Total
Grades 4 and 5

Source	Wilk's Lambda	F
Group	.970	4.53**
Sex	.992	1.11
Race	.816	32.80**
Grade	.727	54.50**
Group X Sex	.983	2.46
Group X Race	.991	1.34
Group X Grade	.985	2.1[illegible]
Sex X Race	.981	2.75*
Sex X Grade	.972	4.25**
Race X Grade	.994	0.85
Group X Sex X Race	.993	1.06
Group X Sex X Grade	.992	1.16
Group X Race X Grade	.996	0.51
Sex X Race X Grade	.992	1.19
Group X Sex X Race X Grade	.999	0.08

---

df = 3, 436
**p < .01
*p < .05

Table 3
MANOVA Summary: Mathematics Computation, Concepts, Total
Grades 4 and 5

Source	Wilk's Lambda	F
Group	.927	11.54**
Sex	.966	5.07**
Race	.851	25.64**
Grade	.706	60.80**
Group X Sex	.984	2.39
Group X Race	.996	0.61
Group X Grade	.964	5.38**
Sex X Race	.996	0.54
Sex X Grade	.989	1.57
Race X Grade	.994	0.82
Group X Sex X Race	.998	0.35
Group X Sex X Grade	.991	1.28
Group X Race X Grade	.999	0.09
Sex X Race X Grade	.998	0.24
Group X Sex X Race X Grade	.999	0.12

df = 3, 438
**p < .01
*p < .05

Table 4
MANOVA Summary: Language Mechanics, Expression, Total
Grades 4 and 5

Source	Wilk's Lambda	F
Group	.935	10.18**
Sex	.970	4.46**
Race	.813	33.69**
Grade	.809	34.48**
Group X Sex	.990	1.41
Group X Race	.984	2.29
Group X Grade	.990	1.42
Sex X Race	.998	0.28
Sex X Grade	.986	2.11
Race X Grade	.996	0.64
Group X Sex X Race	.998	0.22
Group X Sex X Grade	.984	2.34
Group X Race X Grade	.996	0.58
Sex X Race X Grade	.989	1.64
Group X Sex X Race X Grade	.997	0.39

df = 3, 438
**p < .01
*p < .05

Table 5
Summary of Analyses of Variance Involving Group Effect
Grades 4 and 5

Source	df	Variable	MS Hyp	MS Error	F
Group	1,438	Read Vocab	3278.05	1958.23	1.67
		Read Compr	6483.62	2637.66	2.46
		Read Total	969.02	2419.10	0.40
Group	1,440	Math Compu	966.29	2189.22	0.44
		Math Conce	41727.95	1765.91	23.63**
		Math Total	17265.30	1699.78	10.16**
Group	1,440	Lang Mech	46196.06	2608.48	17.71**
		Lang Expr	597.08	2606.36	0.23
		Lang Total	20684.75	3008.96	6.87**
Group X Grade	1,440	Math Compu	35075.18	2189.22	16.02**
		Math Conce	13460.99	1765.91	7.62**
		Math Total	23231.70	1699.78	13.67**

**p < .01
*p < .05

Table 6
Cell Means and Standard Deviations of the
Group X Grade Effect on Mathematics Subtest Scores
Grades 4 and 5

Group	Grade	N	X	S
Math Computation Subtest				
Control	4	111	457.02	39.86
	5	117	498.70	50.68
Experimental	4	107	440.16	48.09
	5	122	516.92	50.06
Math Concepts Subtest				
Control	4	110	457.44	31.90
	5	117	488.32	42.56
Experimental	4	118	421.37	46.16
	5	122	516.88	45.82
Math Total				
Control	4	110	454.32	31.02
	5	117	491.32	42.78
Experimental	4	118	409.06	44.15
	5	122	516.03	45.38

Table 7
MANOVA Summary: Reading Vocabulary, Comprehension, Total
Grades 1 and 2

Source	Wilk's Lambda	F
Group	.996	0.61
Sex	.981	2.72*
Race	.780	40.54**
Grade	.452	174.65**
Group X Sex	.995	0.69
Group X Race	.993	1.02
Group X Grade	.998	0.33
Sex X Race	.998	0.24
Sex X Grade	.999	0.17
Race X Grade	.986	2.04
Group X Sex X Race	.990	1.45
Group X Sex X Grade	.989	1.63
Group X Race X Grade	.999	0.09
Sex X Race X Grade	.991	1.26
Group X Sex X Race X Grade	.980	2.88*

df = 3, 432
**p < .01
*p < .05

Table 8
MANOVA Summary: Mathematics Computation, Concepts, Total
Grades 1 and 2

Source	Wilk's Lambda	F
Group	.980	2.96*
Sex	.982	2.60
Race	.799	36.53**
Grade	.414	205.78**
Group X Sex	.989	1.63
Group X Race	.982	2.64
Group X Grade	.997	0.44
Sex X Race	.993	1.02
Sex X Grade	.985	2.20
Race X Grade	.976	3.53*
Group X Sex X Race	.979	3.12*
Group X Sex X Grade	.990	1.52
Group X Race X Grade	.992	1.09
Sex X Race X Grade	.983	2.46
Group X Sex X Race X Grade	.992	1.22

df = 3, 436
**p < .01
*p < .05

Table 9
MANOVA Summary: Language Mechanics, Expression, Total
Grade 2

Source	Wilk's Lambda	F
Group	.976	1.75
Sex	.934	5.05**
Race	.728	26.79**
Group X Sex	.978	1.57
Group X Race	.977	1.69
Sex X Race	.930	5.35**
Group X Sex X Race	.973	1.97

df = 3, 215
**p < .01
*p < .05

Table 10
Summary of Analyses of Variance Involving Group Effect
Grades 1 and 2

Source	df	Variable	MS Hyp	MS Error	F
Group	1,438	Math Compu	1579.10	605.25	2.61
		Math Conce	6541.93	1085.82	6.02*
		Math Total	2675.10	622.43	4.30*
Group X Race	1,439	Math Compu	4345.57	605.25	7.18**
		Math Conce	1336.15	1085.82	1.23
		Math Total	2239.25	622.43	3.60
Group X Sex X Race	1,438	Math Compu	4103.34	605.25	6.78**
		Math Conce	744.97	1085.82	0.69
		Math Total	1190.22	622.43	1.91
Group X Sex X Race X Grade	1,438	Read Voc	4488.00	684.25	6.56**
		Read Compr	2041.54	1297.50	1.57
		Read Total	1608.64	1375.89	1.17

**p < .01
*p < .05

Table 11
Means and Standard Deviations of Reading Vocabulary Scores
Group X Sex X Race X Grade Effect
Grades 1 and 2

Grade 1

Group	Sex	Race	N	Mean	S
Control	Male	White	44	352.6	27.7
	Male	Black	16	318.5	28.2
	Female	White	38	355.4	23.5
	Female	Black	17	325.3	25.7
Experimental	Male	White	48	353.1	26.4
	Male	Black	18	330.2	23.9
	Female	White	33	359.6	25.0
	Female	Black	11	317.6	36.0

Grade 2

Control	Male	White	45	404.6	27.0
	Male	Black	12	380.3	30.7
	Female	White	43	410.2	20.1
	Female	Black	17	383.0	32.3
Experimental	Male	White	45	411.8	23.6
	Male	Black	9	374.3	34.8
	Female	White	35	406.1	25.2
	Female	Black	19	401.5	27.5

Table 12
Means and Standard Deviations of
Mathematics Computation Scores
Group X Sex X Race Effect
Grades 1 and 2 Combined

Group	Sex	Race	N	Mean	S
Control	Male	White	90	324.5	25.9
	Male	Black	29	286.6	27.3
	Female	White	81	330.6	23.1
	Female	Black	35	308.2	26.0
Experimental	Male	White	93	325.2	21.8
	Male	Black	27	313.6	31.1
	Female	White	67	328.2	21.0
	Female	Black	32	315.5	29.0

References

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

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53, 571-585.

Bates, K. L. (1981). Scoring high on the California Achievement Tests (Books A & B). New York: Random House.

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

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

McPhail, I. P. (1977). A psycholinguistic approach to training urban high school students in test-taking strategies (Doctoral dissertation, University of Pennsylvania, 1976). Dissertation Abstracts International, 37, 5667A.

Mehrens, W. A. & Lehman, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart, and Winston.

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

Random House 1983 school division catalog. (1983). Westminster, MD: Random House.

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

Vernon, P. E. (1954). Practice and coaching effects in intelligence tests. Educational Forum, 18, 269-280.