DOCUMENT RESUME

ED 137 376 TM 006 180

AUTHOR DeVito, Pasquale J.; Long, John V.
TITLE The Effects of Spring-Spring vs. Fall-Spring Testing Upon the Evaluation of Compensatory Education Programs.
PUB DATE [Apr 77]
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Gains; *Achievement Tests; *Compensatory Education Programs; Disadvantaged Youth; Elementary Education; Models; Post Testing; Pretesting; *Program Evaluation; *Student Testing; *Time
IDENTIFIERS *Elementary Secondary Education Act Title I

ABSTRACT

The impact of achievement testing on a Spring-Spring vs. a Fall-Spring basis in evaluation of compensatory education programs is examined. The effects of the summer vacation period and crossing test levels were examined. Significant decreases in achievement test scores were found, at all grade levels, between Spring and Fall test administrations. Substantially larger decreases were noted when testing levels were crossed. The implications of differences in achievement scores, a function of time and level of testing, are discussed in terms of the RMC Title I evaluation models and the current emphasis placed upon longitudinal examination of compensatory programs. (Author/RC)

THE EFFECTS OF SPRING-SPRING VS. FALL-SPRING TESTING UPON THE EVALUATION OF COMPENSATORY EDUCATION PROGRAMS

Pasquale J. DeVito
Rhode Island Department of Education

John V. Long
University of Rhode Island

Paper presented at the annual convention of the American Educational Research Association, New York City, April, 1977.

ABSTRACT

This study examined the impact of achievement testing on a Spring-Spring vs. a Fall-Spring basis in evaluation of compensatory education programs. The effects of the summer vacation period and crossing test levels were examined. Significant decreases in achievement test scores were found, at all grade levels, between Spring and Fall test administrations. Substantially larger decreases were noted when testing levels were crossed.

The implications of differences in achievement scores, a function of time and level of testing, are discussed in terms of the RMC Title I evaluation models and the current emphasis placed upon longitudinal examination of compensatory programs.

Title I of the Elementary and Secondary Education Act of 1965, as amended, (ESEA) authorizes financial assistance to local school districts with concentrations of economically disadvantaged students. These funds are designated specifically for programs designed to meet the special needs of the educationally disadvantaged student. The legislation for Title I, (Public Law 93-380, section 151), requires that the Commissioner of Education "provide for independent evaluations which describe and measure the impact of programs and projects assisted under this title." In order to provide a uniform and standard format of reporting impact, the United States Office of Education (USOE) has awarded two contracts to RMC Research Corporation for the development and refinement of a series of models for evaluating the cognitive outcomes of Title I compensatory education academic programs.

In documents explaining the evaluation models developed under these contracts RMC Research Corporation has identified several "hazards" which, if violated, would invalidate the evaluation findings. One hazard cited, the use of non-comparable (non-identical) pretest and posttest levels, becomes a severe restriction upon the evaluation of local Title I programs when annual (Spring to Spring) testing is used for assessing achievement. Since Spring to Spring testing is common in many school districts, the problem of crossing testing levels becomes acute when examined in light of this hazard. If this limitation is a valid one then the problem becomes even more acute and techniques developed by RMC for aggregating data across school district and states are employed.

A further problem associated with Spring to Spring testing is the

hypothesis of the "summer drop". Since many communities examine program impact on a Fall-Spring basis, the effect of this summer drop will make comparison (or aggregation) of scores from communities which test on a Fall-Spring schedule with those following a Spring-Spring testing schedule extremely difficult if the magnitude of the summer drop is not taken into account.

METHOD

The sample consisted of 238 students classified as educationally disadvantaged in grades two through six in a moderate-sized Rhode Island school district. One hundred twenty-six (126) students were tested with the California Achievement Test (CAT) at two testing times with the same level instruments while one hundred twelve (112) students were given a different level of the test for the second administration. Table 1 presents numbers of students, test dates, levels and forms for the sample in the study.

TABLE 1

Test Levels and Forms by Grade

	Spring, 1974	Fall, 1975
Grade 2 N = 69	Level 1, Form B	Level 2, Form A
Grade 3 N = 60	Level 2, Form B	Level 2, Form A
Grade 4 N = 43	Level 2, Form B	Level 3, Form A
Grade 5 N = 66	Level 3, Form B	Level 3, Form A

RESULTS

Table 2 presents, for each of the four grades examined as part of the study, the Spring, 1974 and Fall, 1975 mean Achievement Development Scale Scores (ADSS). Examination of these scores indicates that there was a decline in mean ADSS at all grade levels between Fall and Spring test administrations. At 3 of the 4 grade levels, these declines were statistically significant at the .01 level.

TABLE 2

Mean ADSS, Standard Deviation, t Values and Significance for California Achievement Test Total Reading

	Spring, 1974		Fall, 1975		t	p
	$\bar{X}_{ss}$	SD	$\bar{X}_{ss}$	SD		
Grade 2 (N=69)	309	21	281	29	8.12	<.01
Grade 3 (N=60)	312	30	305	31	1.78	N.S.
Grade 4 (N=43)	347	34	336	31	2.82	<.01
Grade 5 (N=66)	362	35	344	37	4.48	<.01

One of the key elements in the RMC Research Corporation evaluation models is the emphasis upon a clear definition of "no-treatment expectation", i.e., how well the students would be predicted to achieve without supplementary assistance. Since in Model A, the norm-referenced model, the no-treatment expectation is based upon percentile rank associated with

mean pretest standard score, mean standard scores for the testing in this study were converted to percentiles. Table 3 presents these percentiles.

TABLE 3

Percentile Rank, Both Administrations for California Achievement Test Total Reading

	Spring, 1974 Percentile Rank	Fall, 1975 Percentile Rank
Grade 2	33	12
Grade 3	14	10
Grade 4	20	11
Grade 5	13	8

The examination of these percentiles highlight several points. First, in all cases, Fall percentile scores are lower than Spring percentile scores. Second, where levels of the CAT have been crossed (grade two and four), Spring-Fall declines in percentile scores are greater than where common test levels are used (grade 3 and 5). Third, Fall percentile scores are more homogenous across grade levels than are percentile scores for Spring testing.

DISCUSSION

The preceding section has shown that achievement test scores earned by the same students are different when these students are tested in the Spring and the Fall. As a result of these differences, the use of students' Spring achievement test scores as pretest measures will yield different predicted posttest scores than if their Fall achievement test scores were used as pretest measures. In all cases, predicted posttest performance will be greater if Spring scores are used--hence making it less likely that assessment of program impact will identify "significant improvement" if Spring scores are used as the pretest (basis for predicting expected posttest performance). This phenomenon is particularly noticeable where test levels change from Spring to Fall administration.

Not only does Spring vs. Fall pretest administration appear to impact upon whether or not an individual program is identified as having "significant"impact upon student achievement, but there also appears to be serious implications of Spring-Spring versus Fall-Spring testing models upon comparison, or aggregation, of scores obtained via these different testing models. At all grade levels a program using Fall-Spring testing will yield a more positive growth factor--whether measured in ADSS, percentile or NCE units--than will the same program using Spring-Spring testing. Attempts to compare or aggregate across programs, districts or states, must, therefore, be cognizant of the particular testing model followed. Comparisons or aggregations across testing models would appear to be inappropriate.

IMPLICATIONS

This investigation provided empirical data related to several educational questions. As mentioned earlier, the RMC Title I evaluation models require that the same level of the test be administered for pre-

and posttesting to minimize errors. This requirement has been criticized by many who would like to cross testing levels as students change grade levels. Since greater Spring-Fall decreases were found when test levels were crossed, this investigation tends to support this RMC model requirement.

In addition, this study provides an estimate of the effects of the "summer drop." It is a widely accepted but little researched belief that students lose some of the growth over the summer that they have achieved during the school year. This study provides some empirical evidence, with respect to reading achievement, to support this belief. The logical implication of this finding would be that a longitudinal examination of student achievement growth over long periods of time may well show that "the whole is less than the sum of its parts," i.e. student growth across a period of years may well be substantially below the cumulative total of yearly achievement gains measured by Fall-Spring achievement testing.