

DOCUMENT RESUME

ED 063 337

TM 001 355

AUTHOR Woodbury, Charles A., Jr.; Jacobson, Milton D.
TITLE Recommendations for the Evaluation of Performance Contracting: An a Posteriori View.
PUB DATE Apr 72
NOTE 4p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS Comparative Analysis; Criterion Referenced Tests; Educational Objectives; Evaluation Techniques; Individualized Instruction; *Performance Contracts; Performance Specifications; *Program Evaluation; *Reading Achievement; *Reading Programs; Student Attitudes; *Student Testing

IDENTIFIERS *Virginia

ABSTRACT

Aspects of design and procedures used in the evaluation of a performance contract are discussed, and the project and findings are briefly summarized. The Virginia performance contract in reading was carried out in seven school districts. Evaluation of the project showed that although instructional changes that utilized desirable individualized approaches were used in the experimental program, the experimental group achieved neither less nor more than the control group on ordinary standardized test measures. However, the experimental group had more positive attitudes toward reading than did the control, and in addition mastered the majority of the instructional objectives prescribed. Eleven observations resulting from the evaluation, presented for the guidance of those who contemplate performance contracting, are presented. (DB)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIG-
INATING IT. POINTS OF VIEW OR OPIN-
IONS STATED DO NOT NECESSARILY
REPRESENT OFFICIAL OFFICE OF EDU-
CATION POSITION OR POLICY.

Charles A. Woodbury, Jr.
Bureau of Educational Research
106 Peabody Hall
University of Virginia
Charlottesville, Virginia 22903

Recommendations for the Evaluation of
Performance Contracting: An a posteriori view

Charles A. Woodbury, Jr.
Milton D. Jacobson

Bureau of Educational Research
University of Virginia

Introduction

Accountability and its handmaiden, performance contracting, are very much on the contemporary academic scene with pro's and con's about performance contracting emanating from all quarters--lay, professional, and union.

The Virginia performance contract in reading is believed to have been the first state-wide undertaking of such and its design for evaluation was beyond that of the seemingly singular achievement gain audit. The authors of this paper were respectively director and principal investigator of the evaluation phase of this project during the academic year 1970-1971.

Although the purpose of this paper is to consider aspects of design and procedures used in the evaluation of a performance contract, the project and findings are first briefly summarized.

Project and Findings

Each of seven local Virginia school districts contracted with a firm to establish learning centers and conduct a reading program, guaranteeing basically a 1.7 grade equivalent gain in reading achievement as measured by nationally standardized tests.

The State Department of Education independently contracted with the Bureau of Educational Research to serve as the outside evaluator, performing the following functions: 1) select the tests for; 2) pre- and post-testing to determine G.E. gains; 3) translate and report gains in monetary payoff figures; 4) verify teachers' certification of pupil mastery of interim performance objectives; 5) ascertain participants' attitudes toward performance contracting; 6) assess students' affective response to their special reading experience; and 7) compare gains and affective responses of project and control pupils.

Although massive instructional changes utilizing desirable individualized approaches were installed in the experimental program, the experimental group did not achieve any less nor more than the control group on ordinary standardized test measures. In no school

ED 063337

TM 001 355

or district, experimental or control, was an average G.E. gain of 1.7 realized. The gains ranged from .1 to .8 prorated for attendance. However, the experimental group of students had more positive attitudes toward reading than did the controls (.001%) and in addition mastered the majority of the instructional objectives prescribed for each student.

Although the task of the evaluator is evaluation there are certain observations evolving from the evaluative procedures and findings which should be reported for the future guidance of all who would undertake performance contracting. These observations by the Bureau of Educational Research follow:

1. It is considered essential to the conduct of performance contracting that the evaluator be involved at the very beginning in all planning and decisions, contractual or otherwise: to select the target population, to adjudicate the validity of the instructional program, to establish on some empirical basis standards of achievement in relation to population characteristics, and to select valid tests for measuring learning and program achievement.
2. The content analysis of the instructor's program objectives against the standardized test items used in the Virginia project indicated inadequate content validity of the tests to assess word-analysis skill and some of the study skill objectives in the contractor's program. One way to overcome this situation would be to administer fewer tests and to maximize the use of available testing time by administering a standardized reading test covering a variety of skills, including word analysis in depth, word meaning, comprehension, and study skills.
3. Norm standards such as projected grade equivalent gains should not only be adjusted to instructional periods completed or available but adjusted to accommodate the standard error of measurement present for all populations on all tests. This latter adjustment suggests a band of acceptable G.E. gain around the gain standard rather than the single G.E. score gain standard.
4. If there are to be different G.E. gain standards established for different pupil I.Q.'s, it is suggested that use be made of some version of I.Q. achievement expectancy tables rather than establishing different standards based on the I.Q. break difference of only one I.Q. point, e.g. 0.4 standard G.E. gain for pupils with I.Q.'s 74 and below and 1.7 standard G.E. gain for pupils with I.Q.'s 75 and above.
5. The unfortunate necessity to maintain test security in performance contracting intrudes upon ideal testing procedures whereby the pupils' teachers administer the tests. Short of some procedure requiring program teachers to sign a pledge not to teach the tests used to assess gains, it is recommended that--

- a) trained administrators be selected by the evaluation contractor and approved by the instructional contractor to administer the pre-tests.
 - b) experienced test monitors be appointed by the local school system and approved by the evaluation and instructional contractors to be present at each and every pre-test administration and submit a report on the testing to each and every participating party in the project.
 - c) either the trained administrator-school system monitor arrangement be retained for post-testing or the program teacher-evaluation contractor monitor team be employed inasmuch as post-test security is not necessary, though valid post-test conditions need to be insured.
6. The time when the post-test should be administered could be a factor in affecting the pupils' performances. From the instructional contractor's point of view it is ideal to test close to the very end of the program so that the pupils have maximum instructional time. On the other hand, post-testing the last week of school seems undesirable from the point of view of pupil motivation and possibly discipline. A suggested compromise would be to test one month prior to the close of school assuming program and school termination are on or about the same date. It would be possible to extrapolate on project gains based on actual gains to encompass the full term of the program.
 7. In addition to conducting performance contracting with pupils assessed to be below grade level in some subject, it would be interesting and informative to have some pupils who are at or above grade level in the subject experience the performance contracting program.
 8. Gear-up time should be such and data should be available or secured in order to closely match experimental and control pupils on all pertinent variables prior to pre-testing.
 9. If an instructional contractor's program includes interim assessment of achievement between pre- and post-testing, and it should, then the requirements in terms of sampling, personnel and logistics should be carefully studied during the planning stages, especially when the program is highly individualized and the program centers are geographically dispersed.

10. To account for the variables of program content and sequencing of skills in a given subject, it is recommended that at least two program by two different instructional contractors be contracted for an experiment in performance contracting. Performance contracting requires a common variable of intensive instruction, but the effect of this intensive instruction might well vary by program content and sequence.
11. If standardized tests with their G.E. norms are limited and open to criticism as the evaluation tool and standard for performance contract payoffs, then one alternative is the use of criterion-referenced tasks to measure pupil achievement of objectives. The local agency could develop the subject objectives, the instructional contractor agree to them, the evaluator assess objective-task validity and utilize entry to exit increments with a locally referenced standard to determine contract payoff.

It is generally recognized that curriculum disparities handicap the success of national standardized tests to measure specific instructional objectives and outcomes in most, if not all, subject areas. Generally, achievement test batteries intended for national use endeavor to strike a compromise in terms of coverage and grade placement of cognitive content.

Another limitation of national achievement tests, in the opinion of many, is the fact that single item performance as a mastery measure is lost by summation of such performances to obtain subtest or total scores for normative developmental interpretation.

The investigators suggest that local school systems undertaking performance contracting develop their cognitive objectives, since objectives are non-pluralistic, and then develop or seek help to develop criterion-referenced exercises to measure attainment of these objectives.

Such is not an easy task to accomplish now. It requires local fiscal and personnel resources, expertise and cooperation in curriculum and evaluation design. One thing seems certain: when the demand for criterion-referenced tests approaches the present request for normative-referenced tests, then agencies external to the local school system will increasingly provide the needed assistance and product. It has already begun. However accomplished, this type of evaluation is essential to a valid diagnosis of curriculum success and failure, and needed change.