DOCUMENT RESUME

ED 167 636 TM 008 651

AUTHOR Horst, Donald P.; Wood, Christine T.
TITLE Collecting Achievement Test Data. ESEA Title I Evaluation and Reporting System. Technical Paper No. 13.
INSTITUTION RMC Research Corp., Mountain View, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Oct 76
NOTE 12p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; Data Collection; Elementary Secondary Education; Program Evaluation; *Records (Forms); *Scoring; *Student Testing; *Testing; Test Results

ABSTRACT

There are four steps in obtaining test data, each requiring planning and decision-making: assembling students, administering tests, scoring, and recording scores. In assembling students, two considerations are in order. First, for norm referenced tests, students in the treatment group should take the test at the same time of the year as the students in the norming sample. Second, high absenteeism, differences among test administrators, and in testing environments may significantly affect pretest-posttest scores. Publisher's manuals should be followed exactly to ensure consistency between treatment and comparison groups. From a research viewpoint, trained test administrators should be used rather than classroom teachers. Scoring decisions include selecting an answer form (machine-scorable booklets vs. answer sheets); selecting a scoring agent (school personnel, test publisher, or independent test scoring company); and costs of scoring and of extensive statistical analysis. Finally, when recording scores, data forms should be carefully proofread. All scores should be completely identified and arranged to facilitate analysis. Eleven suggestions for designing test results forms discuss page numbers, testing date, group identification, confidentiality, test name, use of one sheet, arrangement and format of names, identification numbers, number of entries, and arrangement of columns. (CP)

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

RMC Research Corporation
Mountain View, California

Once an evaluation design and an appropriate achievement test are chosen, the most crucial step in the evaluation process is the collection of accurate, complete data. Analysis of the data may be a more technically complex step, but at least, when analysis errors are discovered, they can usually be corrected. On the other hand, if data are distorted or missing, no amount of analysis can adequately correct the problem. If there are too many flaws in the raw data, the entire evaluation becomes meaningless.

There are four steps in obtaining test data, each requiring planning and decisions: (a) assembling the students, (b) administering the tests, (c) scoring the tests, and (d) recording the scores.

ASSEMBLING STUDENTS FOR TESTING

This step, often passed over lightly, is important for two reasons. First, of course, the time of day and the place where students are assembled may affect test scores. The date of testing may also be important. In the norm-referenced model, for example, it is critical that students in the treatment group take the test at the same time of year as the students in the norming sample. Second, any changes in the way the test is given that are made between the pretest and posttest may significantly affect test scores. A change such as testing students in their classrooms rather than in a large assembly hall may or may not make a difference in scores, but the only way to be safe is to use exactly the same procedures each time. Changing from independently administered pretests to posttests administered by classroom teachers because the money ran out--or vice versa because money was left over--is an example of a practice which

should be carefully avoided. Careful planning could avoid all such problems.

It is difficult to generalize about rules for assembling students because of the wide differences among schools. Most important is to minimize the disruption to the students while ensuring that all treatment and comparison students can take both pre- and posttests under similar testing conditions. The major problems in achieving this goal are high absenteeism, differences among test administrators, and differences in testing environments. Where the evaluation simply involves testing project students in their regular project setting, few problems should be encountered. On the other hand, the situation may be more complicated if control students are involved, or if students are to be tested before the project begins or after it ends. Under these circumstances, it is well worth the effort to lay out in detail the number of different tests or test levels to be used, the number of test locations, the time for each test, the number of make-up sessions, the number of special test administrators or supervisors, and so on. Testing often turns out to be a bigger project than anticipated, and, if resources are limited, it is better to simplify both the pretest and posttest rather than to expend so much effort on the pretest that posttesting cannot be accomplished in an adequate fashion.

ADMINISTERING THE TESTS

Ensuring Consistency

It goes without saying that test administration should be orderly, and that cheating and other irregularities are not permissible. But orderliness is not enough. For the purposes of

evaluation it is necessary to have consistency. There are two kinds of consistency to worry about, depending on whether a norm-referenced or comparison-group evaluation design is used. If a norm-referenced design is used, the critical thing is to be sure that the test publisher's procedures are followed exactly. This specifically includes reading instructions, answering questions, doing practice problems, and timing each section.

When a comparison group is used, it is still advisable to follow the publisher's instructions to the letter in order to make supplementary norm-referenced comparisons possible. The most critical thing, however, is to maintain close similarity between treatment- and comparison-group testing situations. The simplest way to ensure comparable situations is to test both treatment and comparison students as a single group. Usually, however, in either norm-referenced or comparison-group designs it will be necessary to test several groups. Then special steps must be taken to make sure that they are tested under as similar conditions as possible. Even here, there are many possible problems; for example, bringing comparison-group pupils into an unfamiliar project lab for testing may put them at a disadvantage.

Training Test Administrators

There are two basic ways of making test administrations comparable. One is to use a few trained test administrators to test all the groups. The other is to train the regular teachers to give the tests to their own students. The latter alternative is much less desirable from a research viewpoint. If teachers must be used, it is advisable to have them test each others' classes to minimize possible biases.

Simply telling teachers or other test administrators to look over the test manual is never

adequate. Each test administrator should be impressed with the importance of following procedures exactly. Each one should at least have "walked through" the entire process, from handing out pencils to collecting the tests, before ever administering the test in an evaluation. Where teacher judgments are involved in scoring student responses (as in oral reading tests), much more training is required.

SCORING THE TESTS

The most important scoring requirement is accuracy, but there are trade-offs of time and money to consider. The important variables are what type of answer form to use and who does the scoring.

Selecting an Answer Form

Most of the major tests can be purchased with machine-scorable booklets or separate answer sheets. Some non-standardized tests may be available only in hand-scored versions. The main factor in choosing among answer forms is the age of the students. Separate answer sheets are usually much easier to process, but young children tend to score lower on these forms, presumably because the forms are confusing to them. In general, separate answer sheets are suitable for above-average fourth graders and all older students. Younger children should use machine-scorable or hand-scored booklets (Harcourt Brace Jovanovich, Inc., 1973).

Selecting a Scoring Procedure

Whichever type of form is used, there are three basic ways of having the test scored. Scoring can be done by: (a) local school personnel,

(b) the publisher of the test, or (c) an independent test scoring company. A choice between the test publisher or an independent company will depend on the local situation and the test that is chosen. Cost, turnaround time, and types and quality of service may vary. Shopping around is in order. The major decision, however, is whether to have the scoring done by either type of service or simply by available school personnel. The major advantages of a good scoring service are accuracy and the variety of analyses provided by computer processing. The major disadvantages are cost, care necessary in preparing the answer forms, and turnaround time. It is also possible that forms may be lost in shipping, or that mishandling or faulty equipment will result in scoring errors. There is little recourse when forms are lost, but spot checks on scoring accuracy should be made after answer forms have been returned.

Cost Considerations

"Ballpark" cost figures for machine-scored forms (taken from one widely used publisher's service) range from $.30 to $.70 per pupil, depending on the type of form and length of the test battery. Hand-scored booklets cost three or four times as much to score, although a lower original purchase price will offset this difference slightly. Clearly, local personnel can do the basic scoring at lower cost, but included in a scoring service's price are a number of features that are costly, time consuming, and prone to error when scoring is done by hand. These include: (a) reports with convenient formats in triplicate for each group (e.g., class), completely identified as to test, date, group, etc.; (b) raw scores, percentile scores (local or national distributions), standard scores, and, in some instances, NCEs for each student on each subtest; (c) mean standard scores

for each group. Several other analyses are available for prices ranging from an additional $.05 to $.12 per student for each analysis. These include score distributions for each class, item analyses, and individual student profiles. Additional statistical analyses are readily available, or, for schools with access to their own computer facilities, the scores are available from the publisher on computer cards or tape.

In short, for very small tryouts with simple analyses it may be desirable to do the entire job locally. Unless local computer facilities are available, however, more extensive evaluations may well be completed more accurately, thoroughly, and economically with the help of a scoring service. All the major services have literature and consultants to provide details and to assist in planning the scoring and analysis.

RECORDING THE SCORES

Recording the scores is the final step in the data collection process, but to ensure that the scores will be usable, the details of recording should be worked out well before pretest time. If you use a commercial scoring service, you may have little control over the recording process. If you decide to do your own scoring, or if you want to transfer scores from computer printouts to a more convenient form, you must consider two important issues: accuracy of the data, and details of the data-recording forms.

Copying scores accurately onto data forms is not a complicated problem for small-scale local studies, but the possibility of errors must not be overlooked. Even the most conscientious recorders make errors. All data forms should be carefully proofread, preferably with one person reading aloud while a second person checks the scores.

The details of the data forms might appear to be of little importance, but, in many school districts, the way in which data have been recorded virtually precludes any reasonable analyses. Two general principles must be observed when deciding upon a standard data format. First, all scores must be completely identified, and second, scores must be arranged in a way that facilitates analysis.

Considerations for Data-Recording Forms

The following considerations should be incorporated into any data-recording form and are illustrated by the building level worksheet forms that accompany each model.

- Page numbers. Most sets of scores require more than one page. A page number at the top should identify each sheet and the "number of pages" helps make sure no pages are missing.
- Testing dates. Test dates are critical, especially in norm-referenced evaluations. Record the date of the original testing and make-up testing sessions for both the pretest and posttest.
- Group identification. Identify clearly the group for which data are recorded near the top of the page to simplify the retrieval of that group's data from a large data base.
- Provision for anonymity. Arrange the page so that it can be photocopied without the student's name. This permits possible later use of the data for research purposes without compromising student privacy.
- Test name. It simplifies analysis greatly to have only one test (pre and post) recorded on each sheet, provided the rules for listing

students suggested below are followed. List the complete name of the pretest and posttest (taken exactly from the test booklets and including publication date).

- Single pre/post data sheet. Identifying students and organizing their names efficiently are the most difficult problems in recording student data. Where evaluations are only for one year and are based on fall and spring testing, the problems can be solved with a little effort and care. But where students must be followed over several years, the problems are more difficult since students come and go from projects, and groups are reorganized every year. The simplest rule is to make sure that the posttest scores are all entered on the same sheet of paper as the corresponding pretest scores. This at least eliminates the problem of trying to find each student's name on two lists.

- Standard order of names. A second rule for listing student names is to establish a standard ordering of the names, and stick to it for the life of the evaluation and for all tests that are used. If a student moves or fails to take some of the tests, then the appropriate entries are blank, but he should not be eliminated from the list. If new students enter the program, their names should be added to the end of the lists for all tests, even those for which no data will be entered. Besides a reduction in confusion, there are some practical advantages to this procedure. For example, a master form can be prepared with only the students' names and identification numbers filled in, and the forms can simply be duplicated when new tests are given. It also makes comparisons or correlations between any two sets of scores relatively easy because any two forms can be laid

side by side and the corresponding names will line up correctly. If there is a compelling reason to change the order of student names in the middle of a project, then either change all forms, or maintain a double set of forms (old and new order).

- Standard form for names. Establish a rule for recording names. "Caldwell, D.E." should never become "Danny Caldwell" on a second list. The simplest procedure is to allow plenty of space and to spell out first names and include middle initials (e.g., Caldwell, Daniel E.).

- I.D. numbers. If I.D. numbers are used, each student should have an I.D. number that identifies him completely. For example, different digits might identify the student either as a member of the project group or a control group, indicate class or sex, and of course, represent the individual student. In some evaluations, other codes (including letters) can be used, but careful planning is necessary in order to permit any desired grouping simply by I.D. number.

- Uniform number of entries. A page should have some reasonable number of entries (e.g., 20, 25, 30), and the number should not vary from page to page.

- Pre/post score columns. Keep pre- and posttest scores in adjacent columns. For example, enter the raw scores for pretest and posttest in two columns, percentile scores for each in the next two columns, etc., instead of pairing each pretest raw score with its standard score, percentile score, etc., followed by each posttest score and its transformations. This greatly simplifies the mechanics of analysis; comparisons are nearly always made between pre- and posttest scores of the same type.