DOCUMENT RESUME

ED 053 181 TM 000 711

AUTHOR	Schriber, Peter E.
TITLE	Cost Benefit Analysis of Comprehensive Achievement Monitoring for Classroom Evaluation.
INSTITUTION	Massachusetts Univ., Amherst. School of Education.
SPONS AGENCY	Charles F. Kettering Foundation, Dayton, Ohio.
REPORT NO	WP-18
PUB DATE	Feb 71
NOTE	11p.; Paper presented at the Annual Meeting of the National Council of Measurement of Education, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement, Behavioral Objectives, *Classroom Research, Comparative Analysis, *Cost Effectiveness, Course Evaluation, *Criterion Referenced Tests, Data Analysis, Decision Making, *Evaluation, Evaluation Techniques, Performance, Teacher Education, Teacher Role, *Test Construction, Testing, Test Results, Tests
IDENTIFIERS	CAM, *Comprehensive Achievement Monitoring

ABSTRACT

Comprehensive Achievement Monitoring (CAM) is a systematic procedure of constructing and administering longitudinal, criterion-referenced tests. CAM has advantages over typical classroom testing of having well-organized tests, providing course evaluation through pretesting, posttesting, and retention measurement; producing data for continuous classroom instructional management; and being based on a curriculum of behavioral objectives. The systematic gathering of comprehensive performance data permits a means of establishing a dollar-and-cents cost analysis for various instructional and curricular alternatives. The costs of typical classroom testing and CAM are compared. (Author)

Project Comprehensive

Achievement

Monitoring

Working Paper No. WP-18 February 1971

COST BENEFIT ANALYSIS OF COMPREHENSIVE ACHIEVEMENT MONITORING FOR CLASSROOM EVALUATION

by

Peter E. Schriber
The University of Massachusetts

(A paper presented at the Annual Meeting of the National Council of Measurement of Education, New York, February, 1971.)

The research and development reported herein was performed pursuant to a grant from the Charles F. Kettering Foundation to the principal investigator, Dwight W. Allen, and under the project director, William P. Gorth, both of the School of Education, The University of Massachusetts, Amherst, Massachusetts 01002.

William P. Gorth
Director, Project CAM
School of Education
The University of Massachusetts
Amherst, Massachusetts 01002

Cost Benefit Analysis of Comprehensive Achievement Monitoring for Classroom Evaluation

by

Peter E. Schriber
University of Massachusetts

I. What Are Evaluation Costs?

The single major category of classroom evaluation costs is professional time. Thus, the cost efficiency of an evaluation procedure is dependent on the value of the hourly productivity of the professional engaged in the various phases of the procedure. The value of this productivity is in large measure dependent on the training of the teachers in evaluation techniques and the teacher's conscientious application of the techniques. The basic components of the teacher's role in evaluation include at least the following:

a. formulation of an evaluation plan for purposes of a check on student learning and the making of instructional decisions;
b. devising and selecting of tests as components of the plan (including teacher-made and local or commercial standardized tests);
c. administration of tests;
d. correction of tests;
e. interpretation and evaluation of test results;
f. use of results to assess student learning and as bases for instructional decision making.

A second category of evaluation costs is clerical time. This includes preparation of test masters and reproduction of test forms. Also, test correction is often done by a person in a clerical position rather than by the teacher.

A third category is that of materials needed. Typically this includes paper, reproduction equipment and, atypically, machine replacements for professional and clerical time such as mark-sense test correction equipment and electronic computers. It also includes the procurement of commercially produced tests.

However, it is of little value to delineate the above cost components and tally their dollar and cents figures for any specific situation for comparison purposes unless two more components necessary for evaluation are established. Referring back to the components of the teacher's role in evaluation, let us first consider the devising and selecting of tests. It is imperative to know the substance of the tests. What does each test purport to measure? The purpose of this question is, succintly, something must be specified in behavioral terms before we can devise an implement to assess learning of the behavior. In other words, at least a large proportion of a course must be based on a performance curricula, i.e., be composed of behavioral objectives, before the measurement of learning can be accomplished by testing procedures.

A second major consideration for an effective evaluative process is a clearcut systematic approach for using test results. If testing is merely to rank students or perhaps in addition, to identify gaps in the recent learning of individuals then a narrow interpretation is made of two more of the components of the teacher's evaluative role, specifically (e) the interpretation and evaluation of test results and (f) the use of results to assess student learning and as bases for decisions for future instruction.

If the evaluation procedure is to provide data for instructional decision making based on other than individual diagnostic data as reflected on a single classroom posttest, then new techniques for compilations of test results are necessary. Data will have to be organized into categories which permit a comprehensive view of student performance with a focus on both student learning and course effectiveness. Here the computer can be of assistance in compiling data into tables which can crosstabulate by student, student group, objective, objective set, item, and other categories deemed pertinent for instructional decision-making.

Thusfar, my purpose in this paper has been to focus on the point that regardless of the dollar and cents figure placed on the evaluation procedure, one must first be certain of what is being bought before comparison can be made to other means of evaluation. More specifically, when the expressed purpose of an evaluation procedure is to assess learning and furnish bases for instructional decision making, the following question must be posed: How completely and to what degree does the procedure accomplish these results?

Before continuing with a comparison of usual classroom testing and Comprehensive Achievement Monitoring (CAM), a word should be said about the use of the computer in making the establishment of performance curricula and test forms less of a chore than it currently is. The use of a computerized bank of objectives and items available to teachers and schools would greatly facilitate the use of behaviorally written curricula. It is even possible at present to use such a bank to compose and reproduce test forms by computer thus curtailing much of the teacher and clerical time formerly necessary for test creation.

II. Comparison of Usual Classroom Testing with Comprehensive Achievement Monitoring (CAM)

A basic premise for this comparison is that achievement testing is based on a performance curriculum and thus a behavior can be observed by some means in the classroom on which to base the assessment of learning of the curricular objectives. The following discussion is a point-by-point comparison by topic considering first the situation in usual classroom testing followed by the situation in CAM.

Evaluation Design

Usual classroom testing lacks a specified design other than the use of posttests for various course segments and a final test along with short quizes for further diagnostic aid. CAM has a specified organization of predetermined test administrations, test form design, and student grouping for test taking all tailored to the teacher and course.

Test Construction

Typically usual classroom testing has loosely constructed tests. The tests are usually not representative of more than a week or two of a course, a conspicuous lack of means of producing trend data is evident, and little emphasis is placed on pretesting or retention measuring.

CAM has well-organized tests which are course representative or representative to the segment of the course covered by the specific CAM system, e.g., a series of CAM tests to be used biweekly for a semester would have test forms representative of that semester-length segment of the course.

The longitudinal testing sequence and the representative nature of the tests provide trend results for individuals and groups for single objectives or selected sets of objectives.

Teacher Training in Evaluative Techniques

The training that the usual teacher has for usual classroom testing often amounts to one college-level evaluation course which did not emphasize data analysis for instructional decisions. CAM requires a short training program of a few days duration for teachers in the areas of assigning items to tests, the construction of student test groups, the interpretation and use of computer data analyses, and the means of introducing students to CAM testing.

Classroom Management

Usual classroom testing affords decision-making potential for a very few objectives at a given point and based solely on immediate posttest data. Class average results are limited only to those items in the short term posttest results. It is rare that even this minimal compilation of data is done by teachers for use in course and curriculum modification. Management functions are usually focused only on diagnostic results for individual students with review and reteaching the only usual instruction decisions relating to change.

CAM presents a vastly larger array of data of not only immediate posttest results but pretest and retention results spanning the entire course segment over which the particular series of CAM tests is being used. The data are organized and compiled by computer and are available on a continuous basis. All the objectives of the CAM test series can be compared for any desired group of students and immediate decisions can be made as to the instructional treatment needed, which objectives are suitable to which students, the pace of learning, the course sequence for the forseeable week or two to come, and the appropriateness of objectives.

Test Analysis

Usual classroom testing rarely includes any type of item analysis or analysis of the test itself regarding its representativeness for the period it is testing. The letter is particularly true of midterm and final tests which contain only a small sampling of course material and a one-shot evaluation. measure.

CAM contains a self-updating component. An item analysis is done on all items of all tests for three time phases: pre-instruction, post instruction, and retention periods. Using the item data individual test forms can be revised and made more comparable.

Summary

The usual classroom test is generally unsystematically constructed and not systematically related to any other test used in the course with the possible exceptions of midterm and final tests. It is focused on immediate post-test data almost exclusively and thus its chief utility is in permitting instruction decisions related to individual student diagnoses based on the course material as reflected, however accurately, on a single test. No systematic updating procedure for the tests such as item analyses are generally employed.

CAM tests are systematically constructed and furnish trend data which expand the range of instructional decision making by means of an encompassing picture of student and course progress composed from the continuous pretest, posttest, and retention measuring. A three-phase item analysis provides a systematic means of test modification.

III. Cost Comparison of Usual Classroom Testing and Comprehensive Achievement Monitoring

When teachers and school officials consider evaluation and evaluation costs they generally perceive only the top of the iceberg. Costs are viewed as centering on test construction, loss of teaching time due to test giving, and test correction and the recording of results. These cost items are not necessarily related to the value of an evaluation but to time management functions in the production of testing implements and the collection of raw test data. Furthermore, the emphasis on how the test is constructed, the question of how much teaching time is wasted due to little feedback for performance assessment, the comprehensiveness of data analyses, and the use of test results for instructional modification are not usually perceived by them as cost considerations.

Generally a teacher needs to obtain answers from classroom evaluation to the following questions:

1. how are my students progressing?
2. how is my course progressing?
3. what do I use to decide where the course goes from here?

Usual classroom testing can supply some data to answer question 1 but it is bits and pieces because its data are piecemeal from discrete testings with prime emphasis only on immediate present posttest data. CAM also furnishes data to answer question 1 but its greatest value and emphasis is in providing data to base answers to questions 2 and 3. Such data are lacking from usual classroom testing.

CAM is a procedure for course evaluation. CAM data compilations, summaries, and analyses allow bases for decisions for answering the following questions about program effectiveness.

1. Are students learning the objectives of the course?

   Pertinent decisions involve the identification of

   (a) poor objectives and whether the objectives need to be omitted, revised, or expanded into several new objectives,

   (b) need for review and whether review was successful,

   (c) program modification involving readjustment of the pace of teaching, modification of curricular emphases, and reordering of objectives.

2. Is each major course segment well-structured?

   Decisions relate to

   (a) coherence and logicalness of course and curriculum,

   (b) objective interactions, i.e., teaching X influenced learning of Y, thus the identification of links between objectives and between course segments which went undetected when the curriculum and course were first organized,

   (c) assessment of instructional mode for particular objectives and student groups.

3. What factors relate to student mastery and to student failure in learning of specific objectives and sets of related objectives?

   Decision bases that can be made from data include

   (a) identification of students with similar learning deficiencies and successes,

   (b) identification of late learning and whether it is related to concurrent teaching,

   (c) pinpointing of incipient forgetting and possible precipitating factors,

   (d) assessing learning gaps and providing alternative approaches to meet specific deficiencies

With the basic outline of the evaluative potential of usual classroom and CAM testing a comparison can now be made in terms of cost differential. We begin with reiterating the first premise for comparison, that for either to have value a performance curriculum is mandatory. In summary form a tally of cost differential on evaluative components follows.

	Cost Differential
Test Items Usual classroom testing (UCT) requires test items for all objectives and since items are rarely reused next year new items will have to be made. CAM requires the same except that items can be reused with the aid of item analyses.	CAM is more efficient and actually requires less time for creation of test items.
Test Construction UCT requires making tests using all the items which are generally used for post-test purposes only and at one time only. Tests are often not reused the following year. CAM uses all items for pretest, posttest, and retention measuring and reuses the items throughout the course. Test forms are systematically revised for the following year. Tests are created before instruction in order to monitor the prescribed program.	Initially effort and time are very comparable but the following year CAM will require updating only.
Test Reproduction UCT requires tests to be reproduced intermittently. CAM has a set time for reproduction and allows a planned usage of clerical aid plus the additional aid of reusable test forms requiring smaller quantities of tests to cover the same population than with UCT.	The time differential is comparable with the added factor of reusable tests making CAM more economical.

Test Administration UCT has intermittent testing. CAM uses pre-set interval points.	Cost Differential Actual test taking time is less per test with CAM due to representative scope. However, real time may be more due to UCT possibly having fewer administrations.
Analysis of Test Results UCT requires little teacher time and training. CAM requires a short training program and several hours per month for analysis of data.	CAM uses more time but provides more usable data for instructional decision making.
Item Analysis This is usually absent in UCT. CAM provides a three-phase analysis which requires a day or two per semester for professional analysis of results for identifying items and objectives in need of modification.	Initially CAM requires several days yearly of teacher or other professional time for item analyses and less for future item analyses.
Data Analysis UCT can have the added cost of standardized testing which usually covers language arts and quantitative skills with the cost at $1 or more annually per student. Such tests provide virtually no data usable by the teacher for instructional management. CAM computer processing costs $1 to $2 per student per subject area depending on the CAM system design, but the data analysis provides bases for stringent course evaluation.	CAM delivers usable data for its processing costs. UCT often provides little data usable for instructional management.

In arriving at a final cost comparison CAM does require more professional time for evaluation purposes than usual classroom testing, but this is not time spent on test construction, correction, and administration, but on making

fruitful use of test results for immediate and long term instructional management. This option is not possible to any significant degree without longitudinal and course representative testing. The actual monetary cost over usual classroom testing for the categories of professional time and computerized data analyses is only slightly more and is tailored to the course of study. However, if consideration is made of the money and teacher and student time wasted because the evaluation system cannot identify a poor course and weak curriculum areas, then CAM is decidedly a savings.

References

1. Gorth, W.P.; Dumont, F. & Wightman, L. Improving educational quality through Comprehensive Achievement Monitoring. Technical Memorandum No. TM-23. School of Education, The University of Massachusetts, Amherst, 1969, Grant No. 642, C. F. Kettering Foundation.

2. Gorth, W.P. & Grayson, A. A program to compose and print tests for instructional testing using item sampling. Educational and Psychological Measurement, 1969, 29, 173-174.

3. Gorth, W.P. & Grayson, A. & Lindeman, R. A computer program to evaluate item performance by internal and external criteria in a longitudinal testing program using item sampling. Educational and Psychological Measurement, 1969, 29, 181-183.

4. Gorth, W.P., Grayson, A. & Stroud, T. A computer program to tabulate a plot achievement profiles of longitudinal achievement testing using item sampling. Educational and Psychological Measurement, 1969, 29, 179-180.

5. Gorth, W.P., Wightman, L., O'Reilly, R.P., & Schriber, P.E. CAM and the computer help teachers improve evaluation. Trend, 1970, Spring, 5-6.

6. Pinsky, Paul D. Feedback information in the comprehensive monitoring of educational achievement. Working Paper No. WP-16. A paper presented at the 11th Annual Meeting of the Institute of Management Sciences, Los Angeles, California, October, 1970, Grant No. 642, C. F. Kettering Foundation.

7. Wightman, L. & Gorth, W.P. CAM: the new look in classroom testing. Trend, 1969, Spring, 56-57.