DOCUMENT RESUME

ED 037 789 CG 005 195

AUTHOR Baker, Eva L.
TITLE Experimental Assessment of the Effects of the Probe System.
INSTITUTION American Educational Research Association, Washington, D.C.; California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE 2 Mar 70
NOTE 14p.; Paper presented at American Educational Research Association Convention, Minneapolis, Minnesota, March 2-6, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$0.80
DESCRIPTORS Academic Performance, *Behavioral Objectives, *Educational Programs, Evaluation, *Inservice Programs, Inservice Teacher Education, *Predictive Ability (Testing), *Student Characteristics, Student Evaluation, Workshops

ABSTRACT

This paper is concerned with objectives-based evaluation and alternative ways in which a system of objectives and test items might contribute to school programs. In the first study, teachers, parents, and students were involved in the needs assessment phase of educational evaluation with the use of behavioral objectives. All three were first asked to rate the importance of each objective for their school situation. Each group was then asked questions pertaining to these objectives. Among the results was a tendency for both parents and students to mis-predict pupil achievement. Teachers made relatively good predictions. The purpose of the second study was to compare the performance of learners taught by teachers trained or not trained by a three day PROBE institute in the use of behavioral objectives. A three day workshop was held for 27 fourth grade social science teachers. To assess the effects of the workshop, a performance test was used where six objectives assembled for fourth grade social science were employed. No significant differences were found between students of the trained and untrained teachers. (KJ)

A Symposium Presentation at the
American Educational Research Association Annual Meeting
Minneapolis, Minnesota, March 2-6, 1970

EXPERIMENTAL ASSESSMENT OF THE EFFECTS OF THE PROBE SYSTEM[1]

Eva L. Baker*
Center for the Study of Evaluation
University of California, Los Angeles

It takes time, lack of reinforcement, and more time to eradicate propositions of the heart. One such, now undergoing extinction by all but zealots, is that the act of providing objectives and coordinate test items to teachers in itself will modify the nature of educational practice. The experience of the regional laboratories and our own anectodal evidence urge the production of a fairly elaborate support system if we wish any innovation, in this case, objectives-based evaluation, to be effectively installed. The research in implementation problems which shall be described represents a preliminary attempt to determine the requirements of such support and to explore alternative ways in which a system of objectives and test items might contribute to school programs.

STUDY ONE: COMMUNITY EVALUATION

While the notion of evaluation almost always implies assessment following some segment of instructional program, the utility of objectives-based evaluation in a needs assessment function was explored in a predominately black junior high school. The intent of the project was to involve parents, teachers and students in identifying objectives of common and discrepant interest and to aid in the school's planning of instructional programs to facilitate achievement of target goals. The procedure actively sought community input but the questions raised were those of values, i.e., what should the goals of the schools be, rather than of means, e.g., how many minority teachers should we have. Objectives from the most complete Collection in the PROBE files were used. One advance limitation was that the subject matter of the Collection was mathematics. But the study has functioned as a procedural prototype for future investigations.

Overview

Teachers, parents and students were involved in the needs assessment phase

---

*Administration of the studies and data analyses were supervised by Ted Dahl of the staff of the Center for the Study of Evaluation.

[1]The research and development reported herein was performed at the Center for the Study of Evaluation, UCLA, pursuant to a contract with the U.S. Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program.

of educational evaluation with the use of behavioral objectives. All three groups were first asked to rate the importance of each objective for their school situation. Teachers also indicated if objectives were among those they ordinarily taught and to estimate their classes' level of performance on the objective. Parents were asked to indicate whether they felt their child could currently master each objective. Learners were asked to predict their own performance on the objectives. To aid in respondents' understanding, each objective was clarified by an example of a test item which would measure it. Students were then tested on the objectives to determine their actual performance.

Procedure

Teacher Data. Ten teachers, each instructing two classes of seventh grade mathematics, were provided with 43 objectives and sample test items taken from the PROBE collection for grades 6, 7, and 8. Teachers rated objectives on a five point scale (1=low, 5=high) in terms of importance, estimated the percentage of pupils who could achieve the objective and indicated if the objective was normally taught in their classes. Of the 43 objectives presented to the teachers, 15 had been previously identified for use in the study (See Figure 1). These dealt with important arithmetic operations, scientific notation, measurement and geometry.

Learner Data. In these teachers' classes a total of 634 students were asked to complete a questionnaire containing the target 15 objectives and sample items. Students were to rate each objective in terms of importance and to indicate whether they felt they could solve problems like the one presented in the sample test item. Following administration of the questionnaire, learners were tested on the 15 objectives which they had rated. In order to limit the time necessary for testing, five separate test forms were devised. On each form three of the objectives were intensively sampled with eight items each, while the other 12 objectives were measured with three items each, resulting in a 60 item test.

Parent Data. Questionnaires were mailed to 164 parents of students. All parents in three complete classrooms were sent letters and four parents from each of the other 17 classrooms were sampled at random. Only 123 of the letters were received, as there were 41 letters with inaccurate addresses. Parents were asked to respond to the same 15 objectives and test items, estimating if their child could achieve the objective, and rating importance from 1 to 5. Parents were also asked to indicate if they felt community participation of this type was useful and whether they would be willing to participate in another survey. Parents who did not respond within three weeks of the first mailing received a second letter, and following four weeks a third letter was sent.

Results

The results of this study are mixed, both in terms of utility and valence. Eighty-two parents, or 67 per cent of the correctly addressed letters, responded to the questionnaire, 29 parents after the first mailing, 36 after the second and 17 after the third letter was sent.

Analysis of variance was conducted on the parents' responses to the objectives according to the time of response to the questionnaire (after the first letter, the second letter, or the third letter). Significant differences were found for responses to only two of the 15 objectives,* which were rated considerably lower by parents who responded to the third request.

Parental ratings of the 15 objectives were uniformly high, with a mean across all 15 objectives of 4.3. Students were less sanguine about the objectives, with a mean rating of 3.9. Teachers' ratings averaged 3.9. An analysis of variance of the three groups' ratings for each objective was conducted. Significant differences were found for 11 of the 15 objectives (Numbers 1, and 5-14) and for the average rating across all objectives. Differences in five of those instances were attributable to the high rating which the parents reported.

(Insert Table 1 about here.)

The meanings of these ratings may be related to the specific nature of each objective considered. Objectives most favorably rated by teachers dealt with set equivalences, whole number arithmetic operations, and translations of fractions when shown a pictorial representation. One interpretation of these results is that the teachers prefer practical and basic concepts in arithmetic, but a less optimistic colleague suggested that the teachers might have preferred what they felt was easiest to teach. Parents generally rated objectives in arithmetic operations highest but also favored a word problem task. Students rated objectives of a somewhat esoteric nature more important. Possibly because they were unfamiliar with the content involved, students favored objectives such as the writing of numerals in scientific and expanded notation, finding the circumference of a circle and finding the area of common geometrical figures. Correlations among the ratings are presented in Table 2 where a tendency for parents and students to agree with each other and disagree with the teachers can be observed.

(Insert Table 2 about here.)

Comparisons of the abilities of teachers, parents and students to predict performance with achievement levels produced findings of considerable interest. (See Table 3.) Both students and parents underestimate the competencies of the students, with parents predicting about 24 per cent performance and students predicting 22 per cent performance. Actual mean achievement of learners' achievement on all objectives was 43 per cent. Teachers, on the other hand, were more optimistic about the ability levels of the students, and predicted their pupils' achievement at around 54 per cent. Such over-prediction might deflate the argument that teachers in predominately black schools tend to under-predict their students' ability.

Separate correlation coefficients were computed between (1) each of the three groups' predictions of student achievement and (2) actual student achievement. The mean correlations for each of the 15 objectives were calculated and are presented in Table 4.

---

*Ability to convert decimals written in expanded notation (F=8.09), Ability to solve word problems dealing with multiplication (F=5.76).

(Insert Table 4 about here.)

The relatively high negative correlations with achievement reflect a tendency for both parents and students to mis-predict pupil achievement. The predictions of parents and students show some consistency. Looking at these correlations conjointly with the means of predictions and achievement of the objectives, one might infer that both students and parents are consistently underpredicting their performance. Teachers, on the other hand, make relatively good predictions of their classes' performance levels and disagree with both student and parent estimates of performance. Parents who responded were positive about this type of involvement in school operations. Eighty-four% indicated that they thought the project was a good idea and 83 % expressed willingness to respond to another questionnaire.

Implications

Results of this investigation have prompted the school to seek specific help in areas of deficiency in student performance. Seven of the 15 objectives received an average rating of four or above by at least two of the groups. Of these, four objectives were the lowest in terms of student achievement.

While it is obvious that only limited, substantive applications can be made from objectives in the field of mathematics, where the nature of the subject matter limits curriculum decisions, both the school and staff of the Center were encouraged, not only by the willingness of the parents, teachers and pupils to participate, but more generally with the potential utility of the procedure. Replications in American History and Black Studies are now underway.

An incidental, but possibly important, result is that the personnel of the school itself, at first reserved about the consequences of specific research projects on their daily operations, has reported positive acceptance of this procedure. Teachers were particularly cooperative and did not feel that the research investigation was an artificial interruption in the normal activities of the school.

STUDY TWO: EVALUATION OF TRAINED TEACHERS' EFFECTS ON THE BEHAVIOR CHANGES THEY PRODUCE IN THEIR LEARNERS

A central issue in the development of the PROBE system of objectives and items is the definition of the support requirements necessary to get the procedure into widespread use. The assumption has always been made in PROBE that some training experience for teachers would be provided so they could begin to understand and capitalize on the use of objectives and test items to improve their evaluations. Such an argument assumes that the purpose of the system is ultimately to improve the effects of educational practice rather than to function primarily to describe the status of educational programs.

The purpose of a second study was to compare the performance of learners taught by teachers trained or not trained by a three day PROBE institute in the use of behavioral objectives. Performance tests as dependent measures have been employed before[2] and although the idea of a short training situation might immediately produce teacher behavior changes strong enough to affect pupil achievement was wildly hopeful, we decided to verify the immediate consequences of such training. In addition, we gathered information regarding a number of other procedures relevant to PROBE, including teacher's responses to the items, the objectives and their use of pretest data.

Subjects

Six school districts within easy testing distance of the Center for the Study of Evaluation were contacted weeks prior to the institute and asked if they would submit the names of at least 10 volunteer fourth grade social science teachers to participate in a three day training session on the use of behavioral objectives. Fifty-four teachers volunteered for the session and, blocking by district, twenty-seven were randomly assigned to participate in the training.

Treatment

While a proper training program would focus on instructional methodology, e.g., the uses of iterative testing procedures, and would optimally provide practice for the teachers in these behaviors in a classroom context, the PROBE staff adopted a "lean programming"[3] strategy rather than attempt to develop a total teacher education program. Limitations of district in-service training resources (for example, paying substitute teachers while regular teachers are undergoing training) encouraged the PROBE training institute to focus on the fewest objectives which we thought could possibly do the job.

A three day workshop was conducted at UCLA during the first week in October, 1969. The workshop was planned to toal 18 hours, but the amount of instructional time actually spent was less than 12 hours. An additional purpose for the workshop was the hope that it would represent a first generation attempt at an instructional training package for eventual exportation to either a network of cooperating schools for dissemination purposes or directly to user districts.

The objectives of the institute called for participants, at its conclusion, to be able to:

1. discriminate between statements of behavioral objectives
2. write behavioral objectives
3. write possible entry and en route behaviors for given instructional objectives
4. discriminate between examples of relevant and irrelevant practice for given objectives
5. produce instances of relevant practice for stated objectives
6. generate additional items for objectives when presented with a sample item

---

[2] Markle, Susan M., Good Frames and Bad, Second Edition. John Wiley and Sons, New York, 1969.

[3] Popham, W. James and Baker, Eva L., "Validation Results: A Performance Test of Teaching Proficiency." Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, February 7-10, 1968.

7. prepare lessons which exhibited the following components:
   a. task analysis of objective
   b. relevant practice
   c. iterative testing and remediation cycles.

All participants were given a pretest in which their ability to perform the objectives was assessed. Following seven hours of instruction, they received a criterion check to monitor their progress toward the objectives, and at the completion of the institute a posttest was given. An 80 per cent criterion level was set to indicate mastery of the workshop's objectives. Our training was not terribly effective, since only 50 per cent of the teachers reached this desired criterion level.

Criterion Measure

To assess the effects of the workshop, a performance test was used where six objectives assembled for fourth grade social science were employed. The objectives focused on the translation and interpretation of graphed data, a task not as yet treated in the participating districts' programs.

Before the instruction began, over 1,600 children in all 54 classrooms were pretested on the six objectives and means of their class for each of the six objectives were reported. All teachers received these data, the objectives and sample items ten days prior to the scheduled instructional period. Following a seven day instructional period, where teachers devoted approximately 30 minutes a day to this topic, children were given a 12 item posttest by the Center staff. Test items measuring these geography objectives had been tried out previous on seven fourth grade classes in another community, critiqued by the seven teachers, revised, readministered to six other fourth grade classes, again critiqued and revised, prior to the administration of the pretest in the actual study. Teacher feedback in all cases was directed to the cohesiveness of the unit, item difficulty, reading level, and the extent to which the items were perceived as adequate measures of the objectives. During the posttest, teachers were asked to complete a questionnaire where they rated the utility of the unit, format of the objectives and items, and described the nature of the learning activities which they used.

Analysis and Results

Analysis of covariance was computed for the posttests of fourth grade students, using pretest scores as a covariate. Total posttest means of 54 classrooms were the entries in the analysis, corresponding to the number of teachers involved in the study. No significant differences were obtained. Looking at the experimental group of teachers only, fourteen of twenty-seven reached the 80 per cent criterion level on the training posttest. For this group the "treatment" as it was conceptualized should show its greatest effects. Analyses of covariance, comparing successful and unsuccessful teachers on the institute (N_1=14, N_2=13) posttest, and successful and control teachers (N_1=14, N_2=27) yielded no significant difference on the total student posttest means. When analysis of covariance was conducted for each of the six objectives, a significant difference was found for objective two.

Considering the number of analyses computed, such a finding is best explained as a random event. While performance between treatment groups was remarkably consistent, there were disparate performance levels on each of the six objectives considered. Performance levels of the learners on each of the objectives are presented in Table 5.

(Insert Table 5 about here.)

Performance for both treatment groups was considerably higher on objectives one and two, objectives which measure recall skill rather than any translation or application of information. For the other objectives, class performance was relatively poor, although there was improvement displayed on each objective. Teachers in the trained treatment reported that they spent approximately 185 minutes on instruction, while the comparison group spent about 200 minutes. The trained teachers reported than an average of 85 additional minutes would be necessary to have students reach a satisfactory criterion level, while the untrained teachers estimated that 400 additional minutes, or almost twice again the instructional time originally allocated for the unit, would be necessary.

Inspecting the process data, that is, the number of activities described in the questionnaire, two research assistants independently judged the activities in terms of relevance to the objectives. Teachers in the trained group produced approximately two and one half times as many relevant activities, but because of the immense variation within groups, this difference was not significant. When teachers' responses to the questionnaire were inspected, a correlation of .36 ($n=54$, $p<.01$) was found between whether the teacher considered the materials useful and total student achievement. Differences in attitude toward use of materials based on treatment condition were not found to be significant.

Discussion

There are a number of plausible interpretations of the negative findings on pupil achievement. Optimistically, one might contend that the time allocated in the criterion task, seven days of instruction, was not sufficient for a teacher to institute teaching, testing, and reteaching cycles and thus "improve" his instruction. Another explanation might clearly indict the impotence of the original training time. Twelve hours of instruction might be insufficient to modify substantially pedagogical habits produced by years of teaching. Yet, we did have at least half of the group attain the mastery level we had hoped for in the training and no differences in their pupils' achievement were found. For a moment, it looked as if we had affected their instructional activities in the reports but that was a mirage. Had we conducted the training in a way purely consistent with the approach we advocated, we would have provided reinforced classroom practice in applying the verbal behaviors the teachers were learning. But such a procedure was precluded by the decision to integrate this training program unobtrusively into usual district practices.

To explain the lack of differences found on the teacher attitude measures, we might examine the teachers who were involved. The districts did not mandate attendance, and thus we assigned both training and control treatments to

volunteers. These teachers were thus already somewhat positive toward behavioral objectives, and lack of differences in attitude data might not be difficult to explain. Both groups averaged 2.5 on a three point scale for a question asking if teachers would be willing to use objectives and items in other subject matter areas.

The last alternative is, of course, that we don't know what the critical components of such a training program really are and that our instruction was wholly inadequate, not just in the lack of practice opportunities for the teachers, but in concept. The results of this study, particularly with regard to the teachers who mastered our objectives, indicate that short-term installation programs for disseminating new practices might be viewed more skeptically. Certainly, PROBE can't get along with a packaged training institute alone.

The studies described were both directed toward the practical problems of helping schools to make use of the resources which PROBE offers. The community-based evaluation study investigated the use of specific statements of goals as a means to involve parents and students in the program decision-making. The second study tried to determine if decisions made by teachers in the use of PROBE were enhanced by a workshop experience. Research of a practical nature will necessarily be continued by the PROBE staff, since our concern is directed to those procedures which can ultimately make a change in the effect of the schools.

Figure 1. Mathematics Objectives

1. To add, subtract, multiply or divide measures.
2. Given a set of numbers, the student will compute the average.
3. To rename a decimal numeral using scientific notation.
4. Given a numeral written in expanded notation, the student will name the decimal numeral for the indicated sum.
5. To identify equal and equivalent sets.
6. To add, subtract or multiply decimals.
7. To find the greatest common factor of a set of numbers.
8. Given a decimal numeral, the student will round it off.
9. To multiply whole numbers.
10. To find area of a rectangle, square, triangle, parallelogram, or rhombus.
11. Given a word problem involving two place multiplication, the student will solve the problem.
12. To add and subtract whole numbers.
13. To write the fraction shown when given a picture.
14. To subtract unlike fractions and reduce the answer to its simplest form.
15. To find the circumference of a circle.

Table 1. Means and Standard Deviations of Ratings of Mathematics Objectives[a]

Objective	Parent Rating		Teacher Rating		Student Rating[b]		Student Rating[c]	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	4.61	.79	2.55	1.36	3.96	1.09	3.98	1.07
2	4.17	.96	3.85	1.73	3.90	1.15	3.74	1.13
3	3.96	1.09	4.30	1.13	4.00	1.15	3.84	1.16
4	4.03	1.11	3.55	1.19	4.02	1.08	3.75	1.15
5	3.74	1.23	4.45	.76	3.03	1.42	3.01	1.38
6	4.66	.63	3.10	1.62	4.10	1.07	4.00	1.07
7	4.26	.85	2.75	1.21	3.63	1.28	3.49	1.22
8	4.19	.89	3.50	.95	3.75	1.21	3.65	1.19
9	4.68	.70	4.75	.64	3.99	1.31	3.87	1.25
10	4.19	.95	2.95	1.28	4.29	1.07	4/07	1.13
11	4.41	.84	3.45	1.40	3.90	1.26	3.78	1.20
12	4.67	.84	4.75	.64	3.59	1.49	3.61	1.40
13	4.32	.91	4.75	.44	3.74	1.21	3.72	1.22
14	4.41	.83	3.50	1.15	4.04	1.09	3.89	1.08
15	4.16	.93	3.85	1.73	4.36	1.22	4.17	1.10

[a]Rating of 5=very important, 1=very unimportant

[b]Ratings of students of parents responding to questionnaire

[c]Ratings of total number of students

Table 2. Mean Correlations for Three Groups' Ratings and Student Achievement across Fifteen Objectives

	Teacher Rating	Student Rating	Student Rating
Parent Rating	-.15	.36	.53
Teacher Rating		-.29	.32
Student Rating			.44

Table 3. Means and Standard Deviations of Parents, Teachers and Students Predictions and Learner Acheivement

Parent Prediction		Teacher Prediction		Student Prediction [a]		Student Achievement [a]		Student Prediction [b]		Student Achievement [b]	
$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
.15	.36	.21	.26	.20	.41	.51	.97	.22	.42	.35	.37
.28	.45	.49	.34	.42	.50	.36	.59	.46	.50	.34	.37
.56	.50	.57	.21	.36	.48	.21	.24	.45	.50	.23	.27
.56	.50	.25	.22	.48	.50	.31	.38	.44	.50	.33	.38
.06	.23	.50	.26	.16	.37	.29	.28	.23	.42	.33	.31
.07	.26	.10	.19	.16	.37	.30	.28	.21	.41	.34	.33
.19	.40	.26	.33	.32	.47	.38	.36	.27	.45	.42	.37
.30	.46	.41	.23	.31	.47	.28	.28	.32	.47	.30	.31
.01	.12	.91	.72	.03	.17	.75	.32	.06	.23	.75	.42
.52	.50	.10	.15	.58	.74	.24	.28	.53	.50	.26	.31
.10	.30	.18	.24	.06	.24	.58	.37	.07	.26	.63	.37
.01	.12	.79	.16	.03	.17	.77	.26	.03	.18	.76	.28
.03	.17	.92	.10	.10	.31	.79	.32	.09	.29	.79	.32
.23	.42	.17	.22	.24	.46	.38	.40	.19	.36	.37	.41
.80	.40	.49	.35	.79	.41	.18	.25	.79	.41	.24	.32

[a]Predictions and achievement of students of parents responding to questionnaire

[b]Predictions and achievement of total number of students.

Table 4. Mean Correlations Across Fifteen Objectives for Predictions and Student Achievement

	Teacher Prediction	Student Prediction	Student Achievement
Parent Prediction	-.37	.88	-.79
Teacher Prediction		-.21	.60
Student Prediction			-.79

Table 5. Pre and Posttest Proportions for Classes of Trained and Untrained Teachers for Six Objectives

Objective	Trained Group		Untrained Group	
	Pretest $\bar{X}$	Posttest $\bar{X}$	Pretest $\bar{X}$	Posttest $\bar{X}$
1. Given a legend of cartographic symbols, the student will select the map which correctly applies the symbols.	.72	.89	.76	.92
2. Given a topographical map containing symbols previously presented, the student will identify the features.	.24	.61	.26	.52
3. Given factual data, the student will select the graph which correctly represents the data.	.41	.81	.44	.83
4. Given a simple graph and statements of interpretation the student will identify the correct interpretation.	.35	.57	.37	.56
5. Given a graphic representation of data and a legend, the student will select the graph which correctly represents the data as interpreted by the legend.	.29	.81	.33	.76
6. Given factual information, the student will write a report according to the following criteria: (1) inclusion of a topic sentence, (2) accurate reference to information, (3) inclusion of a summary sentence.	.10	.23	.09	.27
TOTAL	.29	.64	.30	.64