DOCUMENT RESUME

ED 038 374 SP 003 837

AUTHOR Webb, Jeaninne; And Others
TITLE Observation as a Methodology.
PUB DATE 70
NOTE 12p.; Paper presented at the American Educational Research Association meeting, Minneapolis, March 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$0.70
DESCRIPTORS *Classroom Observation Techniques, *Inservice Education, *Inservice Teacher Education, Program Improvement, Public School Teachers, *School Personnel, *Training Techniques

ABSTRACT

This paper describes the 2-year evolution of an ongoing program of inservice training for public school personnel designed to train observers in the use of a systematic observation technique--the Reciprocal Category System. The 605 participants include teachers, librarians, administrators, and central office personnel. Nine aspects of the four training sessions are discussed with respect to the three major headings of "organization," "training procedure," and "observer competence." Under the category of organization, training time, group size, and training sequence are considered. A training time of two days is developed, while a comparison of results by group size indicates that large groups are as effective as small groups. Training procedural concerns include development of a manual, revised twice; content, which shifts in emphasis from concepts and theory to skills and calculations; and trainer use. Observer competence elements include types of instruments--a concept test and a skill test--and the improvement made in these two instruments. The paper concludes that it has been found possible to train large groups of observers in two days to a responsible level of accuracy in the use of the Reciprocal Category System for classroom observation for purposes of classroom self-analysis. (RT)

OBSERVATION AS A METHODOLOGY

by

Jeaninne Webb
University of Alabama

Ernest L. Bentley
Atlanta Public School System

R. Robert Rentz
University of Georgia

Introduction

Systematic observation is a data collection procedure with recognized usefulness in a variety of settings, ranging from controlled experimental research to self-evaluation efforts of classroom teachers. Reports of research using interaction analysis and other observational techniques have included little discussion of the major problem in using observation--the training of observers. It is generally assumed that training observers to an acceptable level of accuracy takes a considerable amount of time, so as to preclude the use of systematic observation in areas where it would be most appropriate. Besides its obvious function in empirical research, observational systems can be useful tools in the supervision and self-evaluation of teachers.

It is the purpose of this paper to describe the two year evolution of an ongoing program of inservice training for public school personnel designed to train observers in the use of a systematic observation technique, with a focus toward the use of observation as a methodology for teacher self-evaluation. Three primary issues are considered in this discussion. These are:

1. The development of training procedures to produce maximum proficiency of observers with the most economical use of time and training personnel.
2. The development and testing of training materials.
3. The development of methods to assess observer competency.

General Description of the Training Sessions

The Reciprocal Category System (RCS) developed by Richard Ober was chosen as the observational system around which the training program would be designed. Ober's RCS is one of several systems that focus on an analysis and description of classroom verbal interaction.

Four training sessions have been conducted in the Metro Atlanta area during the past two years. A total of 605 trainees have participated in these sessions. The sessions were conducted during the summer of 1968 (S-68), the winter of 1969 (W-69), summer of 1969 (S-69), and winter of 1970 (W-70). Participants have represented the range of public school assignments, including: teachers, librarians, administrators and central office personnel. They were selected for attendance by their respective school systems on both a voluntary and assigned basis. Participants were of both sexes and represented various ethnic groups.

Various aspects of the training sessions have been developed and modified over the period that the training was conducted. Decisions to make modifications were based on both subjective judgments and empirical data. In presenting the evolution of this program, the plan is to present the salient features of each of the four training sessions in terms of nine major aspects of training, focusing primarily on those modifications that seem to have the widest implications and greatest

generalizability. Where available, empirical data are presented to support certain training modifications; at other times, the subjective judgments that led to changes are reported. The nine aspects of training are discussed with respect to the three major headings of Organization, Training Procedures, and Observer Competence. Figure 1 summarizes the salient features of the four training sessions with respect to the nine aspects of training.

Organization

Three aspects of organization were considered--training time, group size, and training sequence. Training time refers to the number of hours required in training participants. After the first session, S-68, which was conducted over three days, training sessions were reduced to 12 hours or two days. Table 1 shows the effect of this time reduction on accuracy of observer codings of a taped sample of classroom

Table 1

Comparison of Three-and Two-Day Training Sessions on Criterion Tape Accuracy*

	S-68	W-69
N	76	208
% Above .60	62%	50%
% Above .50	77%	71%

* Accuracy coefficients were computed using Scott's method for observer reliability. Scott's procedure essentially provides an index of observer agreement. The term accuracy coefficient is used here since each trainee's observations were compared with his trainer's coding of the criterion tape. In this sense the coefficient is more nearly like a validity estimate than a reliability estimate.

	Summer 1968 N=76	Winter 1969 N=228	Summer 1969 N=220	Winter 1970 N=81
Organization				
Training Time	3 days (18 hours)	2 days (12 hours)	2 days (12 hours)	2 days (12 hours)
Group Size	1 Group, Large-Small (3)	4 Groups (N=30 to 80)	6 Groups (1 large N=94 5 small N about 30)	4 Groups (N about 25)
Training Sequence	Fixed	Variable	Standard Sequence	Standard Sequence
Training Procedures				
Manual	Version 1	Version 2	Version 3	Version 3
Content	Concepts and theory: Skills and calculations	Less theory and calculation; more skill in application	Increase emphasis on performance skill	More practice in data collection
Trainer	1 for large group 3 for small group leaders	Team-Teaching Approach	Complete sequence per Trainer	Complete sequence per Trainer
Observer Competence				
Type instruments	Concept Test Criterion Tape Skill Test Final Exam	Concept Test Criterion Tape	Concept Test Criterion Tape	Concept Test Criterion Tape
Concept Test	Version 1-25 items	Version 2-36 items	Version 3-38 items	Version 3-38 items
Criterion Tape	Version 1	Version 1	Version 2 Standard Key	Version 2 Version 3

Figure 1 Summary of the Salient Feature of Four Training Sessions

verbal behavior. The reduction in the level of trainee accuracy was not deemed sufficient to continue the three day schedule. These results should be viewed cautiously since factors affecting accuracy scores, in addition to reduction in training time, included several other changes in procedures.

Group size as a variable has been varied in the following manner: (1) combination of large group-small group, with small groups (approximate n of 25) for skill instruction; (2) small groups varying in size from 15 to 40 participants; (3) large groups (n of 94) versus small groups (approximate n's of 30) for self-contained instruction; and, (4) standard small groups of 20 to 25 participants. Table 2 reports results of a large group, small group comparison where the procedures were as nearly equivalent as possible. While differences for both

Table 2

Comparison of Trainee Performance in Large and Small Group Training Sessions

	Large Group N=94	Small Group N=31
Sample Size		
Concept Test		
Mean	22.3	20.0*
S.D.	5.2	4.6
Criterion Tape		
% Accuracy Scores above .60	55.3%	51.6%

* Difference between the two means significant at the .05 level.

criteria favor the large group, it is more important to note that they do not favor small group size. This implies an important economic savings in training.

Training sequences have involved fixed, variable and standard approaches. Early attempts included a fixed schedule involving all participants in the same activities (though at times as one large group and at other times as small skill instruction groups). During W-69 a variable sequence was employed, because trainers had not been satisfied with procedures of organization for the S-68 session. The variable sequence was achieved by teaming instructors, dividing participants into two groups, and flip-flopping the activities. The groups participated in a different instructional sequence. Subjective trainer judgment and dissatisfaction with this "team teaching" approach led to the "standard sequence" utilized with the S-69 and W-70 participants. Each trainer conducts the entire set of training activities with a single group.

Training Procedures

Three main procedural concerns were development of a training manual, ordering and presenting content in training sessions, and trainer differences. The training manual evolved through three states to its present form. Version 1 was a collection of articles and interpretive comment. Version 2 was written specifically for training purposes based largely on Richard Ober's Working Manual (incorporated in Version 1). Version 3 moved toward a semi-programmed revision of the second manual. Refinement of manual content is

continuing. The manual has become an integral part of the training materials and provides a degree of standardization of training procedures.

Content originally was separated into concepts and theory, and skills and calculation (S-68). Based on trainers' opinions that the emphasis should be on applications of skills, W-69 sessions devoted less time to theory and calculations and more time to coding skills. During S-69 increased emphasis was placed on observer performance skill; the only change in W-70 resulted in an increase in practice time. This additional practice will be examined by comparing S-69 with W-70 for gross indications of this modification as soon as W-70 data can be analyzed.

The effects of individual trainers were appraised in the initial session (S-68). Results indicated that for the original trainers there were no apparent differences. Since these trainers continued to serve as the core instructors through the series no additional investigations have been attempted. Once procedures are standardized, subsequent investigations can examine trainer differences. Division of responsibility has been varied with initial procedures (S-68) requiring one trainer for the total group to handle theoretical considerations and individual group leaders for skill instruction. W-70 sessions utilized special division of tasks such that cooperating trainers handled either data collection activities for all participants (in small groups) or preparation and interpretation activities. By S-69 trainers directed the entire sequence for a specified group of participants; this procedure was continued in W-70 sessions. The

evolution of a complete sequence per trainer approach was based on judgments of trainers in post mortem discussions of the training sessions.

Observer Competence

Instruments for assessing observer proficiency have all along been considered of major importance. Four types of instruments were considered initially (S-68). These included: (1) A Concept Test stressing understanding of the major aspects of the system; (2) Criterion Tapes which focused on extent to which a trainee could accurately code the verbal interaction depicted on the audio tapes; (3) A Skill Test which was concerned primarily with data preparation and interpretation; and, (4) A Final Exam which was essentially a combination of content covered by the Skill Test and the Concept Test.

Analyses of data for S-68 indicated that the two most promising evaluative instruments were the Concept Test and the Criterion Tape accuracy scores. Scores from these two devices showed low correlations (r=.14) indicating different aspects of competency were being assessed. In addition, a substantial correlation (r=.74) between the Skill Test and the Final Exam indicated that in the interest of time one of these instruments could be eliminated. It was decided to eliminate the Final Exam entirely and to incorporate the Skill Test into the training procedures. The subsequent sessions used only the Concept Test and Criterion Tape methods to assess observer competence.

The Concept Test underwent two revisions based on S-68 and W-69 data and another revision is planned based on W-70 data that will involve increasing test length in order to increase test reliability. Table 3 summarized the statistical data for the different versions of the Concept Test.

Table 3

Statistical Data for the Different Versions of the Concept Test

	S-68	W-69	S-69	W-70
Version No.	1	2	3	3
No. Items	25	36	38	38
Mean	16.84	21.52	21.14	23.65
S.D.	3.10	5.06	4.99	4.97
Reliability (KR-20)	.592	.732	.712	.732
Sample Size	76	228	251	81

The development of a Criterion Tape has gone through several stages. The first tape to be used (Version 1) was not only used to assess accuracy in coding observations, but was also used in the training procedures. This practice was somewhat questionable and was discarded after the W-69 session. Another reason for revising the Criterion Tape and evaluation procedures was that in calculating observer accuracy (for sessions S-68 and W-69), each participant was compared with his own group trainer's coding.

To improve on some of these deficiencies a second tape was developed (Version 2) to be used only for final competency evaluation. Additionally, a standard "key" was constructed so that each trainee accuracy score would be based on the same comparison. This "key" was developed from a composite of the ratings of five judges who had served as trainers. The intra-judge reliabilities for these five judges is shown in Table 4.

Table 4

Intra-Judge Reliability* of Ratings for Version 2 of the Criterion Tape

Judge No.	1	2	3	4
2	.66			
3	.85	.73		
4	.69	.76	.77	
5	.80	.65	.86	.65

* Reliability of judges ratings computed by Scott's coefficient procedure

There are several categories of verbal behavior measured by the RCS that were not represented in Version 2 of the criterion tape. It was felt that another tape should be developed that better represented all of the categories in the RCS system. This tape has been developed (Version 3, W-70) but the data have not yet been processed.

Current Status

Two years of schedule refinement have reduced the original procedures from 18 hours of instruction and testing to 12 hours of instruction and criterion performance tasks. Savings in time and paper work resulted largely from the finding that the concept criterion score and accuracy score predicted substantially what was identified through paper and pencil skill tests and final competence checks. A current training schedule is presented as Figure 2. The schedule evidences the critical importance of accuracy in coding interactions. Accuracy exercises on the second day are designed to compare participant coded data with the criterion score on the same audio tape. Only two segments--Data Preparation and Data Interpretation--focus on other aspects

Figure 2

Schedule of the Most Recent Training Session (M-70)

Thursday, January 22, 1970		Friday, January 23, 1970	
8:45	Introdution to PCS	8:45	Collection of live data
	Beginning Data Collection	9:00	Micro-lesson (Simulation)
10:30	Coffee	10:15	Coffee
10:40	Response Discrimination (15's & 16's)	10:25	Data Interpretation
11:30	Intermediate Data Collection	12:00	Accuracy Exercise
12:30	Lunch	12:30	Lunch
1:15	Data Preparation	1:15	Advanced Data Collection
2:30	Intermediate Data Collection	2:30	Accuracy Exercise
3:00	To be continued	3:00	Synthesis

of observer competence. This does not imply lack of importance of data use. Rather, it focuses the participant on "first things first," the task of accurately observing and recording verbal behavior.

Data collection is approached from three complementary veiws: practice coding of prepared audio tapes, practice coding of live simulated situations, and practice in eliciting selected verbal behaviors through individually planned minute lessons (based on *a priori* scripts of sequences of categories) as well as coding these simulated lessons. Comparisons in all cases are against "expert" criterion scores.

Data preparation skill sessions include techniques of summarization (bracketing, plotting, and totaling) and calculations (percentages and ratios). Data interpretation sessions emphasize identification

of important aspects of data sequences and clumpings based on instructional plans of the teacher, (patterns of verbal behavior and concentrations).

After two years of planned successive approximations it is possible to report that large group observers can be trained to a responsible level of accuracy and educated to reasonable theoretical and conceptual understanding within 12 hours of training for purposes of classroom self-analysis through the Reciprocal Category System of Interaction Analysis.