


AN ABSTRACT OF THE THESIS OF

Karen Beal for the M. S. in Family Life and Child Development
(Name) (Degree) (Major)

Date thesis is presented May 11, 1964

Title A COMPARISON OF ORAL AND WRITTEN RESPONSES TO
A CLASSROOM SIMULATION TEST

Abstract approved 
(Major Professor)

The purpose of this study was to compare the quality of data obtained from individual oral and group written administration of a classroom simulation test. The individual oral procedure involved one group of subjects taking the simulator test individually and responding orally, while the group written procedure involved another group writing their responses to the simulator test under group conditions.

The sample consisted of 34 women majoring in elementary education at Oregon College of Education at Monmouth, Oregon. These subjects were matched on the basis of age and years of teaching experience.

In each of the two different test procedures the subjects responded to 11 short filmed episodes of the classroom behavior of 22 sixth grade children. The subjects enacted the role of student teacher and were asked to respond at any time during the filmed episode where they felt some response or action was needed. Each subject was equipped with a stop watch. This was used to indicate the point in time in the sequence of events that a subject chose to initiate a response or formulate a judgment. In both test procedures the subjects were asked to respond as if they were actually the teacher talking to the children portrayed on the film.

Upon completion of each of the episodes, the subjects were asked to answer three questions: (1) What was it in this situation that made you, as the teacher, respond when you did? (2) Why did you say what you said or do what you did? and (3) What did you anticipate achieving, if anything, by the response you made? These questions were used to try and tap the rationale or reasoning for the particular response made.

The protocols, which included responses to the filmed episodes and the answers to the three questions, from the two test procedures were compared in the following ways: (1) the number of words elicited by each method, (2) the point in the episode (time) at which the subject responded, (3) the degree of involvement in the response, (4) the homogeneity of content, and (5) the consistency between the rationale given for a response and the response itself.

Analyses three, four and five involved placement of the protocols in category systems. Reliability of category classification was demonstrated by a measure of percent agreement between two sets of raters before the protocols were classified.

The data were analyzed by the chi-square analysis and the t-test for related samples to determine the similarity of data elicited by the two administrative procedures.

In general, the results of the analyses indicated that the two methods of test procedure, individual oral and group written, elicited comparable data to a classroom simulation test. On the basis of these results, several conclusions seem justified:

1. The group written procedure is economically superior to the individual oral procedure because of its tendency to elicit significantly shorter protocols and because it can be administered to more than one person at one time.

2. Both methods are similar in terms of the time of

initiation of response.

3. The group written procedure and the individual oral procedure are similar in facilitating involvement in the simulator test.

4. Both methods are similar in terms of the nature of the content elicited.

5. Consistency between response and rationale for the response was independent of method of administration.

The results of the present investigation cannot claim superiority for one method over the other, but it points to the equality of both. Additional research involving the use of simulation materials under various methodological conditions needs to be done in order to further explore the potential of this type of test stimuli.

A COMPARISON OF ORAL AND WRITTEN RESPONSES
TO A CLASSROOM SIMULATION TEST

by

KAREN BEAL

A THESIS

submitted to

OREGON STATE UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

June 1964

APPROVED:

[REDACTED]

Associate Professor of Department of Family Life

In Charge of Major

[REDACTED]

Head of Department of Family Life

[REDACTED]

Dean of Graduate School

Date thesis is presented May 11, 1964

Typed by Eileen Humphrey

ACKNOWLEDGEMENT

I would like to express appreciation to Dr. H. D. Schalock, whose capable direction and guidance were invaluable in this work.

I would also like to express thanks to Dr. Bert Y. Kersh, Associate Director of Teaching Research Center, for his assistance and cooperation.

TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION	1
Background and Statement of the Problem	1
Review of Literature	3
Purpose of the Study	6
PROCEDURE	7
Subjects	7
Design	8
Administration of Tests	10
Measurement	12
Rater Reliability	14
THE DATA AND THEIR TREATMENT	17
DISCUSSION	26
SUMMARY AND CONCLUSIONS	31
Summary	31
Conclusions	32
BIBLIOGRAPHY	34
APPENDIX A	35
APPENDIX B	37
APPENDIX C	38
APPENDIX D	41

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	The Distribution of Subjects by Years of Teaching Experience in the Two Treatments	8
2	Classification of Episodes Used in the Simulation Test	9
3	The Involvement Category System	13
4	The Content Category System	14
5	Percentage-Agreement Between Raters A and B in Classification of Responses into Involvement Category System	15
6	Percentage-Agreement Between Raters A and B in Classification of Responses into Content Category System	16
7	Percentage-Agreement Between Raters B and C in Classification of Rationale into Content Category System	16
8	Difference of Number of Words Across Responses Used by Each Subject by Method of Administration	18
9	Difference of Number of Words Across Questions Used by Each Subject by Method of Administration	20
10	Difference Scores Per Episode of Initiation of Response Time by Method of Administration	21
11	Comparison of Frequencies Within the Involvement Category System by Method of Administration	22
12	Comparison of Frequencies in the Content Category System by Method of Administration	23
13	Comparison of Frequencies Occurring Within Sub-Categories in the Content Category System by Method of Administration	24
14	Frequencies of Consistency Between Episode Response and Questions One and Three by Method of Administration	25

A COMPARISON OF ORAL AND WRITTEN RESPONSES TO A CLASSROOM SIMULATION TEST

INTRODUCTION

Background and Statement of the Problem

The area of tests and measurement plays an important part in the search for knowledge about human behavior, for without some form of measurement the classification and prediction of behavior, attitudes and abilities could not proceed. When using psychological or educational tests, however, there is always the question of which administrative procedure to use: individual or group administration. The issue is one of striking a workable balance between economy of administration and utilization of the potential of the test (2, p. 433-454). If it could be demonstrated, for example, that a test administered to a group provided data as adequate as that obtained through administration to single individuals, the test would be strengthened in its bid for widespread use and application. Demonstrating the relative merits of different administrative procedures is a task for research, and it is to this task that the present study is addressed.

Recently Kersh (5, p. 1-12) at the Teaching Research Laboratory at Oregon College of Education, has developed what he has called the classroom simulation test. In general, the process of simulation means the production or reproduction on film of real-life situations. Classroom simulation "refers to the technique of recreating on film, through the use of sound motion pictures and with printed materials, a single classroom of youngsters so that nearly all of the relevant features of the classroom environment may be displayed" (5, p. 1).

In the film the children are facing the camera so that, to the subject viewing the film, it appears as if the children are looking and speaking directly at him or that a situation is occurring directly

in front of him. The person observing the film assumes the role of teacher and responds to the filmed situations as if he were the teacher in the classroom.

The methodology developed by Kersh is currently being used by Schalock in a study of the use of motion pictures as test stimuli. Schalock is concerned generally with the predictability of complex behavior, in this instance the interpersonal behavior of teachers in the classroom, and is testing the hypothesis that as test stimuli increase in complexity, prediction to real-life behavior will increase in accuracy. The research design called for a series of tests, ranging in complexity from a traditional word test (Minnesota Teacher Attitude Inventory) through a motion picture choice test to a classroom simulation test, to be given to student teachers the term before they entered the field for their practice teaching experience. Each student was then observed during her practice teaching experience, and the relationship between her test scores and her actual behavior in the classroom was to be determined.

Before testing the subjects with the classroom simulation test, the method of administration had to be determined. Kersh had developed the simulator initially as an individual test, but since all of the other tests used in this study were designed for group administration, and since the simulation test required an hour and a half to complete, its administration on an individual basis represented a costly and time consuming step within the project. The present study was undertaken to determine the feasibility of adapting the simulator to group administration. This was to be done by comparing the similarity of data coming from an individual oral administration and a group written administration of the classroom simulation test. If it were found that the group written procedure provided data as adequate as that provided by the individual oral procedure, the group

written procedure could be adopted for the parent project.

Because the classroom simulation test for the parent project was still in the process of being developed at the time the present study was undertaken, a modification of the simulation film test developed by Kersh was used in the study.

Review of Literature

There is some indication in the literature that written responses to various media provide just as much useful data (of equally high quality) as do oral responses. A study by Eron and Ritter (3, p. 59-61) on the comparison of two methods of administering the Thematic Apperception Test found, in general, that the individual oral and group written methods were essentially the same. There was marked similarity between the stories contributed by the subjects taking the test under the two conditions. However, there were some differences which exceeded chance expectations. Stories obtained in the oral administration were longer than those obtained in the group written administration. But on the basis of content, that is, actual thematic material elicited, the two groups were almost identical.

A study by Pine (9, p. 24-30) on the comparison of written and oral responses to a projective test dealing with marriage interaction also showed that group written responses were equal to and sometimes superior to either the individual oral or individual written procedures. In terms of eliciting thought units that could be classified as evidencing dissatisfaction, the group written administration was superior to either of the other two administrative procedures. There were no apparent differences between procedures in the effectiveness with which they elicited thought units that could be classified as evidencing satisfaction. Pine found the group written procedure to be economically superior to both of the other administrative procedures on two counts: (1) because of the tendency of the group procedure

to elicit significantly shorter, though just as adequate, protocols, and (2) because it can be administered to a number of persons at once.

With regard to content or quality of response, Metzner and Mann (8, p. 489-491) reported that writing answers to questionnaires and open-end interviews may permit greater freedom on the part of the subject to express asocial or antisocial feelings than oral responses. Along similar lines Lindzey and Heinemann (7, p. 35-40), in a study comparing individual and group administration of the Thematic Apperception Test, reported that there was some very slight evidence suggesting the superiority of the group method. The authors suggest that this might be accounted for by the subjects feeling a greater sense of anonymity in a group situation, and consequently feeling freer to express themselves. There was also the possibility of there being some social facilitation effects in the group situation that could tend to increase motivation for expression.

Lindzey and Heinemann also made the point that for purposes of research, where the investigator is interested in a limited number of characteristics of a large number of subjects, or for purposes of screening a large number of subjects for particular qualities that can then be closely examined, any advantage of the individual method over the group method of administration would be slight. A study by Clark (1, p. 46-50) supports this point of view.

Some evidence, however, suggests the superiority of the oral over the written procedure. Terry (11, p. 507-511) found that written stories showed a significantly lower average level of response than did the oral stories, suggesting that the subjects writing their stories become less involved in the test and thus produce less personal material. Terry reasoned that the value of a projective technique lies in the degree to which the stimuli elicit

characteristic responses from the subject and that it seemed plausible to assume that the projective value of a technique, such as the Thematic Apperception Test, was a function at least in part of the subject's degree of personal involvement.

Studies pertaining specifically to the comparison of administrative procedures in motion picture tests are practically non-existent. Levonian (6, p.250-254) has pointed out in fact that the application of motion picture films to test methodology has been extremely limited. He goes on to say that this is rather surprising in view of the efficiency, control, and flexibility afforded by the use of this medium for the assessment of opinions, attitudes and feelings.

The literature is also limited in the area of motion picture simulation around classroom experiences. Kersh (4, p.41-52), in his exploration of the techniques of simulation, developed a classroom simulation test for use in the pre-service training and education of elementary school teachers. The problem with which he was concerned was that of training a student teacher in the ability to detect, diagnose, and resolve such teaching problems as confusion, inattention, distraction and fatigue on the part of the children in the classroom. In the procedure he developed, a subject viewed the filmed episodes and was asked to respond as if he were the teacher. The reactions of the subject were rated by the experimenter and then discussed with the subject. As a teaching aid, there was evidence to suggest significant gain in understanding by those who participated. Reactions of the student teachers to the simulator were almost unanimously favorable.

In view of the fact that most of the studies that have been done on written vs. oral responses have involved either the Thematic Apperception Test or other similar projective tests, it was felt that the question concerning test procedure for classroom simulation

test must be demonstrated. Two factors make such a study necessary: (1) the stimuli presented in a classroom simulation test are quite different from those used in traditional projective techniques, and (2) the nature of the demands made upon the subject by this media are assumed to be unique.

Purpose of the Study

The purpose of this study was to compare the quality of data obtained from individual oral and group written administration of a classroom simulation test. The individual oral procedure involved one group of subjects taking the simulator test individually and responding orally, while the group written procedure involved another group writing their responses to the simulator test under group conditions. Since economy of test procedure was a major concern of the study, adequacy of individual written administration was not considered.

The major hypothesis to be tested in the study was that there would be no significant difference in quantity or quality of data obtained from individual oral and group written administrative procedures.

PROCEDURE

Subjects

Thirty-six subjects took part in the investigation originally, but one subject and her matched partner were later dropped because of incomplete protocols. This left a total of 34 subjects on which the analyses were made. These were all women who were attending summer school at Oregon College of Education at Monmouth, Oregon, and majoring in elementary education. The use of women, and only those majoring in elementary education, was in keeping with the design of the parent project. In order to make reliable generalizations from this study to the parent project, it was desirable to keep the subjects as similar as possible.

Participation was on a voluntary basis. The qualifications for participation were either (1) some experience teaching at the elementary level, or (2) junior or senior standing in college classwork. From an information sheet asking questions concerning grade level taught, age, years of teaching experience if any, and the like, it was possible to obtain a base for pairing the subjects. After considering the characteristics on which pairing could take place, teaching experience was chosen as the variable most needful of control. To this end, each pair of subjects was matched as nearly as possible on this trait. One member of each pair, chosen from the two by random means, was assigned to the individual oral treatment while the other member was assigned to the group written treatment. Table I shows the distribution of subjects by years of teaching experience in each of the two treatments.

Although the primary characteristic for matching the subjects was years of teaching experience, the age of the subjects was also taken into account when the original matching took place. The range of age differences between matched subjects was anywhere from one

to ten years. The average difference in age between two subjects in a matched pair was five years. The majority of the participants were 30 to 40 years of age.

Table 1. The Distribution of Subjects by Years of Teaching Experience in the Two Treatments

Years of Teaching Experience	Method	
	Individual Oral	Group Written
0-3 years	6	6
4-10 years	5	5
11-20 years	5	5
over 20 years	1	1

Design

The classroom simulation test used in this study consisted of a series of 11 filmed episodes of classroom behavior which were assumed to represent a sampling of situations a teacher might encounter in a classroom. The episodes revolved around issues of management, the development of content, and communication as these appeared in a sixth grade class. The classification of these episodes, as identified by Kersh (4, p.24), are presented in Table 2.

Each of these 11 episodes, which were called "problem episodes", lasted approximately 30 seconds and was prefaced by a verbal description of the setting in which the problem occurred. The description included such factors as the time of day, the assignment on which the students were working, the instructions which had been given them, whether or not the supervising teacher was present in the classroom, and the like. These descriptions may be found in Appendix A. After the setting description was read the episode was shown and the subject or subjects would respond

according to instructions.

Table 2. Classification of Episodes Used in the Simulation Test

Classroom Behaviors Represented	Total
Management problems	
Disorderly behavior	3
Distracting behavior	3
Baiting-testing	2
Fatigue-reaction	1
Communication problems	
Inattention	1
Content	1
Total problems	11

The 34 subjects were randomly assigned to one of the two treatments; individual oral or group written. Each subject in the individual treatment responded to the filmed episodes orally and alone. These responses were tape recorded and then typed for coding at a later time. The written responses obtained under the group conditions were coded directly from the written protocols.

A variable of relevance which needed to be acknowledged and controlled was the timing factor. The concern was not with the length of time involved in making or writing a response, but rather at what point in the sequence of events did a subject choose to initiate a response or formulate a judgment. Put in another way: at what point did the subject, in the role of teacher, feel a response or action was needed? To control this, a record of the time lapse before a response was made or written was recorded by means of a stop watch. During the individual oral sessions this recording of initiation of response was noted by the subject and written down by the experimenter. In the group written treatment, each subject was equipped with a stop watch and each noted his own time of

initiation and recorded it.

Administration of Tests

Before taking the test each subject was given a five-minute orientation or "practice episode" in order that she might become adapted to the particular procedure demanded in the simulator. Refer to Appendix A for the description of the practice episode. The practice conditions corresponded with the conditions under which the test episodes were given; that is to say, each subject in the individual oral treatment received an individual oral practice episode while those in the group written treatment received group written practice. The administrative procedures are described below:

Individual Oral: Each subject, during the administration of the individual oral test, was alone with the experimenter. The subjects were told that their responses were to be tape recorded. They were given the following instructions:

This film, called a simulator, is of a single classroom of 22 sixth graders. Their regular teacher is Mr. Land. This simulator is comprised of a variety of short filmed episodes each involving some aspect of classroom activity. These episodes you will see start near the beginning of a school day and continue right through until the late afternoon.

You will be enacting the role of student teacher. I want you to respond at any point in the episode, at any time which you feel the situation requires some action or response from you as the teacher in charge. It is not necessary to wait until the episode is completed. Try to respond as freely and openly as you would were you really in this situation; as though you were actually the teacher talking to these children.

The experimenter would then read the setting description. The filmed episode would begin and the subject would respond accordingly.

Group Written: For purposes of this administration the

subjects came together as a group to take the test. Their instructions were similar to those given the other subjects. They were told to respond at any point in the episode where they felt a response of some nature was warranted from the teacher, whose role they were enacting. They were asked to write their response the moment it seemed most appropriate for them to make it, and to ignore what followed in the episode and attempt not to be distracted by it.

Concurrently with this, instructions were given to all subjects, whether in the individual oral or group written modes, for the operation of the stop watch. Practice in its use was also given. In the group mode, the subjects were asked to stop the watch at the moment they began to write their response, and after they had completed their response, to record on their paper the reading from the stop watch. The subjects taking the test individually and orally were asked to stop the watch at the moment they began to verbally respond to the episode. The experimenter recorded the time reading after the completion of the episode.

After a subject had finished responding to a particular episode in either treatment she was asked to answer three questions. These questions were included as part of the test to try and tap the rationale or reasoning for the particular response made. The questions were: (1) What was it in this situation that made you, as the teacher, respond when you did? (2) Why did you say what you said, or do what you did? (3) What did you anticipate achieving, if anything, by the response you made?

After a subject, or subjects, had responded to the filmed episodes, had answered the three questions, and the response initiation time had been recorded, the experimenter would read the instructions for the next episode.

Measurement

The raw data for the study appeared in written form from the transcribed tapes and the written protocols. These had to be submitted to content analysis in order to arrive at data permitting a comparative analysis of treatments. Two separate category systems were developed for this purpose: a Content category system and an Involvement category system. The development of these two systems progressed independently, first from just looking at the data and identifying distinguishing characteristics, then organizing typical responses into groups, and finally enumerating, defining and describing the separate category classifications. In other words, the category systems evolved from the data themselves. The category systems are described below:

The Involvement Category System: This category system evolved as a result of the concern about the degree of personal involvement on the part of subjects in each mode. It was established as an attempt to discover which of the two modes best facilitated the process of a subject projecting herself into the situation, of becoming involved to the point of responding as if actually in the situation portrayed in the film. The categories involved in the system appear in Table 3. Examples from the data representing each category appear in Appendix B.

The Content Category System: As mentioned earlier, the classroom simulation test involved a sequence of 11 episodes that had been selected from a series of films developed by Kersh. Since each of the 11 episodes were filmed around a certain aspect of classroom behavior this dictated to a degree the direction for the development of this particular category system. That is to say, the Content category system developed to a large extent as a consequence of the nature of the filmed episodes themselves. For example, some

Table 3. The Involvement Category System

Degree of Involvement	Definition
Projection	Respondent projecting into the situation by directly addressing the children.
Description	Respondent describing what she would do or not do in a situation. Her response is more like that of an observer looking in on the classroom.
Combination	Respondent both describes what she would do and supplies a projective response.
No Response	Respondent choosing not to respond.

episodes dealt specifically with a child's personal problems or needs; others were concerned with the provision of structure for an activity, either by requiring the respondent to give instructions or by seeing that the children followed through on assignments previously given. Still other episodes had as their focus behavior which represented a deviation from expectations and the respondent had to deal with this in some way. The category system followed this breakdown of episodes in its development.

This particular system was established in order to compare the quality of data obtained from the two methods of test administration. A brief description of the system appears in Table 4. A complete description of the Content category system with examples of data fitting each category may be found in Appendix C.

Table 4. The Content Category System

<u>Philosophy</u>			
Method of directly interacting with a child in meeting needs and understanding problems and behavior.			
<u>No Response</u>	<u>Rigid</u>	<u>Understanding</u>	<u>Nurturing</u>
chose to wait	operates solely from an adult point of view	empathetic, but still firm in own direction	extends herself to sympathize and assist
<u>Approach</u>			
Method of providing physical and/or psychological setting to facilitate learning in the classroom.			
<u>No Response</u>	<u>Authoritarian</u>	<u>Democratic</u>	<u>Child-Determined</u>
chose to wait	unquestioning obedience and compliance to rules	considers child's point of view in decisions	child's request or appeal determines action
<u>Control</u>			
Method of stopping or modifying behavior after a deviation has occurred in the classroom.			
<u>No Response</u>	<u>Power: Strict</u>	<u>Power: Considerate</u>	<u>Power: Choice</u>
chose to wait	unqualified power; punishment	qualified power; persuasion or suggestion	gives child a choice or alternative to correct behavior

Rater Reliability

Before analysis of the data could be undertaken the reliability of raters with the category systems had to be established. Separate reliability checks were computed for (1) the Involvement classification of responses, (2) the Content classification of the responses, and (3) the Content classification of the rationale for the response, that is, the answer to the three questions raised in relation to each

response. A statement of the essential directions that were used in the classification of responses and rationale appear in Appendix D.

One graduate student in the area of family life and child development and one assistant on the parent research project worked with the writer in establishing reliability. Practice sessions were held to familiarize the raters with the category systems.

For the reliability measure of the classification of responses into the Involvement category system, raters A and B each classified independently the responses to 91 episodes. These were taken from eight individual oral protocols and nine group written protocols that had been picked randomly from the total pool of data. The classifications of the two raters were analyzed to determine the consistency with which they agreed. The formula used for the computation of reliability was:

$$\text{percent of agreement} = \frac{\text{agreements}}{\text{agreements} + \text{disagreements}}$$

These data appear in Table 5.

Table 5. Percentage-Agreement Between Raters A and B in Classification of Responses into Involvement Category System

Total Responses Classified	Total Responses Disagree	Total Responses Agree	Percent of Agreement
91	5	86	94.5%

Reliability of classification with respect to episode responses, using the Content category system, was also demonstrated by raters A and B. Each classified independently the responses from three complete individual oral protocols and two complete group written protocols. The formula used in computing this reliability measure was the same as that used in relation to the Involvement system.

These data appear in Table 6.

Table 6. Percentage-Agreement Between Raters A and B in Classification of Responses into Content Category System

Total Responses Classified	Total Responses Disagree	Total Responses Agree	Percent of Agreement
55	7	48	87.2%

A third and final reliability check was made on the accuracy of classification with respect to the rationale or answers to the questions asked around each episode. Here raters B and C each independently classified a total of 40 questions, taken from four group written protocols and four individual oral protocols, which had been randomly picked from the total pool of data. This reliability measure, using a different rater, was found to be nearly as high as it was for the response data in Table 6. These data appear in Table 7.

Table 7. Percentage-Agreement Between Raters B and C in Classification of Rationale into Content Category System

Total Responses Classified	Total Responses Disagree	Total Responses Agree	Percent of Agreement
40	5	35	83%

On the basis of these data it was concluded that the categories were sufficiently clear in definition and the writer was sufficiently reliable in categorization that the analysis of the data could proceed under independent classification by the writer. In view of the amount of time and effort that would be involved in establishing validity of these categories and in conjunction with the fact that no further implementation of these specific category systems was planned, no attempt was made to establish their validity.

THE DATA AND THEIR TREATMENT

The purpose of this study was to compare the data obtained from two methods of administering a classroom simulation test: individual oral and group written. The comparison of the data from these two methods was done in the following ways: (1) the number of words elicited by each method, (2) the point in the episode (time) at which the subject responded, (3) the degree of involvement in the response, (4) the homogeneity of content, and (5) the consistency between the rationale given for a response and the response itself. The specific hypotheses tested were:

1. The number of words in the response and in the rationale for the response will be significantly greater for the individual oral method of administration than for the group written method.

2. There will be no significant difference in the timing of initiation of response for the two methods.

3. There will be no significant difference in the degree of involvement for the two methods.

4. There will be no significant difference in the nature of the content elicited by the two methods.

5. There will be no significant difference in the consistency between response and rationale of the two methods.

These data will be considered in the order of the hypothesis listed.

In order to test the first hypothesis, that the number of words in the responses and in the rationale was greater in the individual oral method than in the group written method, a t-test for related samples was computed to determine the significance of difference in the number of words elicited by the two methods. In this respect two analyses were run: one for the number of words elicited in the responses to the episodes and one for the number of words elicited in the questions relating to each response. The data pertaining to

the number of words used by each subject in her responses in each of the two administrative procedures appear in Table 8.

Table 8. Difference of Number of Words Across Responses Used by Each Subject by Method of Administration

Paired Subjects	Method		y Difference
	Individual Oral	Group Written	
1	154	149	5
2	331	245	86
3	121	240	-119
4	317	286	31
5	193	355	-162
6	714	192	522
7	175	216	-41
8	264	167	97
9	239	282	-43
10	290	154	136
11	144	228	-84
12	275	253	22
13	175	276	-101
14	195	205	-10
15	189	138	51
16	103	289	-186
17	203	231	-28
Total	4,082	3,906	176

$$\begin{array}{l}
 n = 17 \\
 \Sigma y = 176 \\
 \bar{y} = 10.352
 \end{array}
 \quad
 \begin{array}{l}
 (\Sigma y)^2 = 30,976 \\
 \frac{(\Sigma y)^2}{n} = 1822.117 \\
 \Sigma y^2 = 408,528.000 \\
 SS = 406,705.883
 \end{array}
 \quad
 \begin{array}{l}
 s^2 = 15419.117 \\
 \frac{s^2}{n} = 907.00 \\
 \sqrt{\frac{s^2}{n}} = 30.116
 \end{array}$$

$$t = \frac{\bar{y} - 0}{\frac{s}{\sqrt{n}}} = \frac{10.352}{30.116} = .3437$$

16 d. f.

It will be seen from these data that the t value was .3437 which in no way approached the level required for significance at the .05 level of confidence. On this basis no significant difference was obtained between methods in terms of the number of words elicited in the responses to the classroom simulation test.

A similar analysis was computed to test if the number of words across questions (rationale) was greater in the individual oral method of administration than in the group written method.

The data pertaining to the number of words used by each subject, in responding to the three questions, by administrative procedure appear in Table 9.

Upon inspection of these data it was found that the t value was 2.0549, which was significant at the .05 level of confidence. From this result it was clear that the individual oral method of administration elicited significantly longer answers to the three rationale questions than did the group written procedure.

The next analysis undertaken was that required in relation to the timing of initiation of response. The hypothesis to be tested was that there would be no significant difference in the timing of initiation of response for the two methods. The t - test for related samples was also used in testing this hypothesis. The t value derived therefrom was only .6751 which was not significant at the .05 level of confidence. The data relevant for this analysis appear in Table 10. From these data it may be concluded that the act of responding to the filmed episodes, whether verbally and alone or by writing in a group situation, was independent of the test-taking procedure so far as the specific point or time within an episode that a response was made or formulated.

The two methods of administration were next compared in terms of the personal involvement elicited in the response to the episodes. Personal involvement was defined operationally in terms

Table 9. Difference of Number of Words Across Questions Used by Each Subject by Method of Administration

Paired Subjects	Method		y Difference
	Individual Oral	Group Written	
1	1049	465	584
2	641	386	255
3	886	1060	-174
4	675	492	183
5	782	540	242
6	2283	646	1637
7	870	318	552
8	846	438	408
9	826	582	244
10	1504	481	1023
11	674	324	350
12	684	431	253
13	730	347	383
14	576	418	158
15	975	400	575
16	1238	455	783
17	866	469	397
Total	16,105	8,252	7,853

$$\begin{array}{lll}
 n = 17 & (\Sigma y)^2 = 213,388.56 & s^2 = 388710.89 \\
 \Sigma y = 7853 & \frac{(\Sigma y)^2}{n} = 25,522.68 & \frac{s^2}{n} = 22,865.35 \\
 \bar{y} = 461.94 & \Sigma y^2 = 6,244,897.00 & \sqrt{\frac{s^2}{n}} = 151.212 \\
 & SS = 6219374.32 &
 \end{array}$$

$$t = \frac{\bar{y} - 0}{\sqrt{\frac{s^2}{n}}} = \frac{461.94}{151.212} = 2.0549 \text{ (16 d. f.)}$$

of the use of the first person in making a response in contrast to making a response which simply described what a teacher might or should have done. The assumption underlying this definition was that the use of the first person form in responding represented a measure of greater involvement than did simple description. It

will be recalled that the categories used in this analysis were Projection, Description, a combination of both Projection and Description, and No Response. Definitions and examples of these categories appear in Appendix B.

To test the hypothesis that the degree of involvement would be the same for both methods, a chi-square test of independence was used. The data involved in this analysis appear in Table 11.

Upon inspection of Table 11, it will be noted that the chi-square value of 3.058 with 3 degrees of freedom was not significant

Table 10. Difference Scores Per Episode of Initiation of Response Time by Method of Administration

Episodes	Method		y Difference
	Individual Oral	Group Written	
1	154	162	-8
2	114	123	-9
3	115	104	11
4	473	389	84
5	222	244	-22
6	82	74	8
7	263	273	-10
8	233	199	34
9	141	223	-82
10	248	249	-1
11	410	297	113
Total	2,455	2,339	118

$$\begin{array}{l}
 n = 11 \\
 \Sigma y = 118 \\
 \bar{y} = 10.72
 \end{array}
 \quad
 \begin{array}{l}
 (\Sigma y)^2 = 13,924 \\
 \frac{(\Sigma y)^2}{n} = 1265.82 \\
 \Sigma y^2 = 28,620.00 \\
 SS = 27,735.18
 \end{array}
 \quad
 \begin{array}{l}
 s^2 = 2773.52 \\
 \frac{s^2}{n} = 252.138 \\
 \sqrt{\frac{s^2}{n}} = 15.879
 \end{array}$$

$$t = \frac{\bar{y} - 0}{\sqrt{\frac{s^2}{n}}} = \frac{10.72}{15.879} = .6751 \text{ with } 10 \text{ d. f.}$$

at the .05 level of confidence. This led to the conclusion that there was no significant difference in the frequency with which the various types of involvement appeared with the individual oral and group written administrative procedures.

Table 11. Comparison of Frequencies Within the Involvement Category System by Method of Administration

Category	Method		Total
	Individual Oral	Group Written	
Projection	128	120	248
Description	34	46	80
Combination	15	15	30
No Response	10	6	16
Total	187	187	374

$\chi^2 = 3.058$ Not significant at the .05 level of confidence

It should do well at this point to recall that the major concern of the study centered around the quality of the content of the data elicited by the two methods of administration. It is to this analysis that we now turn.

From the protocols, each response to the filmed episode plus the subject's answer to question number two, "Why did you say what you said or do what you did?", was submitted to analysis using the Content category system. The response data and the answer to question two were combined because they tended to be complementary and thereby permitted a more adequate basis for categorization. Question two had direct bearing on the initial response, that is to say, it was a specific query concerning the reason or rationale for the response given.

The data coming from this categorization were used as the major test of the hypothesis that there would be no significant

difference in the nature of the content elicited by the two administrative methods. These data appear in Table 12.

Table 12. Comparison of Frequencies in the Content Category System by Method of Administration

Categories	Methods	
	Individual Oral	Group Written
Philosophy	17	17
Approach	51	51
Control	120	120
Total	188	188

Upon inspection of the data it was found that there was an identical number of response frequencies falling into the three major headings of the Content category system. This was an expected result in view of the fact that the episodes were designed around specific behavioral problems and the development of the Content category system gave added distinctness to this arrangement. For this reason, no statistical test was run on these data, and it was concluded that the content of the response to the simulator was independent of the method of administration.

In spite of the apparent similarity in category frequencies, there was still a possibility that this all-over comparison may have masked internal differences. In view of this a comparison of the frequency of responses appearing within the three major classifications was made. The data involved in this analysis appear in Table 13.

It will be seen from these data that no significant differences appeared in the sub-category frequencies within the three major categories, giving added confidence to the interpretation that content of response was independent of administrative procedure.

Table 13. Comparison of Frequencies Occurring Within Sub-Categories in the Content Category System by Method of Administration

Category and Sub-Category	Degrees Frequency		Chi-Square Value	Significance Level
	Individual Oral	Group Written		
Philosophy				
No Response	3	2		
Rigid	2	3		
Understanding	1	3		
Nurturing	11	9	1.6000	----
Approach				
No Response	0	0		
Authoritarian	14	16		
Democratic	28	29		
Child-Determined	9	6	.7542	----
Control				
No Response	9	12		
Power: Strict	38	35		
Power: Considerate	63	66		
Power: Choice	10	7	1.1515	----

In an effort to test the hypothesis concerning whether any difference existed between methods of administration on the consistency between the rationale given for a response and the response itself, analysis was done in the following way. Since question two had already been included in combination with the response on an earlier classification, it was not analyzed again. Questions one and three were separately classified by means of the same Content category system as was used in the analysis of the response and question two.

It must be pointed out at this time that in the use of the Content category system first with the response and question two, and second with questions one and three, several qualifying conditions existed. These were: (1) coder reliability was established

separately and with different coders for the two separate classifications, (2) the two classifications were separated by four month's time, with the second reliability check being made just prior to the last classification, and (3) no reference was made at any time in the second classification to the classification obtained in the first.

It can be seen from the data, presented in Table 14, that there was remarkable consistency of content orientation between episode response and the rationale given for the response between the two methods. The data were so similar in fact that no statistical test of difference was deemed necessary. This particular measure of consistency accounted for only a little over half of the 187 possibilities for consistency.

Table 14. Frequencies of Consistency Between Episode Response and Questions One and Three by Method of Administration

Response and Questions	Method	
	Consistency in Individual Oral Mode	Consistency in Group Written Mode
Response and both questions	88	92
Response and question 1	114	116
Response and question 3	106	105

DISCUSSION

Five comparisons were made of the data coming from the individual oral and group written administrations of the classroom simulation test: (1) the number of words in the response and the rationale elicited by each method, (2) the point in the episode (time) at which the subject responded, (3) the degree of involvement in the response, (4) the homogeneity of content elicited by each method, and (5) the consistency between the rationale given for a response and the response itself. The results of the study indicated in general that there was little difference in the nature or quality of the data coming from the two administrative procedures on any of these analyses. For summary purposes, however, the results of each analysis will be reviewed and discussed below.

In testing the hypothesis relating to the length of the protocols, no significant difference was found for administrative procedure with respect to the number of words used by the subjects in responding to the filmed episodes. There was a significant difference found between administrative procedures, however, in the number of words used by the subjects in answering the three questions pertaining to rationale. Answers obtained orally were consistently longer than those obtained through writing.

From these data it appeared that in terms of just responding to an episode, that is, in dealing with classroom deviations, communicating or giving instructions, one obtained as much information from a group written procedure as from an individual oral procedure. Put in another way, in order to meet certain demand situations like those required of the subjects who enacted the role of student teacher, one says what needs to be said and that is all. When it came to

telling "why" one responded as he did, however, or giving reasons for responding at a particular time in a particular way, the individual oral procedure elicited a much more lengthy response than did the group written procedure. This finding was an expected outcome in view of the difference between writing in longhand as opposed to dictation. These are noteworthy data in view of the fact that this analysis made the distinction between the response to the stimulus conditions and the rationale for that response. This distinction has not always been made in other studies, thus the individual oral procedure is often viewed uncritically as providing more extensive and often more complete data.

Length of response obviously is not the primary criterion by which one is to judge the adequacy of a particular administrative procedure, and from the results of the subsequent analyses pertaining to the qualitative aspects of the data, it should not be a criterion at all.

The second hypothesis tested related to the time of initiation of response, or the lapse of time prior to making a response. The results of this analysis indicated that no essential difference existed between the two methods of administration as to time of response initiation. This was an encouraging finding in view of the possibility of error being introduced through a possible "contagion effect", and suggests the power of the demand qualities of the episodes themselves. Apparently, the behavioral cues in the filmed episodes were of such a nature that some response was "called for" at given points in time, independent of method of administration.

The hypothesis regarding the degree of involvement in the response by each method also failed to disclose a significant difference between the two administrative procedures. Subjects were able to become sufficiently involved in the test, independent of

administrative procedure; they were able to use first person in making their responses rather than merely describing actions to be taken. There was just as much evidence in the group procedure of subjects responding to the episodes as though they were actually the teacher talking to the children as there was in the individual procedure.

The fourth hypothesis tested furnished the primary support for the conclusion that the two methods provide similar or homogeneous data. This hypothesis dealt with the content of the data, content being defined operationally in terms of three major categories: Philosophy, Approach, Control. Within the Philosophy category there were sub-categories of Rigidity, Understanding, Nurturing; within the Approach category there were sub-categories of Authoritarian, Democratic, Child-Determined; and within the Control category there were sub-categories of Power: Strict, Power: Considerate, Power: Choice. Upon analysis it was found that the two methods of administration provided essentially the same data in all major categories and in all sub-categories.

While these categories were not exhaustive of the dimensions of teacher behavior that could be tapped within the protocols, they did represent a sufficiently wide range to permit the tentative conclusion that a group written administrative procedure provides data with as much richness and power as does an individual oral procedure.

The final hypothesis tested in the study pertained to the consistency between the rationale given for the initial response and the response per se. This analysis derived from the anticipation that one or the other administrative procedure might invite inconsistency between a response and its rationale. In order to arrive at a measure of consistency between the response and the rationale, the Content category system was applied to both, independently, and

the coding for episodes was compared. Consistency was judged by the number of occurrences of identical category classifications for the response and the rationale data. This provided a measure of whether a subject's reasons for giving a certain answer were congruent with the response she made. Inconsistencies occurred when there was a discrepancy of classification for the response and for any of the answers to any of the questions. Although the data would permit a consistency-inconsistency judgment in only approximately one half of the episodes in each administrative procedure, consistency between the response and its rationale was as frequent in the written procedure as it was in the oral procedure.

On the basis of these data it seemed permissible to conclude that individual oral and group written methods of test administration elicit comparable data when used in relation to a classroom simulation test. This conclusion supports generally the results of similar investigations with more traditional projective devices (3, 9). These results have important implications for the users of tests of this nature, for the great saving of time that is possible through group administration makes practical the widespread use of such tests, especially when used for screening large numbers of subjects for particular qualities.

It should be pointed out, however, that the use of the individual method may still have greater relevance for diagnostic purposes because of the information that can be obtained through observing a subject in the process of taking such tests, or because of the opportunity to encourage and/or inquire during test administration.

Although these findings suggest the equality of the two administrative procedures, the results were conditioned by the circumstances of the study. It would do well here to point up some of the factors within the study which may have modified the results in some

way. No attempt was made, for example, to judge the adequacy of the subject's response, that is, the appropriateness of the response to the situation, its effectiveness in disciplining, or the like. In other words, the present study does not indicate how "adequate" or "relevant" a subject's responses were to the demands of the episodes. The method of analysis used in the study may not have been appropriate to make an adequate comparison of the two methods. For example, the measure of certain independent variable such as clarity, completeness, and succinctness may have yielded different results than those obtained by the category systems.

While these are relevant questions about analysis, it seems that they represent suggestions for further research rather than placing question on the results of the present study.

SUMMARY AND CONCLUSIONS

Summary

The purpose of this study was to compare the quality of data obtained from individual oral and group written administration of a classroom simulation test. The individual oral procedure involved one group of subjects taking the simulator test individually and responding orally, while the group written procedure involved another group writing their responses to the simulator test under group conditions.

The sample consisted of 34 women majoring in elementary education at Oregon College of Education at Monmouth, Oregon. These subjects were matched on the basis of age and years of teaching experience.

In each of the two different test procedures the subjects responded to 11 short filmed episodes of the classroom behavior of 22 sixth grade children. The subjects enacted the role of student teacher and were asked to respond at any time during the filmed episode where they felt some response or action was needed. Each subject was equipped with a stop watch. This was used to indicate the point in time in the sequence of events that a subject chose to initiate a response or formulate a judgment. In both test procedures the subjects were asked to respond as if they were actually the teacher talking to the children portrayed on the film.

Upon completion of each of the episodes, the subjects were asked to answer three questions: (1) What was it in this situation that made you, as the teacher, respond when you did? (2) Why did you say what you said or do what you did? and (3) What did you anticipate achieving, if anything, by the response you made? These questions were used to try and tap the rationale or reasoning for the particular response made.

The protocols, which consisted of the responses to the filmed episodes and the answers to the three questions, from the two test procedures were compared in the following ways: (1) the number of words elicited by each method, (2) the point in the episode (time) at which the subject responded, (3) the degree of involvement in the response, (4) the homogeneity of content, and (5) the consistency between the rationale given for a response and the response itself.

Analyses three, four, and five involved placement of the protocols in category systems. Reliability of category classification was demonstrated by a measure of percent agreement between two sets of raters before the protocols were classified.

The data were analyzed by the chi-square analysis and the t-test for related samples to determine the similarity of data elicited by the two administrative procedures.

Conclusions

In general, the results of the analyses indicated that the two methods of test procedure, individual oral and group written, elicited comparable data to a classroom simulation test. On the basis of these results, several conclusions seem justified:

1. The group written procedure is economically superior to the individual oral procedure because of its tendency to elicit significantly shorter protocols and because it can be administered to more than one person at one time.

2. Both methods are similar in terms of the time of initiation of response.

3. The group written procedure and the individual oral procedure are similar in facilitating involvement in the simulator test.

4. Both methods are similar in terms of the nature of the content elicited.

5. Consistency between response and rationale for the

response is independent of method of administration.

The results of the present investigation cannot claim superiority for one method over the other, but it points to the equality of both. Additional research involving the use of simulation materials under various methodological conditions needs to be done in order to further explore the potential of this type of test stimuli.

BIBLIOGRAPHY

1. Clark, Ruth Milburn. A method of administering and evaluating the Thematic Apperception Test in a group situation. *Genetic Psychology Monographs* 64: 3-55. 1944.
2. Cronbach, Lee J. *Essentials of psychological testing*. New York, Harper and Brothers, 1949. 457 p.
3. Eron, L. E. and Anne M. Ritter. A comparison of two methods of administration of the Thematic Apperception Test. *Journal of Consulting Psychology* 15: 55-61. 1951.
4. Kersh, Bert Y. *Classroom simulation: a new dimension in teacher education*. Monmouth, 1963. 101 numb. leaves. (Teaching Research, Oregon State System of Higher Education. Final report on U. S. Department of Health, Education and Welfare Grant Number 7-47-0000-164)
5. Kersh, Bert Y. Simulation in teacher education. In: *Programmed learning and teacher education: Symposium held at the Annual Convention of the American Psychological Association, St. Louis, 1962*. p. 1-12.
6. Levionian, Edward. The use of film in opinion measurement. *Audio Visual Communication Review* 10: 249-255. 1962.
7. Lindzey, G. and Shirley Heinemann. Thematic Apperception Test: individual and group administration. *Journal of Personality* 24: 34-55. 1955.
8. Metzner, H. and Floyd Mann. A limited comparison of two methods of data collection: the fixed alternative questionnaire and the open-ended interview. *American Sociological Review* 17: 486-491. 1952.
9. Pine, LaVern. A comparison of written and oral responses to a projective-picture test. Master's thesis. Stillwater, Oklahoma Agriculture and Mechanical College, 1956. 46 numb. leaves.
10. Siegel, Sidney. *Nonparametric statistics for the behavioral sciences*. New York, McGraw-Hill, 1956. 312 p.
11. Terry, Dorothy. The use of a rating scale of level of response on Thematic Apperception Test stories. *Journal of Abnormal and Social Psychology* 47: 507-511. 1952.

APPENDICES

APPENDIX A

SETTING DESCRIPTIONS FOR CLASSROOM SIMULATION TEST

Orientation Episode

It is about midmorning and Mr. Land is passing out papers to individuals in the class, calling them by name. You have just entered and you are waiting to be noticed by Mr. Land. Now Mr. Land has interviewed you previously and is expecting you to come today.

Episode 1

The students are studying at their seats while you are monitoring the class. It is around 10 o'clock in the morning, very close to recess time. Carol comes up and says something to you about this.

Episode 2

Carol has just notified you that it is five minutes past recess time. She now continues by saying something about how Mr. Land handles similar situations.

Episode 3

Remember Carol has just announced that the class is five minutes late for recess, and then added what Mr. Land does when this sort of thing occurs. Now at the beginning of this episode you have just said that the class can go immediately.

Episode 4

The class has been asked to study spelling at their seats. Mr. Land is working with a youngster across the room, and you are supposed to be helping the other children with their spelling as required.

Episode 5

The class is now engaged in seat work and a small group is working at the bulletin board changing a display. You are looking at the group at the bulletin board. Eddie, in the yellow shirt, says something to you.

Episode 6

The class is changing from committee work to regular seating arrangement. You are in charge of the class and you have just instructed them to begin moving their tables into the regular position.

Episode 7

The class is just returning from afternoon recess. They take their seats as usual in preparation for the next school activity. You

are in charge of the class.

Episode 8

You are working with a small group of children near the blackboard while the rest of the class studies. Linda is reading while the others follow along in their textbooks.

Episode 9

This is the same situation as before--in the reading group. Linda is still reading to the group.

Episode 10

This same reading group has now been in session about 20 minutes. Linda has finished reading; someone else is reading.

Episode 11

Douglas has been asked to study at his seat. He is observed leaving the reading group. The others continue their reading lesson.

APPENDIX B

INVOLVEMENT CATEGORY SYSTEM WITH EXAMPLES
FROM THE DATA

Degree of Involvement	Definition	Examples
Projection	Respondent projecting into the situation by directly addressing the children in the classroom.	"Boys and girls, will you put your things away and we'll get ready for recess." "David, let's keep out hands to ourselves."
Description	Respondent describing what she would do or not do in a situation. Her response is more like that of an observer looking in on the classroom.	I'd say, I'd ask if anybody had heard the bell. Because perhaps the clock is wrong or perhaps there had been a change in the timing for the day. I would acknowledge the fact that it was past recess time and tell her to sit down and we'd get ready.
Combination	Respondent both describes what she would do and supplies a projective response.	Well, I'd ask them, what's going on here? I know what I would do, I'm not used to having a student teacher but I would tell her this; "Right now Carole, I am your teacher and as soon as we're ready to be dismissed, I will dismiss you."
No Response	Respondent choosing not to respond.	

APPENDIX C

CONTENT CATEGORY SYSTEM WITH EXAMPLES
FROM THE DATA

DEFINITION

Philosophy

This concerns the method whereby a respondent directly interacts with a child in meeting his needs, in understanding his problems or in coping with his personal behavior problems.

 Examples

No Response

Respondent chose to wait.

Rigid

Respondents manner is abrupt, her opinions are firmly fixed and she operates solely from the adult point of view or frame of reference.

Well, if he had spoken out so it bothered the class I probably would have asked him to refrain from speaking, but apparently it's the teachers desk he's on and I think I'd be more apt to respond to that--he's not sitting where he's suppose to be, if he feels like sitting down he should be at his desk.

Understanding

Respondent empathizes with the child but still remains firm in terms of her own direction and in upholding standards of behavior.

"Why are you so tired, Eddy? Perhaps if you would bring the display to me while the others take them down, we could be through more quickly."

Nurturing

Respondent extends herself to protect or sympathize. She often tries to find out the reasons for a child's behavior.

"Eddy, why don't you go and sit at your desk for a moment and put your head down--see if that will help."

DEFINITION

Approach

The respondent provides or seeks to provide physical and/or psychological setting to facilitate learning; or dictates expectations or instructions in relation to an activity or regular routine.

 Examples

No Response

Respondent chose to wait.

Authoritarian

Respondent expects or demands unquestioning obedience; she considers herself in sole command. Despite interruptions, appeals and the like, she still maintains her own direction.

"Carole, wait until recess and then we'll discuss it."

Democratic

Respondent answers child's question or inquiry. She considers the child's point of view and interests and/or recognizes child's contribution when determining procedure. She is cognizant of children's moods and behavior.

"Thank you for reminding me, Carole. I'm sure we're all ready to take a break after studying so quietly, so you may put aside your books and quietly line up at the door."

Child-Determined

The child's or children's request or appeal ultimately determines subsequent action or response of the teacher.

"Yes--class get ready for recess now. I'd forgotten the time."

DEFINITION

Control

This is the method whereby a respondent seeks to stop or modify behavior after a deviation has occurred. Subject controls or seeks to control by imposing conditions, requirements, or by maintaining established rules for behavior and conduct.

Examples

No Response

Respondent chose to wait.

Power: Strict

Respondent uses unqualified power to initiate change. This may include scolding, criticizing or punishment. Respondent expects quick results. The obedience to the rules is the most important consideration.

"O. K. , that's enough. Johnny let's step out in the hall for a moment. "

Power: Considerate

Respondent uses explanation, or suggestion to initiate change. She utilizes a determined yet positive manner in stipulating conditions or requesting a change in behavior. Respondent may change an activity to cope with behavior deviations.

"Alright children let's get back to work and let's not do any more of that throwing. If you need to put something in the wastebasket just get up from your seat quietly and put it in. "

Power: Choice

Respondent presents a choice or alternative to initiate change in behavior. Children are often given a chance to explain their behavior or respondent may discuss such problems with those concerned.

"Boys, Johnny has been sent to his seat because he was unable to control himself in this group. Now I can't interrupt this group any longer but I think you will want to think twice before you decide to join him in his trouble making. It may be you won't want to join in his punishment. "

APPENDIX D

INSTRUCTIONS FOR CATEGORY PLACEMENT

The instructions for the classification of the responses in terms of the Involvement Category System were self-explanatory in view of the examples given in the definition and description of the system. (See Appendix B)

Instructions for the classification of responses in terms of the other category system, Content, were a little more detailed. It was necessary to tell the raters that each of the 11 episodes were filmed around specific kinds of classroom behavior. Thus, the development of the Content category system grew from the nature of the filmed episodes themselves. Some episodes dealt specifically with a child's personal problem or needs; this became the Philosophy category which was designed to cover this kind of interaction. Other episodes were concerned with the providing of structure for an activity either by requiring the respondent to give instructions or to see that the children followed through on assignments previously given. These kinds of interactions lead to the development of the Approach category. The development of the Control category resulted from a series of episodes showing behavioral deviations. That is to say, some deviations of behavior had occurred and the respondent had to deal with this in some way.

The classifying of subjects responses in terms of this Content category system was done by reading the initial response to the episode plus reading the subject's answer to question number two, Why did you say what you said, or do what you did? Then the initial response in combination with the rationale was classified under the appropriate heading according to the nature of the episode.

The classification of questions one and three in terms of the Content category system followed the same procedure as above

with the exception that each question was classified separately. All of the question ones were classified before the question threes, so there would be no carry-over effect.