A COMPARISON OF FILMS, PARTICIPANT OBSERVATION, AND SYSTEMATIC OBSERVATION IN ASSESSING BEHAVIOR

by

PATRICIA ANN WALKER

A THESIS

submitted to

OREGON STATE COLLEGE

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

June 1960

APPROVED:

[redacted]

Assistant Professor of Family Life

In Charge of Major

[redacted]

Head of Department of Family Life and Home Administration

[redacted]

Chairman of School Graduate Committee

[redacted]

Dean of Graduate School

Date thesis is presented July 24, 1959

Typed by Barbara Jean Metzger

ACKNOWLEDGMENT

The writer wishes to express thanks to Dr. Henry D. Schalock for his valuable assistance throughout this research project. To Mrs. Katherine H. Read goes appreciation for the suggestions and advice prior to the initiation for this research. To Dr. Jerome C. Li the writer would like to express appreciation for his advice concerning the statistical analysis of this study.

Margaret Shea, whose research study was conducted as a parallel study, provided support and laughter through many trying hours. Thanks also go to Jane Coverstone and Diane Goodman for help in data collection.

Dr. Jack Edling provided invaluable technical assistance in his knowledge of photography and his understanding of children's behavior. Special thanks go to him for his excellent advice and his photographic work.

TABLE OF CONTENTS

Table of Contents, continued

LIST OF TABLES

LIST OF FIGURES

A COMPARISON OF FILMS, PARTICIPANT OBSERVATION, AND SYSTEMATIC OBSERVATION IN ASSESSING BEHAVIOR

INTRODUCTION

Purpose of the Study

This study had as its purpose the exploration of the differences and similarities in four observational methods for assessing children's behavior in a free-play situation. The methods compared included participant observation, non-participant observation with two observers using systematic techniques, and film observation. The study was designed with the hope that it would provide directions and areas for further research in the field of observational methodology.

Importance of the Study

Present-day research in child behavior employs all four methods of observation investigated in this study, and there has been much speculation as to which method has the most value as a research tool. Also, there has come from continued experience with these methods, statements as to the relative strengths and weaknesses of each. At present, however, there is so little empirical evidence that is based on research on the methods themselves that statements as to the relative value of or the nature of the contribution of these methods are not meaningful. This study was designed to initiate research that ultimately will provide an answer to the question of the relative merit of each approach.

Background of the Study

There are various methods for observing behavior and events: participant or non-participant uncontrolled observation, participant observation with controls for the observer, non-participant observation with systematic controls for the observer, and observation by films. Each would seem to have its own particular advantages and disadvantages.

Observation without controls for the observer, whether participant or non-participant, is that in which the observer has no particular aspects of behavior or events to focus upon. Thus, the observer is not controlled as to what to see. There is also a lack of systematic controls as to the reliability of the observer in seeing what is to be seen and in recording what has been seen. However, an individual could not live without using observation to some extent; he observes people, witnesses events, and participates in various experiences and then draws conclusions from these.

Participant observation involves the observer's assuming an active role in the group which he is observing. The group may or may not be aware of his role as observer. The participant observer has access to the same experiences of the group and hence is able to record behavior of the group as he, a member of the group, views and experiences it. He, thus, is able to gain important insights concerning the members of the group. Participant observation makes it possible to secure data about individuals in actual life situations as the members of the group are involved. It is possible for the

participant observer to maintain an active-passive role as a member of the group by establishing rapport with the group while still retaining objectivity about the members.

When a participant observer enters a group its structure is changed; consequently, it is a different group because of the observer's presence and interaction with its members. Then, too, if the group knows about the observer's reasons for being there, the behavior of the group may be changed so that the observer receives a distorted picture of the group and its behavior. Finally, though the observer's behavior and feelings may become like that of the other group members with whom he is participating and whom he is observing, it is possible that the observer may become too emotionally involved with the group for his observations to remain objective.

In non-participant observation the group may or may not be aware of the observer's presence. The observer may be hidden behind a one-way visual screen or in an observation booth. However, if the observer is in view and remains wholly passive, the group may become suspicious and uncomfortable if his role is not clearly understood. In social research this is often solved by the observer's taking both a participant and a non-participant role for different activities.

Both non-participant and participant observers may utilize various systematic techniques for control of the observer. A definite technique for recording data in terms of time intervals, symbols for recording, a rotation schedule for observing different members, prepared observation sheets, and a decision as to the unit of behavior

to be recorded may be developed and used by the observer.

Films are a means of securing information which can later be observed, analyzed, and the information evaluated. A film can be observed innumerable times so that a behavior or an event missed the first time might be discovered on the second, third, or fourth viewing. Films, however, are an expensive means of observation, especially if there is only one specific occasion for their use. Also, there is a question as to what a film can record that an observer cannot record or what an observer can record that a film cannot.

Various questions arise as a consequence of the differences inherent in these methods. For example, which method will result in achieving the greatest quantity of information? In what situations can each method be used to its full advantage? What method is the most effective in securing the necessary information? These and other questions are subjects for research.

REVIEW OF THE RELATED LITERATURE

Observation of social behavior is probably as old as man himself. Throughout man's history, the novelist, poet, historian, and philosopher have been interested in describing and explaining behavior in terms of what they have seen. From observation comes interpretation concerning the motives and feelings behind overt behavior.

Systematic Observation

The use of a human observer as a measuring instrument in the behavioral sciences was a development of the 1920's and 1930's. According to Jersild (48, p. 472), there was a desire for a more objective technique which could aid in studying social and emotional behavior of children in natural settings. The rise of the centers for research in child development gave impetus to the use of systematic observational techniques. The scientific utilization of observation is known by various names such as controlled or systematic observation and time-sampling.

Definition of Time-Sampling

Arrington defines time-sampling as "a method of observing the behavior of individuals or groups under the ordinary conditions of everyday life in which observations are made in a series of short time periods so distributed as to afford a representative sampling of the behavior under observation. It is a method of sampling, the

validity of which is primarily a function of the amount and distribution of the time spent in observation or of the number, length, and distribution of the separate observations or time samples. It is a form of controlled observation in which the observer, the method of recording, and the manner of selecting the behavior to be observed are subject to control rather than the situation in which observations are made. It is a method whose essential function is accurate measurement of the incidence of specific behavior acts or patterns under specified conditions." (5, p. 82)

Early Pioneers in Systematic Observation

The major contributions to the development of the technique of systematic observation have come from the work of Olson, Goodenough, and Parten at the University of Minnesota and from Dorothy Swaine Thomas at Columbia University and at the Yale Institute of Human Relations. At Minnesota the studies were concerned with developing a technique to aid in the measurement of social behavior. At Columbia and Yale, Thomas and her associates studied the observer as a recording instrument and attempted to break down behavior into units which could be defined, recorded, and quantified. (5)

Olson (64) first applied the term "time-sampling" to a study of the nervous habits of children. Parten's study (65) of the social participation of children in an indoor free play situation was similar to that of Olson. However, it differed in that the time unit of observation was shortened from five minutes to one minute. Using

a more refined system and under more controlled conditions, Goodenough (34) and her graduate students measured the behavior of a group of children in free play. Behavior was recorded in terms of symbols at fifteen second intervals. Under the direction of Dorothy Thomas at Columbia University, Ruth Arrington (4) and Loomis (57) made significant steps in further development of systematic observational techniques.

Requirements of Systematic Observation

From the investigations of the earlier researchers and from the work of those following has come information concerning the technique and findings designed to make systematic observation a more adequate and reliable instrument for measurement. According to Jersild, researchers have found that the following are requirements if systematic observation is to be used properly:

"1. There must be systematic recording in objective terms of behavior which is in the process of occurring in a manner that will yield quantitative, individual scores.

2. There must be definitions in terms of overt actions of the units or patterns or contexts of behavior that are recorded and scored.

3. There must be measurement of the objectivity of the definitions which are used.

4. There must be control of the observer.

5. There must be measurement of the reliability or fidelity of the observer.

6. There must be appropriate timing and distribution of the observation periods.

7. There must be observation periods that are sufficient in duration and number to give a reliable sampling of the behavior with which the study purports to deal and for which quantitative results are reported." (48, p. 472-482)

Applications of Systematic Observation to the Study of Children's Behavior

Systematic observation has been used to measure many different aspects of children's behavior. In these studies the length of observation periods used, the time units, the category systems, the rating systems, etc., have varied considerably. Generally, systematic observation has been applied to studies of the behavior of nursery school, kindergarten, or first grade children although there are some applications of systematic observation to the behavior of older children and adults. (5, p. 87)

The following are some of the areas which have been investigated by using systematic observational techniques: physical contacts (57), aggression (29), conflict (47), ascendance (43), friendship and quarrels (22, 37), domination and integration (2), domination and submission (70), social patterns (45), initiation of social contacts (10), social participation in play (65), social adjustment (46), and language patterns (1). For a more comprehensive review, see Arrington (5), Jersild (48), Bott (18), Gellert (30), and Strodtbeck (72).

Applications of Systematic Observation to the Study of Groups

Systematic observation has been applied to diagnostic and descriptive studies of small groups of older children and adults whose members are in face-to-face contact with each other. Bales (6) has published a set of observational categories for studies concerned with the interaction and function of groups in face-to-face contact. Heyns (55) developed a category system for studying problem-solving in a conference situation. Fouiezos, Hutt and Guetzkow (55) developed a category system to observe decision-making conferences in business and industry. Carter, Haythorn, Meirowitz, and Lanzetta (55) systematically observed leadership while Polansky, Lippitt, and Redl have systematically observed behavioral contagion of a group of boys in a camp setting (55).

Types of Observer Systems

Basically, there are two observer systems which can be applied to systematic observation. These are the category and rating scale systems. "A category system consists of one or more categories or statements describing a class of phenomena into which observed behavior may be coded." (42, p. 388-389) Category systems differ in the extent to which the observer can record all overt behavior in the categories and can make inferences as to the motives and feelings of those being observed.

Rating scales have been used to describe the behavior of individuals as well as the activities of an entire group. They may be

used to record behavior at frequent intervals throughout an interaction period, or to assess the nature of an entire social event after it has ended. (42, p. 393) Rating scales are often used in exploratory studies to help refine definitions of behaviors under investigation. A rating system was used for purposes of measurement in this study. Properties of scales, their uses, etc. will be discussed in detail at the end of this chapter.

Methods of Recording Observations

In many of the studies, the units of behavior were defined in advance and the observer could then enter a check or symbol on the record each time an item of a given category occurred. (4, 64, 65) In other studies a system was used of recording behaviors simultaneously as they occurred by a series of symbols. (57, 58, 63) Mechanical aids, too, have been found to be effective. The advantages and limitations of such aids as the Bales and Gerbands Interaction Recorder, sound recordings, stenograph, and stenotype machine are reviewed by Carter (21). For a discussion of motion pictures used in observation see the section in this chapter dealing with films.

Observers have used time units ranging from five seconds to several minutes. Single observation periods have ranged from less than a minute to several hours. Scores may represent a tally of the frequency of a given response within a time unit, or they may represent the number of separate time units in which a given response has occurred.

Reliability of Observation

The problem of the reliability of observers was studied by Dorothy Thomas (76) and Budzik (19). Thomas used motion picture records as a check on observer reliability but found these to be expensive and impractical and not a true picture of reliability unless the observer has the same vantage point as the camera. Budzik found that other things being equal the fewer the behavior items or categories included in the record, the more precise the definition of these items; and the simpler the recording process, the more reliable will be the observations (19, p. 40). Other factors which affect reliability of the observers are the training of the observers, the length of the time interval, and the amount of inference of the observers in the behavior under observation.

Two methods have been used to measure the reliability of observation, percentage of agreement method and the correlation method. The percentage method measures observer agreement on each item while the correlation method measures gross agreement of the total frequency per observation period. The correlation method is deceptive when the behavior under observation occurs so rarely that the scores of some categories or points on the scale are zero. Several studies have found that observer bias may also affect the comparability of observers' data (4, 57).

Advantages and Limitations of Systematic Observation

Florence Goodenough has summarized the advantages and limitations of systematic observation as a technique.

"1. The individual is measured in terms of actual samples of his everyday behavior as displayed under ordinary conditions.

2. While the method involves different distribution of time than that of the ordinary test series, the total amount of time seems to be little in excess of that required to secure measurements of corresponding reliability by the more usual procedures.

3. The method appears to be applicable to the measurement of widely varying types of behavior; many of which have hitherto proved exceedingly difficult to express in quantitative terms.

4. Measurements can be taken without interfering with or interrupting the usual activities of the subject; by the exercise of a little care they can be taken without his knowledge.

5. It is unsafe to assume that norms or standards derived from a limited group may be applied to another group for whom conditions may be quite different even though they appear to be similar....Comparisons should be limited to the individuals included in the particular group studied." (35, p. 235)

Participant Observation

"Among the most widely practiced and least codified procedures of social research are those comprised by the large-scale collection of observation and interview data in communities." (59, p. 304) No studies of participant observation of young children in nursery school, play group, or school room settings came to this reviewer's attention. However, several field studies of children have been made. Wayne

Dennis (54) in his study of the child-rearing patterns of the Hopi Indians systematically observed the play of the Hopi children as they played under a shelter constructed to provide shade for his family. Kluckhohn's study (54) on Navaho infancy and early childhood used some controls in observation and sampling procedure as she investigated child training practices. Another study which used a combination participant-systematic approach was one conducted by Henry (54). All of the Pilaga children in the community were studied and observations were recorded as they reacted in a doll-play situation.

In discussing field methods and techniques, Oscar Lewis (54) mentions that interest in control and experiment in field work is relatively recent and that much remains to be done in the future. "Since 1930 only 7 articles dealing directly with field methods have appeared in the AMERICAN ANTHROPOLOGIST and four of these were concerned with learning native languages." (54, p. 460) In the past twenty years there has been an increased concern with procedures, use of questionnaires, schedules, and informants. Photography, taping, and recording of interviews have been used. The most recent trend in field studies is the increasing use of quantification of data and objectivity in recording of the data. Lewis suggests the increase in the use of such controls as sampling and systematic observation. He also would see more than one observer on the field so that data can be compared. All of this is necessary if participant observation is to secure meaningful, objective, and reliable data (54).

Definition and Development of Participant Observation

Kluckhohn defines participant observation as the "conscious and systematic sharing, in so far as circumstances permit, in the life activities and, on occasion, in the interests and affects of a group of persons. Its purpose is to obtain data about behavior through direct contact and in terms of specific situations in which the distortion that results from the investigator's being an outside agent is reduced to a minimum." (51, p. 331)

The greatest contribution to the development of the technique of participant observation has come from the field of social anthropology. In the study of communities and primitive societies, participant observation has been particularly applicable because of its usefulness in exploratory studies, in discovering local customs, or in developing hypotheses where there was a lack of formulation. Examples of field studies in primitive societies are the studies of Dennis (54) and Malinowski (66). Examples of field studies in community research are those of Warner (79), Whyte (59), Anderson (3), and Hollingshed (50). An example of a study using participant observation in a small group is one done by Schwartz (71).

Types of Participant Observation

Researchers have employed many variations of participant observation when studying the life of a community. An example of a study in which the researcher established a predetermined role is that of Hortence Powdermaker in her study of Cottonville (50). She obtained

permission to use the title "Visiting Teacher" from the Department of Education at Jackson, Mississippi, prior to her entrance into the community of Cottonville. John Whiting (66) in a study of the Kwoma arrived with police officials and government officials, and the natives assumed that he was a part of the government also.

Malinowski (66) lived among the Trobriand Islanders and participated in every native activity. Holmberg (66) participated in the activities of the Sirioni of Bolivia to the extent that like the natives he ate his food at night to avoid the begging of starving natives. He went with the natives hunting for food, slept in huts, and went without salt. Nels Anderson (3) lived for twelve months in Chicago's hobohemia and collected sixty life histories.

Another level of participation is exemplified by Lynd's study of Middletown (50), an American community. The research team lived in the community and participated as fully as possible. Interviewing was used. In addition, observation and interviews were supplemented by all types of documentary research. Child (50) studied second generation males of Italian origin using a combination of participant observation and systematic interviewing.

According to Schwartz (71, p. 347-350), the participant observer's role is a combination active-passive one. Although the observer never can completely separate his emotions from those of the persons in the community, he must detach himself as much as possible. He must interact with the group members but must be aware of the dangers

of too much interaction. He must realize the possibilities and limitations of his role in the community. If he fails to limit his participation, then a failure to uncover clues in data collection may result. The reason for participation is to better experience the life of those members of the community so that they can be observed with more understanding. All of the interactive roles between the observer and observed must be taken into account as the data are interpreted (78, p. 355). The participant observer must constantly evaluate himself in all instances.

Factors Which Affect the Reliability and Validity of the Data

It would seem that several factors might affect the reliability and validity of the participant observer's data. Because it is impossible to record conversation as it occurs and events in perfect chronological order, distortions in the data may result. The observer will usually have to depend upon his memory or on hastily noted words and phrases for his observational data. Bias and a lack of controls may alter the data. To counteract such tendencies, there is a feeling that a student in field work should have a theoretical grasp of field work, and he should be exposed to field work experiences early. A further help is when several students enter the same field work experience and work as a team, observing, recording, and comparing their data and checking for biases and inaccuracies (44, p. 512).

Advantages and Disadvantages of Participant Observation

The advantages and disadvantages of participant observation as a technique have been reviewed by Schwartz (71), Paul (66), Lehman (56), Miller (61), and Kluckhohn (51). The following are some of the advantages:

1. Through participation in a community or culture, the observer gains insights about the life of the people.

2. It is possible for the participant observer to interact with his subjects while still maintaining objectivity. Thus, the observer may take part in activities of the community and observe other activities.

3. Participant observation makes it possible to secure data about individuals in actual real-life situations as the members of the group are involved.

4. By participating in a community, the observer experiences the same events and feelings as the members of the community experience.

5. Data is no longer completely observational. It is supplemented by interviews, surveys, case studies, questionnaires, and so on.

Some of the disadvantages of participant observation as a technique are:

1. When the observer assumes a role in the community, the structure of the group is changed. Behavior and events may be altered because of his presence.

2. If the observer becomes too emotionally involved with his subjects, biases, judgments, and interpretations may enter into his data.

3. Often, the observer cannot record events, behavior, or conversation as it occurs and must depend on his memory of the situation.

4. The participant observer's role requires rapport with his subjects combined with objectivity.

Film Observation

Definition of Cinematography

Cinematography includes all of the applied sciences which are concerned with the recording and the reproduction of moving pictures. "A series of separate images are recorded on the same continuous light sensitive ribbon and are exposed at standard intervals of time to represent successive phases of movement. When these are exhibited in rapid sequence above the fusion of human vision, the separate images persist long enough in the mind of the observer to reproduce the appearance of continuous motion." (60, p. 91) For a complete review of scientific cinematography, its historical development and its literature, see Michaelis (60).

Application of Film Observation to the Study of Human Behavior

Gesell and his associates (60, 40) have applied the principle of cinematography to various scientific studies of human behavior. Gesell has called his technique of film analysis "cinemanalysis". (8, p. 216) "Cinemanalysis consists in the analytic study of individual frames. As a method of investigation it is made possible by five features:

1. The film is propelled at a known speed and minutely records time values and sequences.

2. Simultaneously and minutely the film records space relationships and configurations.

3. The film records these spatial and temporal data in a series of discrete, instantaneous registrations.

4. These registrations can be serially reinstated at normal, retarded, and accelerated rates.

5. Any single registration can be individually studied, in terms of time and space as a delineation pattern of a behavior event." (32, p. 4)

At the Yale Psycho-Clinic Gesell and his co-workers constructed a special domed observation hemisphere which provided one-way vision, transmission of sound, ventilation, and reflection for internal lighting. It was made of mesh screen and steel ribs through which films could be taken. A projection table was also made to help in the study of the film. The projector had a one-inch lens and could be operated backward and foreward by hand or by power. The projector was mounted beneath the table so that with a mirror it projected the film on a ground glass enameled plate set in the corner of the table top. Each movement of a subject could be timed and traced (60, p. 242-243).

Gesell's major work, AN ATLAS OF INFANT BEHAVIOR (31), contains 3,200 photographs, all single frames from his motion picture films. From this work he has derived standards and norms for child development. Thus, using cinematography as an instrument of observation and quantitative analysis, Gesell investigated infants from four to fifty-six weeks old and classified their behavior patterns, observed new-born infants through the first two weeks of life, obtained data of the normal infant through the first year of life, studied growth phenomena, compared identical twins from infancy through adolescence, and compared twins filmed at different times.

Other Film Observation Studies of Children

There have been other contributions to the method of film observation. Behrens; Langmuir, Stone and Bucher; Butlohn; and Fries (60) have all done filmed biographical records of children. At the Caroline Zachry Institute of Human Development (60), play techniques of children with blocks, paints, crayons, clay, and water were filmed. McGraw (60) at Columbia University has done many film records of children's behavior and development. Spitz (60) photographed the reactions of different babies during the first fifteen months of life and the development of the emotional life of babies. Spitz (60) also filmed the effect upon infants of prolonged separation from the mother and also compared the behavior of children. Bowlby, Robertson, and Rosenbluth (60) filmed the reaction of a two-year old, Laura, who was separated from her mother while she awaited an abdominal operation in the hospital.

In an experimental situation using projective play techniques, Langmuir, Stone, and Bucher (60) filmed the use children made of a stick given to them after toys had been taken away. Escalona and Leitch (60) at the Menninger Foundation filmed infants eighteen through twenty-five weeks old after they were exposed to prolonged perceptual stimulation.

Advantages and Disadvantages of Film Observation

The application of cinematography to the study of human behavior has various advantages and disadvantages (39, p. 119-120; 60, p. 9-12).

The following are some of the advantages of film observation as a technique:

1. Using an instrument to record observable behavior make possible precise investigations of behavior.

2. "In thoroughness of detail, in scope, objectivity, permanence, and trustworthiness, the cinema record far outclasses the human observer." (39, p. 119)

3. The camera records every detail of behavior within its range and never lapses in its attention to the scene which it is recording.

4. The camera is indifferent and disinterested in that which it is recording.

5. The film can be viewed an unlimited number of times and can be projected at its normal speed, slow speed, or in successive still frames.

6. Patterns of behavior can be translated into time and space values. Thus, "time sampling is inherent in cinematographic records". (60, p. 10)

7. The camera can be concealed, will operate automatically, and is almost insensitive to atmospheric and climatic changes.

Various disadvantages of the film as a technique of observation can be noted also:

1. The camera position must be chosen by the camera man. The position of the camera as it is photographing behavior is a most relevant factor. Thomas (74) found when using films as a check on observer reliability that unless the observer has the same perspective as the camera, the comparison of the film and the observer records does not provide a true picture of observer reliability.

2. Often, the time interval which must lapse between development of the film record and actual projection is a long one.

3. Photographic equipment and film are rather expensive. Actual filming requires time, effort, and some training.

4. Often a film record of behavior is impractical.

Rating Scales

Definition of Rating Scales

A graphic rating scale system was used in this study to assess the behavior of children in a free-play situation. The process of rating may be defined as "an assigning to a person or datum a rank, score, or mark that implies some quantity. The rank, score, or mark may assert the presence or absence of a quality called a qualitative rating or it may assign a place on a quantitative scale. A rating scale is a device by which a rater can record the estimated magnitude of the trait or quality rated." (25, p. 437-438)

Types of Rating Scales

Graphic Rating Scales. The graphic rating scale is a form for recording a rating according to the magnitude of some quality or trait. "It consists of a straight line representing the entire range from a conceivable maximum to a conceivable minimum strength of the variable in question. The relative strength of the specimen being rated is indicated by a check mark at a proportional distance from the ends. Scale points with brief descriptions may be indicated along the line to serve as a guide to the judge in localizing his rating." (25, p. 437-438)

Itemized Rating Scales. In this type of rating scale the rater selects one of a limited number of well-defined categories which are ordered in terms of their scale positions (44, vol. 1, p. 203).

Comparative Rating Scales. The comparative rating scale consists of a scale in which "a ratee is compared with a group of individuals who have been chosen to exemplify different degrees of a particular trait. The ratee receives the score or rank of the person he most resembles in the trait in question." (25, p. 437-438)

Use of the Rating Scale

The rating scale can be used on the spot as an observation or interview is taking place, or it can be used after subjects of a study have responded to the interviewer. Ratings can be used with interviews, observations, projective materials, and questionnaires.

Rating scales have been used to rate elementary teachers (77), personality (69), training and experience of employees (20, 62), self-ratings vs. group ratings (80), relationships between insight and measures of projection and distortion (26), and behavior of a pupil (11, 24, 67).

The Validity of Rating Scales

In the 1950's the United States Armed Forces began an evaluation of the use of the rating scale in its personnel training program. A research program involving 400 student officers was so designed as to study the effect of different rating techniques, rater characteristics and rating conditions on the validity of efficiency ratings. No differences in the validity of the ratings were found when: raters were required to identify themselves as opposed to remaining anonymous, a variety of graphic scales and forced-choice forms were used, lenient

raters were contrasted with hard raters. Ratings were more valid if: made earlier in a rating session involving several ratees rather than at the end of the session, an average of a number of ratings was used rather than a single rating per ratee, the raters were high in ability and achievement (9).

The Reliability of Rating Scales

A. W. Bendig (12, 13, 14, 15, 16, 17) has done many research studies involving rater reliability and other factors. In a study of rater reliability and judgmental demoralization, Bendig found that judgmental demoralization results when the total rating task presented to the judge is too extensive or complex (13). In a study of rater reliability and heterogeneity of scale anchors, Bendig found that, when rating for preference value two lists of ten foods each which varied in the heterogeneity of the verbal anchors defining the end categories, raters were significantly less reliable for the more homogeneous list and more reliable when the end anchors were heterogeneous (12).

Bendig also studied judgmental fatigue and rater reliability. One hundred twenty subjects were asked to rate a total of forty-five stimuli for preference value using a nine-point scale. The forty-five stimuli were randomly divided into individual lists containing 10, 15, or 20 foods and each group received one of the possible orders of list length. Rater reliability was significantly different between the individual lists but was not affected by either the

length of the list or by the temporal order of the list in the series. It was concluded that judgmental fatigue does not affect rater reliability or bias when the subjects report food preference self-ratings (14).

In studying reliability and the number of rating scale categories, Bendig used two hundred thirty-six subjects to rate 20 foods as to preference using 2, 3, 5, 7, and 9 categories. Test reliability (summed ratings for each subject) was constant over the entire range of categories. Rater reliability (summed ratings for each food) was constant from 5 to 9 categories but was slightly lower at 2 and higher at 3 categories (15).

In further studies (16, 17), Bendig found that rater bias tended to decrease with longer scales (7 and 9 categories) and rater reliability increased with longer scales (7 and 9 categories) and with higher educational levels.

Effect of Number and Order of Ratings on Reliability and Validity

In a study by Karcher and King, the first rating was consistently most reliable and the fourth rating made in the same period was consistently least reliable and valid. There was a sizeable increase in the reliability estimate for four ratings over the reliability of any single rating. There was a sizeable increase in the validity estimate based on the sum of four ratings over that of the first (49).

Structure of Rating Scales

Guilford suggests the following rules for scale construction:

"1. Each trait should occupy a page by itself.

2. The line should be at least five inches long but not much longer.

3. The line should have no breaks or divisions.

4. The "good" and "poor" ends should be alternated in random order so as to avoid a constant motor tendency to check at one side of the page.

5. Use three or five descriptive adjectives--two extremes and one or three intermediates.

6. The descriptive phrases should be in small type with considerable white space between them.

7. Only universally understood descriptive terms should be used.

8. Decide beforehand upon the probable extremes of ability to be found in the group or groups in which the scale is to be used.

9. The end phrases should not be so extreme in meaning as to be avoided by the raters.

10. Have the extreme phrases set flush with the ends of the lines.

11. The average or neutral phrases need not be evenly spaced.

12. Descriptive phrases need not be evenly spaced.

13. In the scoring use a stencil which divides each line into several sections to which numerical values are assigned.

14. The divisions of the scoring stencil need not be equal.

15. Do not require any finer distinctions in rating than are used in scoring." (38, p. 271-272)

In a study by Howard Clarke (23), it was found that physical features of a scale which facilitate recall of the actual behavior

will increase the accuracy of ratings. Longer objective descriptive statements will be more effective than simple phrases in defining the steps on an adjectiveal type rating scale.

The Rater

In a study concerning the training of raters Delmar Landen found that raters will vary in the accuracy of ratings given in direct proportion to the number of previous relevant contacts with the ratee. The rater will make more accurate ratings when he has been forewarned concerning the types of activity to be rated since this will facilitate his more properly focusing attention on such behavior (53). Falk and Bayroff (27) found that when the same rater was involved agreement was greater if the same technique was employed than when different techniques were used. When different raters were employed, it made no difference whether the same or different techniques were used.

The Scale Used in This Study

The rating scales used in this study were graphic scales especially designed for measuring the behavior of children in a free-play situation. The behavior assessed consisted of five main variables: tension, avenues used in the expression of tension, body movement, outgoingness, and purposefulness. These five variables were subdivided further until there were 20 variables to rate. Each variable was placed on a four-, five-, or six-point continuum with brief descriptions of the scale points along the line to serve as a guide for the rating. For a more complete description of the rating scale

see Appendix "A" and the discussion pertaining to the measuring instrument in the section entitled Method and Procedures.

The rating system for measuring behavior of children was used for all three of the approaches to observation described above. It was used within the framework of systematic observation, participant observation, and film observation.

METHOD AND PROCEDURES

Subjects

Sixteen children were chosen to participate in this study. These children were selected because of their age and availability. All were registered by their parents as applicants for enrollment in the Oregon State College Nursery Schools. At the time of observation, the children were between the ages of ten months and thirty-four months.

The subjects were divided into four groups consisting of four children each. The first group was composed of ten to twelve month old children, the second group of sixteen to seventeen month old children, and the third group of twenty-three to twenty-five month old children, and the fourth group of thirty-two to thirty-four month old children. The age range within each group was kept small so that the behavioral characteristics within each group would be more homogeneous.

From each of the four age groups one child was randomly selected to be observed during his scheduled observation hour. Each of the four children was observed by four variations of observational methods.

Procedure

Contact of Subjects. The parents of the children in the study were initially contacted by a letter (see Appendix "B") explaining the research project and asking the parents to consider the

possibility of their child's becoming a subject for the study. This letter was followed by a telephone call. At that time, any question of the parent was answered, and a decision was reached as to whether the parent would cooperate in the study. If the parent agreed to participate, the mother was told the date and time which she and her child were to be at the nursery school. Each mother was asked to accompany her child and stay with him during the observation hour. This was done in order to give the child the security of his mother's presence in an unfamiliar nursery school setting with strange children and adults.

Nursery School Setting. Each group of four children with their mothers came to the nursery school laboratory for one hour of free play. The nursery school laboratory was used as the setting for the free-play situation. Around the room various toys and equipment attractive to children were placed. Since the playroom was large, the room was divided by a series of folding screens so that a space approximately 20' x 16' was utilized by the children for play. The mothers were seated in chairs at one side of the room so that they could communicate easily with the children and so that the children could readily reach their mothers if they wished.

In a small alcove adjoining the playroom, the systematic observer, camera observer, and the photographer with his camera were seated. The observers in the alcove were separated from the play group by a series of waist-high folding screens. The observers and camera were

visible to the children at all times, but interaction between the two groups did not take place.

A participant observer was at all times with the children in the playroom. This observer actually had contact with the children while at the same time observed the child in the total situation.

The Measuring Instrument for Assessing Behavior. A system of graphic rating scales was constructed for this study to measure the children's behavior. A rating scale was chosen as the method of data collection for several reasons:

1. The study was to be exploratory in character and the rating scale is often used as a means of "obtaining the experience which will be helpful in formulating relevant hypotheses and for discovering the possibilities for different types of research." (44, vol. 1, p. 33)

2. "Rating scales are probably one of the most commonly used forms for scaling traits and attributes." (68, p. 227)

3. The rating scale is a relatively simple instrument by which one can observe and record the presence or absence of a given behavior.

4. The rating scale was suitable and easily adaptable to all of the methods of observation employed in this study.

The children's behavior was measured in terms of five variables: tension, avenues used in the expression of tension, body movement, outgoingness, and purposefulness. Each of the main variables was subdivided until the complete Behavior Rating Scale (see Appendix "A") consisted of twenty variables of behavior. Each of the variables was placed on a 4, 5, or 6 point continuum with brief descriptions of the scale points along the line to serve as a guide for the observer's rating of the behavior. The following is a skeletal outline of the

five main variables and the subdivisions of each. (See Appendix "A" for a further definition and explanation of the variables and rating scale.)

I. Tension

1. Extent to which a child evidences tension during the course of the observation

2. Extent to which the situation is inferred to be tension-producing

II. Avenues used in the expression of tension

3. Bodily expression

4. Vocal expression

5. Physical aggression expressed in positive, acceptable (non-destructive) ways toward materials appropriate for this expression

6. Physical aggression expressed through attack on materials

7. Aggression through attack on people

8. Expression of tension through the seeking of nurturance

9. Expression of tension through showing-off behavior

III. Body Movement

10. Speed of movement

11. Expansiveness of movement

IV. Outgoingness

12. Frequency with which a child initiates interaction with materials

13. Frequency with which a child initiates interaction with adults

14. Frequency with which a child initiates interaction with children

15. Proportion of adult overtures to interaction accepted by a child

16. Proportion of child overtures to interaction accepted by a child

17. Amount of time spent in interaction with materials

18. Amount of time spent in interaction with adults

19. Amount of time spent in interaction with children

V. Purposefulness

20. Purposefulness

In addition to the fact that the scales varied as to the number of scale points, fourteen of the twenty were single dimensional needing only a horizontal rating (see Figure 1) and six scales required the rater to make both a horizontal and vertical rating (see Figure 2), or two-dimensional rating.

6. Physical Aggression Expressed Through Attack on Materials

Much use of this avenue (40 - 100%)	Considerable use of this avenue (26 - 39%)	Some use of this avenue (11 - 25%)	Slight use of this avenue (1 - 10%)	No use of this avenue

FIGURE 1.
A Single Dimensional Rating Scale with a Horizontal Dimension

2. Extent to Which the Situation is Inferred to be Tension-Producing

	All of the time (99-100%)	A large proportion of the time (40 - 99%)	A considerable proportion of the time (26 - 39%)	Some of the time (11 - 25%)	A small proportion of the time (1 - 10%)	None of the time
Stimuli are extremely tension-producing						
Stimuli are tension-producing						
Stimuli are somewhat tension-producing						
Stimuli are slightly if at all tension-producing						

FIGURE 2.

A Two-Dimensional Rating Scale with Both Vertical and Horizontal Dimensions

Methods of Observation. Four variations of observational methods were employed in this study. These variations included the participant observer, the non-participant systematic observer in the observation alcove, the non-participant observer with the camera, and the impartial coder of the films. All four observers used the Behavior Rating Scale in their assessment of the child's behavior.

From each of the four age groups, one child was randomly selected to be observed during his scheduled observation hour. During this hour each of the observers concentrated on the selected subject.

The participant observer assumed an active role with the group of children initiating activities when it seemed necessary and observing the child in the total situation. After each observation hour the participant observer rated the subject according to what had been observed over the entire hour. Each child selected for observation from the four different age groups was rated once by a participant observer.

At each of the four observation periods, a camera was located in the observation alcove and focused on the child selected for observation during each of the four periods. The camera recorded each of the subjects' behavior for a period of twelve minutes. The twelve-minute sample of the child's behavior was staggered throughout the entire observation hour. The film of the child was later analyzed and rated by an impartial film observer who was unfamiliar with the child and the free-play situation. The film observer, in assessing the child's behavior from the film, used the Behavior Rating Scale

as did the other three observers.

As the camera photographed the behavior of the subject selected for filming, a non-participant observer stayed close to the camera and observed the child selected for observation from the same vantage point as the camera. This non-participant observer observed only as the camera was in operation or for the twelve-minute period in which the camera was sampling the child's behavior. The camera observer used the Behavior Rating Scale also.

In summary, a child was selected randomly from each of the four age groups for purposes of observation. During this observation, his behavior was rated by a participant observer, a non-participant systematic observer in the observation alcove, and a non-participant observer stationed with the camera. A film of the child was taken at this time and was later analyzed and rated by a film observer. In order to make it possible to compare the differences and similarities between the four different ratings, the film observer, non-participant observer with the camera, and the participant observer rated each of the four children only once, while the non-participant systematic observer rated each of the four children three times with the three ratings being pooled and averaged. For comparison purposes, then, there was one rating for each of the four methods of observation for each of the four subjects selected for study.

Observation Schedule. Each group was scheduled to come to the nursery school laboratory for one hour of observation. From each group one child was randomly selected to be observed. Each observer

had a different observation role for each hour of observation; that is, no observer was camera observer more than once, participant observer more than once, etc.

A system of rotation for observation was devised as shown in Figure 3. This system of rotating the observers was necessary in order to compare the differences and similarities in the ratings obtained by the four different methods of observation used. The different methods of observation that were used in the study are shown along the top by NPSO, the non-participant systematic observer; PO, the participant observer; CO, the camera observer; and FO, the film observer. The letters A, B, C, and D shown within the large squares illustrate the rotation of the four observers of the study. Each observer assumed a different role for each observation hour.

		Methods of Observation			
		NPSO	PO	CO	FO
Children and Age Groups	1 a	A	B	C	D
	2 b	D	A	B	C
	3 c	C	D	A	B
	4 d	B	C	D	A

FIGURE 3.
Observation Rotation System

Along the side of Figure 3 are arabic numerals and lower-case letters a, b, c, and d. The numerals 1, 2, 3, and 4 stand for the four children randomly selected from the age groups for observation. The letters a, b, c, and d stand for the different age groups of the study: a, 10 to 12 months; b, 16 to 17 months; c, 23 to 25 months; and d, 32 to 34 months.

Reliability of Observers. Reliability of observation was established on three groups of children who came to the nursery school laboratory for observation. These children were only used for purposes of establishing reliability. The setting of the nursery school was the same as that of the study. All of the children used in the reliability study were chosen from those registered for Oregon State College Nursery Schools.

Each of the three groups consisted of three children with the first group being made up of ten and eleven-month-old children; the second, sixteen to seventeen-month-old children; and the third, twenty-seven to twenty-nine-month-old children. It was assumed that if the observers were reliable for a group from twenty-seven to twenty-nine months of age, they would also be reliable for a group from twenty-three to twenty-five months of age and a group thirty-two to thirty-four months of age since twenty-seven to twenty-nine months is near the middle of both of these age groups.

Reliability scores are presented in Tables I, II, and III. Table I indicates Individual Category Reliability for the three age groups, and Table II shows Individual Observer reliability for the

three groups. Table III indicates Average Total Observer Reliability for the three age groups. The reliability scores as presented in Table I were computed for each age group and for each individual variable of the Behavior Rating Scale. In determining reliability, it was decided to compute percent of agreement from complete agreement, when observers checked the same scale point, and to compute percent of agreement from an agreement based on disagreement by one scale interval. The formula used for these computations was:

$$\text{Percentage of agreement} = \frac{\text{Number of agreements}}{\text{Agreements plus Disagreements}}$$

Table II shows the individual observer reliability scores according to every possible combination of two observers. Here, too, reliability was computed for complete agreement and agreement when differing by only one scale interval. The same formula was utilized for these computations.

Table III indicates the average total observer reliability for the entire Behavior Rating Scale. Here, each of the reliability scores of the observers for the variables was totaled and averaged. These figures were used in computing the observer's total reliability for the rating scale. Complete agreement and agreement when differing by one scale interval were computed.

It will be noted in Table I that the reliability scores for several of the variables are low, indicating a lack of understanding of the variables by the observers. The scores on these variables were not as valid or satisfactory as scores on those variables indicating

higher agreement. With this fact in mind, it was decided to keep these variables showing low reliability in the study, remembering that the data for these variables would be questionable.

TABLE I.
INDIVIDUAL CATEGORY RELIABILITY

VARIABLES	AGE GROUPS					
	10 - 12 months		16 - 18 months		27 - 29 months	
	Complete Agreement	Disagree. by 1	Complete Agreement	Disagree. by 1	Complete Agreement	Disagree. by 1
1. Extreme tension	88%	96%	100%	100%	100%	100%
2. Considerable tension	88	88	83	100	79	87
3. Some tension	38	58	66	83	41	75
4. Little or no tension	38	41	66	100	58	88
5. Stimuli are extremely tension-producing	88	88	100	100	100	100
6. Stimuli are tension-producing	46	70	100	100	66	87
7. Stimuli are somewhat tension-producing	16	50	89	100	41	96
8. Stimuli are slightly if at all tension-producing	12	41	66	83	62	66
9. Bodily expression of tension	41	50	83	83	16	33
10. Vocal expression	46	71	66	83	66	87
11. Physical aggression expressed in positive, acceptable ways towards materials which are appropriate for this expression	75	88	100	100	87	100
12. Physical aggression expressed through attack on materials	88	100	100	100	100	100
13. Aggression through attack on people or materials	88	100	83	100	58	75
14. Expression of tension through seeking of nurturance	79	83	100	100	75	87
15. Expression of tension through "showing-off" behavior	83	79	66	100	100	100
16. Speed of movement	50	100	100	100	46	100
17. Expansiveness of movement	100	100	61	100	58	100

Table I, continued

VARIABLES	AGE GROUPS 10 - 12 months Complete Agreement	10 - 12 months Disagree. by 1	16 - 18 months Complete Agreement	16 - 18 months Disagree. by 1	27 - 29 months Complete Agreement	27 - 29 months Disagree. by 1
18. Frequency with which a child initiates interaction with materials	37%	83%	33%	55%	71%	100%
19. Frequency with which a child initiates interaction with adults	25	75	77	100	33	87
20. Frequency with which a child initiates interaction with children	33	75	77	100	66	96
21. Proportion of adult overtures to interaction accepted by a child	79	83	83	83	50	50
22. Proportion of child overtures to interaction accepted by a child	58	62	77	83	79	79
23. Amount of time spent in interaction with materials	62	88	22	66	54	75
24. Amount of time spent in interaction with adults	66	83	38	100	33	75
25. Amount of time spent in interaction with children	50	58	66	100	41	87
26. Level of involvement with materials	29	50	38	66	33	75
27. Level of involvement with adults	41	75	38	60	25	29
28. Level of involvement with children	25	38	55	71	29	38
29. Extreme persistence	96	96	100	100	100	100
30. Persistence	71	71	100	100	75	75
31. Some persistence	66	71	83	83	46	50
32. Little or no persistence	88	88	83	83	46	50
33. Extreme frustration	91	91	100	100	100	100
34. Frustration	71	71	55	60	75	75
35. Some frustration	46	46	83	83	33	38

TABLE II.
INDIVIDUAL OBSERVER RELIABILITY

OBSERVER COMBINATIONS	AGE GROUPS					
	10 - 12 months		16 - 18 months		27 - 29 months	
	Complete Agreement	Disagree. by 1	Complete Agreement	Disagree. by 1	Complete Agreement	Disagree. by 1
A and B	60%	80%	67%	83%	64%	79%
A and C	53	71	66	83	54	75
A and D	55	71	60	82	67	87
B and C	60	70	86	95	60	76
B and D	58	69	82	93	65	77
C and D	68	83	84	93	57	76

TABLE III.
AVERAGE TOTAL OBSERVER RELIABILITY

	AGE GROUPS					
	10 - 12 months		16 - 18 months		27 - 29 months	
OBSERVERS	Complete Agreement	Disagree. by 1	Complete Agreement	Disagree. by 1	Complete Agreement	Disagree. by 1
A	55%	74%	64%	83%	61%	80%
B	59	73	78	91	63	74
C	60	75	78	90	57	75
D	60	74	75	87	62	80

RESULTS

The purpose of this study was to compare the differences and similarities in four methods of observation used to assess the behavior of children in a free-play situation. The four observational methods that were compared were participant observation, non-participant systematic observation, film observation, and non-participant observation from the same vantage point as the camera.

Statistical Treatment of the Data

A 4 x 4 Latin Square design was used in the statistical analysis of the data. This design is illustrated in Figure 4. In this figure, the columns marked 1, 2, 3, and 4 pertain to the four children of the study who were randomly selected for observation. The rows lettered SO, PO, CO, and FO indicate the four methods of observation employed.

METHOD	CHILDREN			
	1	2	3	4
SO	B	C	A	D
PO	C	A	D	B
CO	D	B	C	A
FO	A	D	B	C

FIGURE 4.
The 4 x 4 Latin Square Design Used in This Study

It will be recalled that a system of rotating the observers was devised so that no observer assumed the same role twice. Each child was exposed to all of the observers but in a different serial order from the other children. The letters A, B, C, and D represent the four different observers in the study. These are arranged by rows and columns to illustrate the system of rotation.

Average scores for each of the 20 variables appearing in the rating scale were found for the four methods of observation. An analysis of variance was applied to these data to determine whether differences appeared in the average scores for the different approaches to observation. Duncan's new multiple range test was used to identify the main source of variation whenever an F-value resulting from the analysis of variance was significant.

Results of the Analysis

The data coming from the analysis of variance appear in Table IV. Five of the twenty variables proved to vary significantly between observational methods. These were: Extent to Which a Child Evidences Tension During the Course of the Observation, Extent to Which the Situation is Inferred to be Tension-Producing, Bodily Expression of Tension, Frequency with which a Child Initiates Interaction with Children, and Amount of Time Spent in Interaction with Materials. A detailed analysis of these variables appears in Tables V through IX.

TABLE IV.
BEHAVIORAL VARIABLES, THEIR F-VALUES, AND LEVEL OF SIGNIFICANCE FOR METHODS OF OBSERVATION

	Names of Variables	F-Values
1.	Extent to which a child evidences tension during the course of the observation	5.56*
2.	Extent to which the situation is inferred to be tension-producing	9.19*
3.	Bodily expression of tension	4.39*
4.	Vocal expression of tension	1.64
5.	Physical aggression expressed in positive, acceptable (non-destructive) ways towards materials appropriate for this expression	1.00
6.	Physical aggression expressed through attack on materials	1.00
7.	Aggression through attack on people	.09
8.	Expression of tension through seeking of nurturance	.54
9.	Expression of tension through showing-off behavior	1.00
10.	Speed of movement	2.40
11.	Expansiveness of movement	.64
12.	Frequency with which a child initiates interaction with materials	.50
13.	Frequency with which a child initiates interaction with adults	1.14
14.	Frequency with which a child initiates interaction with children	8.90*
15.	Proportion of adult overtures to interaction accepted by a child	1.26
16.	Proportion of child overtures to interaction accepted by a child	.64
17.	Amount of time spent in interaction with materials	10.30**

Table IV, continued

Names of Variables	F-Values
18. Amount of time spent in interaction with adults	1.08
19. Amount of time spent in interaction with children	.36
20. Purposefulness	2.17

* Significant at five percent level.

** Significant at one percent level.

The mean scores for the variables appearing in Tables V through IX are arranged according to size. Means which are not significantly different as measured by Duncan's test are underlined; means which are significantly different from each other are not underlined. Thus in Table V the mean score of the systematic observer is significantly higher than the camera observer and the mean score of the participant observer is significantly higher than those of the film observer and the camera observer. The mean scores of the film observer and systematic observer, camera observer and film observer, and systematic observer and participant observer are not significantly different.

It will be seen from these data that no consistent trends or patterns appear for methods. For the variable Extent to Which a Child Evidenced Tension During the Course of the Observation, the participant observer had the highest mean score while the camera observer had the lowest mean score. For the variable Extent to Which the Situation is Inferred to be Tension-Producing, the

systematic observer had the highest mean while the film observer had the lowest. This lack of patterning characterizes the data throughout.

TABLE V.
EXTENT TO WHICH A CHILD EVIDENCED TENSION
DURING THE COURSE OF OBSERVATION

	METHODS OF OBSERVATION			
	CO	FO	SO	PO
MEANS	1.775	2.325	4.250	4.450

TABLE VI.
EXTENT TO WHICH THE SITUATION IS INFERRED
TO BE TENSION-PRODUCING

	METHODS OF OBSERVATION			
	FO	CO	PO	SO
MEANS	1.500	1.750	2.575	3.300

TABLE VII.
BODILY EXPRESSION OF TENSION

	METHODS OF OBSERVATION			
	SO	CO	PO	FO
MEANS	1.425	1.550	1.900	2.300

TABLE VIII.
FREQUENCY WITH WHICH A CHILD INITIATES
INTERACTION WITH CHILDREN

	METHODS OF OBSERVATION			
	SO	FO	CO	PO
MEANS	1.625	1.725	1.800	1.975

TABLE IX.
AMOUNT OF TIME SPENT IN INTERACTION WITH MATERIALS

	METHODS OF OBSERVATION			
	FO	SO	PO	CO
MEANS	1.150	1.350	2.200	2.250

DISCUSSION

Discussion of the Findings

The purpose of this study was to compare four observational methods for assessing the behavior of children in a free-play situation. Of the twenty aspects of behavior studied, only five differed significantly between methods. From these results it would seem that there would be no particular advantage in choosing one method over another, at least as far as these results were concerned. However, there are instances in which time, money, and effort could be saved if a particular method were chosen. In terms of these results two examples can be cited in which a choice of one method would seem more advantageous than a choice of another. First, instead of asking a systematic observer to observe the behavior of a nursery school group the teacher could assume the role of a participant observer and could obtain essentially the same results. Second, a systematic observer might be used to gain information about the behavior of a group rather than attempting to film that behavior.

In reviewing the results, it is apparent that there is no consistent pattern in the direction or trend of the data; that is, one method does not consistently yield a higher mean score or a lower mean score than any other. However, the five variables which were significantly different for methods of observation do provide some interesting bases for speculation. It is important to look at these differences and consider what might explain or account for them.

The first difference which is noteworthy is the difference between the participant observer's mean scores and the scores of the other three observers. In no case was the participant observer's scores significantly lower than the scores for the other methods. In comparison with the film observer, the participant observer's scores were significantly higher for four of the five variables. The participant observer also had higher mean scores than the camera observer and the systematic observer on two variables. These results may be accounted for by two advantages which the participant observer had over the other three observers. First, was the length of observation time. The participant observer observed an hour while the other three observers observed a maximum of fifteen minutes during the hour. This greater sampling of behavior clearly could lead to the observed differences in variables such as Amount of Time Spent in Interaction with Materials or the Frequency with Which the Child Initiates Interaction with Children. A second advantage was the opportunity for the participant observer to interact with the child under observation, whereas the other three observers never had direct contact with the child. This advantage would seem particularly great when it came to observing tension or other variables involving a multitude of cues which may be picked up only through close contact with a child.

A second difference that is noteworthy is that between the film observer and the camera observer. It will be recalled that both of these observers observed the child's behavior for the same period of

time, the only difference being that the film observer watched a film of the child's behavior while the camera observer watched the child as the camera was in operation. As a consequence of this design whatever differences observed would have to be a function of chance, error variance, or method. Interestingly enough, the differences observed between these two approaches were not great, as one variable for each approach had a significantly greater mean score than the other. On this basis it would seem that an observer cannot gain any more information about behavior by observing a film than could an observer observing a child in an actual free-play period for the same amount of time.

A look at the two variables which differed significantly between methods leads to a slightly different conclusion. The film observer had a significantly higher mean score than the camera observer on the variable Bodily Expression of Tension. This is a characteristic of behavior which is not always easy to make quick judgments about, but which is relatively easily captured on film. With the film the observer could see this behavior many times and thereby perhaps make a more accurate judgment than could the camera observer who had to make much quicker judgments about the same behavior. It is also interesting to note that this same relationship held between the film observer and the systematic observer for Bodily Expression of Tension.

This same reasoning should hold for the variable Amount of Time Spent in Interaction with Materials, but this was not the case as the camera observer obtained a significantly higher mean score on

this variable than did the film observer. From these results, the soundest interpretation may be that the differences observed between the camera and film observers were a function of chance or error variance.

The third difference that is noteworthy is the difference between the systematic observer's data and the data coming from the film observer and the camera observer. The mean score for the systematic observer was significantly higher than the mean scores of both the film observer and the camera observer for the variable Extent to Which a Situation is Inferred to be Tension-Producing, and higher than the camera observer for the variable Extent to Which a Child Evidences Tension during the Course of Observation. In rating the variables relating to tension, the systematic observer had the advantage of seeing the entire free-play situation and thereby was able to utilize in his rating the many cues which were outside the range of the camera and, theoretically, the camera observer. In no instance were the systematic observer's scores higher than those of the participant observer.

Limitations of the Study

Three limitations of the study seem paramount when considering reasons for the lack of pattern observed among the four methods of observation. The first has to do with the use of only four children in the study. Subjects were limited to this small number because the study was an exploratory one and because of the cost of filming.

Certainly further research along these lines will require a larger number of participants.

A second limitation has to do with the shortness of the time sample. Had the number of hours of observation been greater the likelihood of finding significant differences between methods would have increased appreciably.

The third limitation has to do with the measuring instrument. It is possible that it did not measure what it purported to measure and surely the reliability of measurement was not high. Confidence cannot readily be placed in observational data when observers were no more reliable than they were for this study.

Research Implications

One of the purposes of this study was to identify possibilities for further research in methods of observation. Several directions seem justified on the basis of this study.

1. A careful comparison of participant and systematic observation to see if the advantages suggested by this research for participant observation are real.

2. A careful comparison of film observation with participant and systematic observation to see if the disadvantages suggested by this research for film observation are real.

3. The development of a valid and reliable means for measuring differences between observational methods. Until such an instrument is developed, it is possible to obtain only a rough estimate as to

the relative value of or the contribution of the various methods under study.

SUMMARY AND CONCLUSIONS

Summary

This study had as its purpose the exploration of the differences and similarities in four observational methods for assessing children's behavior in a free-play situation. The methods compared included participant observation, non-participant observation with two observers using systematic techniques, and film observation.

Sixteen children were chosen to participate in the study. The children's ages ranged between ten months and thirty-four months. The subjects were divided into four groups consisting of four children each: Group 1, ten to twelve months; Group 2, sixteen to seventeen months; Group 3, twenty-three to twenty-five months; and Group 4, thirty-two to thirty-four months. Each group of four children came to the nursery school laboratory for one hour of free play. From each of the four age groups one child was randomly selected to be observed during his scheduled observation hour. Observation of these four children provided the data for the study. Reliability of observation was demonstrated for all observers under systematic observation conditions. Observer reliability was found by computing the percentage of agreement between observers on an item-by-item comparison of ratings made simultaneously but independently.

During the child's scheduled observation, his behavior was rated by a participant observer, a systematic observer in the observation alcove, and an observer stationed with a camera. A film of the child's

behavior was taken at this time and was later analyzed and rated by a film observer. Each observer had a different observation role for each hour of observation; that is, no observer was camera observer more than once, participant observer more than once, etc. The children's behavior was measured in terms of five variables: tension, avenues used in the expression of tension, body movement, outgoingness, and purposefulness. In assessing the behavior of the child, the observers each used a system of graphic rating scales designed to measure twenty aspects of behavior. For comparison purposes, there was one rating for each of the four methods of observation for each of the four subjects of this study.

A 4 x 4 Latin Square design was used in the statistical analysis of the data. Average scores for each of the twenty variables appearing in the rating scales were found for the four methods of observation. An analysis of variance was applied to these data to determine whether differences appeared in the average scores for the different approaches to observation. Duncan's new multiple range test was used to identify the main source of variation whenever an F-value resulting from the analysis of variance was significant.

Conclusions

There was no consistent pattern or direction in the differences observed between methods, although there was evidence to suggest the superiority of participant observation and the inferiority of film observation.

BIBLIOGRAPHY

1. Adams, Sidney. A study of the growth of language between two and four years. Journal of Juvenile Research 16:269-277. 1932.

2. Anderson, Harold H. The measurement of domination and of socially integrative behavior in teacher's contacts with children. Child Development 10:73-89. 1939.

3. Anderson, Nels. The hobo. Chicago, University of Chicago Press, 1923. 302 p.

4. Arrington, Ruth E. Interrelations in the behavior of young children. New York, Teachers College, Columbia University, 1932. 156 p. (Child Development Monograph No. 8)

5. Arrington, Ruth E. Time-sampling studies of human behavior. Psychological Bulletin 40:81-124. 1943.

6. Bales, Robert Freed. Interaction process analysis. Cambridge, Addison-Wesley, 1951. 203 p.

7. Barker, Margaret. A technique for studying the social-material activities of young children. New York, Teachers College, Columbia University, 1930. 69 p. (Child Development Monograph No. 3)

8. Barker, Roger, Jacob Kounin and Herbert Wright. Child behavior and development. New York, McGraw-Hill, 1943. 625 p.

9. Bayroff, A. G., Helen R. Haggerty and E. A. Rundquist. Validity of ratings as related to rating techniques and conditions. Personnel Psychology 7:93-112. 1954.

10. Beaver, Alma Perry. The initiation of social contacts by preschool children. New York, Teachers College, Columbia University, 1932. 65 p. (Child Development Monograph No. 7)

11. Behavior observation record. San Diego, San Diego Public Schools, 1949. 127 p. (California. State Department of Education. Bulletin No. 6)

12. Bendig, A. W. Rater reliability and heterogeneity of the scale anchors. Journal of Applied Psychology 39:37-39. 1955.

13. Bendig, A. W. Rater reliability and judgmental demoralization. Journal of Applied Psychology 41:66-68. 1957.

14. Bendig, A. W. Rater reliability and judgmental fatigue. Journal of Applied Psychology 39:451-454. 1955.

15. Bendig, A. W. Reliability and the number of rating scale categories. Journal of Applied Psychology 38:38-40. 1954.

16. Bendig, A. W. The reliability of adjustment ratings and the length of case histories. Journal of Consulting Psychology 19:463-467. 1955.

17. Bendig, A. W. and Janine Sprague. Rater experience and the reliability of case histories. Journal of Consulting Psychology 18:207-211. 1954.

18. Bott, Helen. Methods in social studies of young children. Toronto, University of Toronto, 1933. 110 p. (University of Toronto Child Development Series No. 1)

19. Budzik, Jo Ann. Reliability of observation as a function of recording procedure, number of categories, and length of time interval. Stillwater, Oklahoma A and M College, Department of Family Relations and Child Development, 1956. 53 p.

20. Burke, Laverne K. and Erwin K. Taylor. Rating training and experience. Journal of Applied Psychology 34:381-383. 1950.

21. Carter, Launor et al. A note on a new technique of interaction recording. Journal of Abnormal and Social Psychology 6:258-260. 1951.

22. Challman, Robert C. Factors influencing friendships among preschool children. Child Development 3:146-158. 1932.

23. Clarke, Howard. An experimental investigation of theorems relating to the structure and content of rating instruments. Dissertation Abstracts 17:1116-1117. 1957.

24. Conrad, Herbert Spencer. The California behavior inventory for nursery school children. Berkley, University of California Press, 1933. 74 p.

25. English, Horace B. and Ava Champney English. A comprehensive dictionary of psychological and psychoanalytical terms. New York, Longmans, Green, 1958. 59 p.

26. Fabian, Walter Albert. An investigation of the relationship between measures of insight and measures of projection and distortion in ratings. Dissertation Abstracts 14:711. 1954.

27. Falk, Gloria and A. G. Bayroff. Rater and technique contamination in criterion ratings. Journal of Applied Psychology 38:100-102. 1954.

28. Festinger, Leon and Daniel Katz (ed.) Research methods in the behavioral sciences. New York, Dryden Press, 1953. 660 p.

29. Fite, Mary Delafield. Aggressive behavior in young children and children's attitudes towards aggression. Genetic Psychology Monographs 22:151-319. 1940.

30. Gellert, Elizabeth. Systematic observation: a method in child study. Harvard Educational Review 25:179-195. 1955.

31. Gesell, Arnold. An atlas of infant behavior. New Haven, Yale University Press, 1934. 922 p.

32. Gesell, Arnold. Cinemanalysis: a method of behavior study. Journal of Genetic Psychology 47:3-16. 1935.

33. Goode, William J. and Paul K. Hatt. Methods in social research. New York, McGraw-Hill, 1952. 386 p.

34. Goodenough, Florence Laura. Interrelationships in the behavior of young children. Child Development 1:29-47. 1930.

35. Goodenough, Florence Laura. Measuring behavior traits by means of repeated short samples. Journal of Juvenile Research 12: 230-235. 1928.

36. Graves, Elizabeth A. A study of competitive and cooperative behavior by the short sample technique. Journal of Abnormal and Social Psychology 32:343-351. 1937.

37. Green, Elise Hart. Friendships and quarrels among preschool children. Child Development 4:237-252. 1933.

38. Guilford, Joy Paul. Fundamental statistics in psychology and education. New York, McGraw-Hill, 1950. 633 p.

39. Halverson, H. M. An experimental study of prehension in infants by means of cinema records. Genetic Psychology Monographs 10: 110-286. 1931.

40. Halverson, H. M. A projection table for studying motion films. American Journal of Psychology 43:124. 1931.

41. Heyns, Roger W. and Ronald Lippitt. Systematic observational techniques. In: Gardner Lindzey's Handbook of Social Psychology. Vol. 1. Cambridge, Addison-Wesley, 1954. p. 370-404.

42. Heyns, Roger W. and Alvin F. Zander. Observation of group behavior. In: Leon Festinger and Daniel Katz's Research methods in the behavioral sciences. New York, Dryden Press, 1953. p. 381-417.

43. Jack, Lois M. An experimental study of ascendant behavior in preschool children. Iowa City, University of Iowa Press, 1934. p. 7-65. (University of Iowa Study in Child Welfare No. 3)

44. Jahoda, Marie, Morton Deutsch and Stuart Cook. Research methods in social relations. New York, Dryden Press, 1951. 2 vols.

45. Jersild, Arthur Thomas. The consistency of certain behavior patterns in young children. American Journal of Psychology 45: 125-129. 1933.

46. Jersild, Arthur Thomas and Mary D. Fite. The influence of nursery school experience on children's social adjustments. New York, Teachers College. Columbia University, 1939. 112 p. (Child Development Monograph No. 25)

47. Jersild, Arthur Thomas and Frances V. Markey. Conflicts between preschool children. New York, Teachers College, Columbia University, 1935. 181 p. (Child Development Monograph No. 21)

48. Jersild, Arthur Thomas and Margaret F. Meigs. Direct observation as a research method. Review of Education Research 40:472-482. 1939.

49. Karcher, E. Kenneth and Samuel H. King. Effect of number and order of ratings on reliability and validity. American Psychologist 5:333. 1950.

50. Katz, Daniel. Field studies. In: Leon Festinger and Daniel Katz's Research methods in the behavioral sciences. New York, Dryden Press, 1953. p. 56-97.

51. Kluckhohn, Florence R. The participant-observer technique in small communities. American Journal of Sociology 46:331-343. 1940.

52. Kroeber, A. L. Anthropology today: an encyclopedic inventory. Chicago, University of Chicago Press, 1953. 966 p.

53. Landen, Delmar Leo. An experimental investigation of theorems concerning the training of raters. Dissertation Abstracts 17: 1120-1121. 1957.

54. Lewis, Oscar. Controls and experiments in field work. In: A. L. Kroeber's Anthropology today: an encyclopedic inventory. Chicago, University of Chicago Press, 1953. p. 452-475.

55. Lindzey, Gardner (ed.) Handbook of social psychology. Vol. 1. Cambridge, Addison-Wesley, 1954. 588 p.

56. Lohman, Joseph D. The participant observer in community studies. American Sociological Review 2:890-897. 1937.

57. Loomis, Alice Marie. A technique for observing the social behavior of nursery school children. New York, Teachers College, Columbia University, 1931. 100 p. (Child Development Monograph No. 5)

58. Merrill, Barbara. A measurement of mother-child interaction. In: William E. Martin and Celia B. Stendler's Readings in child development. New York, Harcourt-Brace, 1954. p. 346-357.

59. Merton, Robert K. Selected problems in field work in the planned community. American Sociological Review 13:304-312. 1954.

60. Michaelis, Anthony R. Research films in biology, anthropology, psychology, and medicine. New York, Academic Press, 1955. 490 p.

61. Miller, S. M. The participant observer and over-rapport. American Sociological Review 18:97-99. 1952.

62. Mosel, James N. Validity of rational ratings on experience and training. Personnel Psychology 5:1-10. 1952.

63. Moustakas, Clark, Irving Siegel and Henry D. Schalock. An objective method for the measurement and analysis of child-adult interaction. Child Development 27:109-134. 1956.

64. Olson, Willard Clifford. The measurement of nervous habits in normal children. Minneapolis, University of Minnesota Press, 1929. 97 p. (University of Minnesota Child Welfare Monograph No. 3)

65. Parten, Mildred Bernice. Social participation among preschool children. Journal of Abnormal and Social Psychology 27:430-440. 1933.

66. Paul, Benjamin D. Interview techniques and field relationships. In: A. L. Kroeber's Anthropology today: an encyclopedic inventory. Chicago, University of Chicago Press, 1953. p. 432-451.

67. Roberts, Katherine E. and Rachel Ball. A study of personality in young children by means of a series of rating scales. Journal of Genetic Psychology 52:79-149. 1938.

68. Rummel, J. Francis. An introduction to research procedures in education. New York, Harper and Brothers, 1958. 413 p.

69. Russell, David Lawrence. A comparison of rating, test, and sociometric methods of personality measurement. Dissertation Abstracts 14:552-553. 1954.

70. Schuler, Edgar A. A study of the consistency of dominant and submissive behavior in adolescent behavior. Journal of Genetic Psychology 46:403-432. 1935.

71. Schwartz, Morris S. and Charlotte Green Schwartz. Problems in participant observation. American Journal of Sociology 60: 343-353. 1955.

72. Strodtbeck, Fred L. and A. Paul Hare. Bibliography of small group research, 1900-1953. Sociometry 17:107-178. 1954.

73. The making of an infantry man. American Journal of Sociology 51:376-379. 1946.

74. Thomas, Dorothy Swaine. An attempt to develop precise measurements in the social behavior field. Sociologus 8:436-456. 1932.

75. Thomas, Dorothy Swaine et al. Some new techniques for studying social behavior. New York, Teachers College, Columbia University, 1929. 203 p. (Child Development Monograph No. 1)

76. Thomas, Dorothy Swaine, Alice Marie Loomis and Ruth E. Arrington. Observational studies of social behavior. New Haven, Institute of Human Relations, Yale University Press, 1933. 437 p.

77. Tolle, Edwin Randall. A critical analysis of the forced-choice rating technique when used in rating elementary classroom teachers of a metropolitan school system. Dissertation Abstracts 17:1021-1022. 1957.

78. Vidich, Arthur J. Participant observation and the collection and interpretation of data. American Journal of Sociology 60: 354-360. 1955.

79. Warner, William Lloyd and Paul S. Lunt. The social life of a modern community. New Haven, Yale University Press, 1941. 460 p.

80. Webb, Wilse B. Self-evaluation compared with group evaluation. Journal of Consulting Psychology 16:305-307. 1952.

A P P E N D I X E S

APPENDIX "A"

BEHAVIOR RATING SCALE

I. TENSION

There are two aspects of this variable to be rated, the extent to which a child evidences tension during the course of the observation and the extent to which the situation is inferred to be tension-producing.

1. Extent to which a child evidences tension during the course of the observation.

As used here tension is defined in terms of behaviors which suggest the presence of a tenseness, fearfulness, or a surplus of energy which needs to find an immediate release. Evidence for the rating of tension comes from such behaviors as typical anxiety symptoms, non-provoked attack, responses that are obviously out of proportion to their stimulation, rushing about aimlessly with much shouting and "letting off of steam", etc.

Evaluation of the level of tension is based on behavioral cues which suggest that intensity of the tension the child is experiencing. The degree to which a child evidences tension in the situation has been defined into four levels. These are:

1. Extreme tension--Evidence for this rating of this level of tension comes from such behavioral cues as rigidity of muscles; nervous mannerisms such as profuse thumb sucking, nail biting, tics; harsh, prolonged sobbing, screaming, yelling, cursing, or talking rapidly in shrill tones; forceful hitting, throwing, kicking with intent to injury or harm; intense clinging to the mother, persistent striving to sit on her lap, hiding behind the mother's skirt, or drawing self up into tightly curled position.

2. Considerable tension--Evidence for the rating of this level of tension comes from behaviors differing from those used in rating extreme tension only in their level of intensity. This decrease in the level of intensity leads to a judgment of considerable tension rather than judgment of extreme tension.

3. Some tension--Evidence for the rating of this level of tension comes from behaviors differing from those used in rating considerable tension only in their level of intensity.

This decrease in the level of intensity leads to a judgment of some tension rather than a judgment of considerable tension.

4. Little or no tension--Evidence for the rating of this level of tension comes from such behavioral cues as an apparent relaxed face and body; calm tone of voice, i.e. one that is not loud, shrill, harsh, or whining; a free, easy flow of words; interaction with people and materials which shows no aggressive, destructive or attacking behavior; little seeking of nurturance, and freedom from other typical anxiety symptoms.

In judging the level of tension the rater is to consider only the intensity of the cues suggesting tension. The amount of time a child spends in evidencing a particular level of tension and number of channels a child uses to express his tension are not to be considered in judging tension level. While this is the case, it should be noted the number of channels a child uses in expressing his tension does play a part in arriving at a judgment of tension level in that the greater the number of cues observed the more adequate the basis for such judgment.

It has been observed that the level of tension a child is experiencing is closely associated with the behavior patterns he exhibits. For this reason it is necessary to assess a child's behavior at each of the levels of tension he experiences. For example, if a child experiences "little or no tension" for a part of an observation period but also experiences "some tension" for another part of the period, it is necessary to evaluate the child's behavior as it appears in relation to each of these levels of tension. The major difficulty in such a procedure is that the rater has to observe a child experiencing a given level of tension for a considerable period of time to get enough cues to make a behavioral assessment at a particular level of tension meaningful. In order to decide when the observer likely has enough cues to make a rating possible, a two-way classification of the tension variable involving the level of tension and the amount of time a child exhibits a particular level of tension is made. If a child experiences a given level of tension for a "considerable proportion of time" or longer (see scale values below), the child's behavior will be rated as it is evidenced at that particular level of tension. This means then that if a child experiences a particular level of tension for a "large proportion of time", another level of tension for "a considerable proportion of time", the child's behavior will be rated separately for each of these levels of tension. However, if a child evidences a particular level of tension for a "large proportion of time" but evidences another level of tension for a "small proportion of time" or even "some of the time", the child's behavior will be rated only for the level at which the child spent a "large proportion of time".

The large range in the percentage of time under each of the scale points makes it possible to have several different combinations of scale values for any one period of observation. For example, since the scale point "a large proportion of time" has a percentage range of 40 to 99 percent, it is possible to have a child spend a large proportion of time at two levels of tension. In this way the number of possible combinations of scale values is limited only in that the total minimum percentages of time cannot exceed 100 percent and the total maximum percentages of time can equal 100 percent.

In making this two-fold rating each of the levels of tension is to be marked irrespective of whether all levels of tension are evidenced.

In summary, the factors one needs to consider in making this rating are:

1. Both the level of tension and the time spent at a particular level of tension needs to be rated.

2. In judging the level of tension the rater is to consider only the intensity of the cues suggesting tension. The number of cues suggesting tension and the amount of time the child spends in evidencing this tension should not be considered in rating the level of tension.

3. In judging the length of time spent at a particular level of tension, the rater will simply estimate the period of time rather than keeping track of the time systematically.

4. In making the two-fold rating each of the levels of tension is to be marked irrespective of whether all levels of tension are evidenced.

	All the time (99-100%)	A large proportion of the time (40-99%)	A considerable proportion of the time (26-39%)	Some proportion of the time (11-25%)	A small proportion of the time (1-10%)	None of the time
Extreme tension						
Considerable tension						
Some tension						
Little or no tension						

2. Extent to which the situation is inferred to be tension-producing.

As used here a situation is tension-producing when it contains elements or events which may be expected to be stressful.

Evidence for the rating of this variable comes from cues which suggest the extent to which the elements or events in a situation are capable of producing tension, i.e., whether these events are potentially tension-producing, tension-producing, or extremely tension-producing. In rating this variable the manifestations of tension in the child's behavior are not to be considered. It is the potential of the situation to produce tension that is to be judged.

Evaluation of the degree to which the situation may be tension-producing is based upon the severity of the events which occur in the situation. The degree to which the situation is tension-producing has been defined into four levels. These are:

1. A situation which is extremely tension-producing is one in which there are stimuli which would ordinarily be extremely tension-producing to a child. Examples of situations of this nature are: a child working a puzzle and two other children taking some of the pieces of puzzle out and throwing them at the child; an infant playing pat-a-cake with his mother and falling off her lap; a child falling off a tricycle and receiving a severe blow on the head.

2. A situation which is tension-producing is one in which there are stimuli which would ordinarily be tension-producing but not extremely tension-producing. Examples of situations of this nature are: a child trying to work a puzzle which is above his developmental level; an infant playing pat-a-cake with his mother while a sibling tries to push the child off the mother's lap; a child riding a tricycle around the playyard and another child trying to take the tricycle away from him.

3. A situation which is somewhat tension-producing is one in which potentially disturbing or stressful stimuli occur. Potentially disturbing stimuli are defined as events which could be stressful for some children, but which would not be expected to be inevitably stressful for all children. Examples of situations of this nature are: a child working a puzzle by himself and another child trying to interfere by offering help; a child riding a tricycle around the playyard and being bumped from behind by a child on another tricycle; an infant playing pat-a-cake with his mother when an unexpected loud noise occurs.

4. A situation in which the stimuli produced little or no tension is one in which there are no apparent, unexpected or unusual events which could be thought of as stressful. Examples of situations of this nature are: a child working a puzzle which is within his developmental level; an infant playing pat-a-cake with his mother; a child riding a tricycle around the playyard.

In order to determine the relationship between the amount of time spent by a child evidencing a given level of tension and the amount of time one would predict the child to evidence a given level of tension, with the prediction being based upon the amount of time the situation is tension-producing, it is necessary to judge the amount of time which stimuli of varying levels of intensity operate during an observation period. In order to do this we have employed a two-fold classification involving the extent to which a situation is tension-producing and the amount of time which stimuli of varying levels of intensity operate. Judgment of the percentage of time a situation is tension-producing to some degree is determined strictly on the basis of time. The number of tension-producing incidents does not play a role in making the rating except in so far as it contributes to the total amount of time which is tension-producing to some degree.

The procedure the rater uses in assigning scale values for the amount of time the stimuli are tension-producing at a given level may be found on pages 67 and 68.

In making this two-fold rating, each of the levels of tension-producingness is to be marked irrespective of whether all levels are evidenced.

In summary, the factors one needs to consider in making this rating are:

1. Both the extent to which the situation is tension-producing and the amount of time the stimuli are tension-producing at a particular level need to be rated.

2. In judging the extent to which a situation is tension-producing, the rater is to consider only the intensity of the unexpected or unusual events which occur and not the frequency with which they occur nor the length of time which they continue.

3. The percentage of time which is tension-producing to some degree is determined strictly on the basis of time, i.e. the number of tension-producing incidents does not play a role in making the ratings except in so far as they

contribute to the amount of time the situation is tension-producing, to some degree.

4. A rating needs to be made for all levels of tension-producingness irrespective of whether there are events in the observation which were representative of all levels of tension-producingness.

Stimuli are extremely tension-producing						
Stimuli are tension-producing						
Stimuli are somewhat tension-producing						
Stimuli are slightly if at all tension-producing						
	All of the time (99-100%)	A large proportion of the time (40-99%)	A considerable proportion of the time (26-39%)	Some of the time (11-25%)	A small proportion of the time (1-10%)	None of the time

****

A SPECIAL PROCEDURE FOR SCORING is needed when a child spends a <u>considerable proportion or more</u> of the observation time at more than one level of tension. When this occurs, it becomes necessary to assess the child's behavior as it is evidenced at each of these levels of tension involved.

To facilitate this multiple recording, colored pencils are used. A plain lead pencil will be used when there are stimuli which are only slightly if at all tension-producing; a red lead pencil will be used when the stimuli are somewhat tension-producing; a blue lead pencil will be used when stimuli are tension-producing; and a green lead pencil will be used when the stimuli are extremely tension-producing. In this way it will be possible to make as many ratings as are needed on a single rating scale.

It will be recalled that if a child spends only some or a smaller proportion of the time at a given level of tension during the period of observation, a separate rating of the child's behavior as it appeared at these levels of tension is not required. This is due to the relatively short period of time the child is experiencing these particular levels of tension with the consequence being that the observer is not able to get enough cues to make a rating possible.

****

II. AVENUES USED IN THE EXPRESSION OF TENSION

This variable is defined in terms of the nature or locus of the behaviors which serve the child in the expression of tension. Seven avenues have been defined: bodily expression; vocal expression; physical aggression expressed in positive, acceptable (non-destructive) ways toward materials which are appropriate for this expression; physical aggression expressed through attack on materials; aggression through attack on people and/or their materials; expression of tension through the seeking of nurturance; and expression of tension through showing-off behavior.

In rating each of these avenues a judgment as to the extent to which the child used the avenue is needed. This has been defined into five levels: much use of the avenue; considerable use of the avenue; some use of the avenue; slight use of the avenue; no use of the avenue. When making the judgment as to the extent of use of a particular avenue the judgment is to be based upon the percentage of time a child expresses his tension through a particular avenue relative to the total amount of time this child is evidencing tension rather than the total amount of time the child is being observed. This means for example, if a child evidences tension for a total of five minutes and in this time he expresses his tension for the full five minute period through the bodily expression avenue, he expresses tension vocally for three minutes, and through physical attack on materials one minute, he would receive a rating of much use of the Bodily Avenue and the Vocal Avenue, and some use of the Physical Attack on Materials Avenue.

If a child exhibits little or no tension throughout the course of the observation, then ratings under this heading are not needed. If any tension is evidenced in the course of the observation, irrespective of the proportion of time it lasts, then all of the scales under this heading need to be scored.

3. Bodily expression of tension.

Evidence for the rating of this variable comes from such behaviors as tenseness of body and facial muscles; muscular rigidity; trembling; nervous mannerisms such as thumb sucking, picking the nose, etc.

Much use of this avenue (40-100%)	Considerable use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

4. Vocal expression of tension.

Evidence for the rating of this variable comes from such behaviors as shouting, yelling, cursing, crying, loudness and rapidity of talking, etc. Any vocal expression evidencing tension, other than a vocal attack on a person, will be considered as evidence for this rating.

Much use of this avenue (40-100%)	Considerable use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

5. Physical aggression expressed in positive, acceptable (non-destructive) ways toward materials which are appropriate for this expression.

Evidence for the rating of this variable comes from such behaviors as pounding, hitting, or any other forceful, releasing behavior that is of a non-destructive nature. Some of the materials which may be used in this non-destructive, yet forceful way, are blocks, pounding clay, pounding boards, etc.

Much use of this avenue (40-100%)	Considerable use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

6. Physical aggression expressed through attack on materials.

Evidence for the rating of this variable comes from such behaviors as hitting, throwing, kicking, or any other behavior which has as its goal the destruction or mutilation of an object.

Much use of this avenue (40-100%)	Considerable use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

7. Aggression through attack on people and/or their materials.

Evidence for the rating of this variable comes from such behaviors as kicking, spitting, throwing of objects toward a person, cursing, or any other behavior, either verbal or physical, which has as its goal the hurting of people.

Much use of this avenue (40-100%)	Considerable Use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

8. Expression of tension through the seeking of nurturance.

Evidence for the rating of this variable comes from behaviors which suggest the seeking of care, comfort, protection, or support. Specific examples of such behaviors are crawling into the mother's lap, hanging onto the mother's skirt, holding the mother's hand, standing close to an adult, etc.

Much use of this avenue (40-100%)	Considerable use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

9. Expression of tension through showing-off behavior.

Much use of this avenue (40-100%)	Considerable use of this avenue (26-39%)	Some use of this avenue (11-25%)	Slight use of this avenue (1-10%)	No use of this avenue

III. BODY MOVEMENT

This variable is defined in terms of the quickness with which the child moves and the way in which he uses his body generally.

10. Speed of movement.

Evidence for the rating of this variable comes from the general quickness with which a child moves. Hand movements, body movements, and general gait are considered here.

Extremely quick, rapid movements	Quick movements	Neither particularly quick nor particularly slow movements	Slow movements	Extremely slow, lethargic movements

11. Expansiveness of movement.

Evidence for the rating of this variable comes from the way in which a child uses his body. Large, wide, free unrestrained body movements will be taken as evidence of expansiveness while smaller, more restricted, more careful, more restrained movements will be taken as evidence of restraint.

Extremely expansive movements	Expansive movements	Neither particularly expansive nor particularly restrained movements	Restrained movements	Extremely slow, restrained movements

****

IV. OUTGOINGNESS

This variable is defined in terms of a child's interaction with materials and/or people. Interaction as used in this scale may be behavioral interchange between two or more people or it may be the active directing of behavior toward one or more individuals without reciprocity, or it may be the active directing of behavior toward materials. The interaction may be verbal, motor, or visual but the interaction we are focusing upon is the interaction involved in or around the activity, people, or material toward which attention seems to be centrally focused. Attention may be pulled away from this activity periodically but as long as it is returned, the activity retains its central focus. Interaction occurring tangentially to the central focus of the activity, i.e. behavior which involves looking away from the task, talking to someone outside the group, or interacting with someone outside the group motor-wise will be considered as incidental interaction.

Outgoingness is broken down into four components, frequency with which a child initiates interaction, proportion of overtures to interaction accepted by a child, the amount of time spent in interaction, and the level of involvement in the interaction. These four components are rated with reference to interaction with materials and interaction with people with the exception of the child's response to overtures to interaction which is rated only for interaction with people.

A. Frequency with which a Child Initiates Interaction.

As used here the initiation of interaction is defined in terms of undertaking or attempting to undertake an activity or conversation with materials or people. In this sense the initiation of interaction requires more than simple watching. That is, it requires an active, clear-cut verbal or physical overture to interaction. Incidental interaction, as defined above, will not be considered as an attempt to initiate interaction.

12. Frequency with which a child initiates interaction with materials.

Evidence for the rating of this variable comes from such behavior as a child talking with a doll, working a puzzle, or building with blocks.

Very frequently (7 or more initiations)	Frequently (4 to 6 initiations)	Sometimes (2 to 3 initiations)	Rarely (1 initiation)	Never (no initiations)

13. Frequency with which a child initiates interaction with adults.

Evidence for the rating of this variable comes from such behaviors as a child inviting another adult to play with him in the doll corner, a child with a rope handing one end to another adult, a child asking an adult to read a story to him or to play the piano for him.

Very frequently (7 or more initiations)	Frequently (4 to 6 initiations)	Sometimes (2 to 3 initiations)	Rarely (1 initiation)	Never (no initiations)

14. Frequency with which a child initiates interaction with children.

Very frequently (7 or more initiations)	Frequently (4 to 6 initiations)	Sometimes (2 to 3 initiations)	Rarely (1 initiation)	Never (no initiations)

B. Proportion of overtures to interaction accepted by a child.

As used here an overture to interaction is synonymous with the initiating of interaction as defined in "A" above. An acceptance of such an overture is defined as a response to the overture which serves to stimulate further interaction. Non-acceptance of an overture is defined as a response which tends to discourage or block interaction.

15. Proportion of adult overtures to interaction accepted by a child.

Evidence for the rating of this variable comes from such behaviors as a child accepting an adult invitation to join in a group of children sliding on the slide, a child responds to an adult's initiation of conversation by further conversation, an infant rolls a ball back to an adult who has initiated this play.

A child responds to a large proportion of adult overtures (76-100%)	A child responds to a considerable proportion of adult overtures (51-75%)	A child responds to some adult overtures (26-50%)	A child responds to a small proportion of adult overtures (1-25%)	A child responds to no adult overtures

No adult overtures_____

16. Proportion of child overtures to interaction accepted by a child.

Evidence for the rating of this variable comes from such behaviors as a child beginning to play in the doll corner after being invited, a child swinging after being asked, a child entering into a singing group around the piano after the adult asks him.

A child responds to a large proportion of child overtures (76-100%)	A child responds to a considerable proportion of child overtures (51-75%)	A child responds to some child overtures (26-50%)	A child responds to a small proportion of child overtures (1-25%)	A child responds to no child overtures

No child overtures_____

C. Amount of time spent in interaction and level of involvement in interaction.

Amount of time spent in interaction is defined in terms of the proportion of time a child spends in verbal and motor interchange with materials and/or people. In making this rating, incidental interaction i.e. the directing of attention to something other than that which occupies the central focus of the action is not to be considered. Evidence for rating the amount of time spent in interaction comes only from the time spent in focused interaction with materials and/or people.

The level of involvement in interaction is defined in terms of the amount of incidental interaction evidenced during the course of observation. If the child's attention is never taken from the activity or task at hand, we will consider his level of involvement to be deep. If the child directs his attention very frequently to factors other than that which commands a central focus, we will consider his level of involvement to be slight.

Frequency of incidental interaction is transposed to a percentage estimate of the total interaction time. This percentage estimate will then enable us to judge a child's level of involvement relative to the total amount of time spent in interaction. The necessity for this procedure lies in the fact that a simple frequency count of the occurrence of incidental interaction has no meaning unless the time spent in interaction is known. For example, three occurrences of incidental interaction during a total of five minutes of interaction suggests a considerably different level of involvement than three occurrences of incidental interaction in fifty minutes of interaction.

In making this two-fold rating, only one of the statements describing the amount of time spent in interaction will be checked. Consequently, there will be only one rating of the level of involvement.

In summary, the factors one needs to consider in making the rating are:

1. Both the amount of time spent in interaction and the level of involvement a child evidenced in this interaction are to be rated.

2. In judging the amount of time the child spends in interaction, the rater is to consider the total time spent in interaction with people or materials independent of the level of involvement of the child in his interaction.

3. In judging the level of involvement of the child in interaction, the rater is to consider the frequency with which the child engages in incidental interaction.

4. In judging the level of involvement, the rater needs to transpose the frequency of incidental interaction to a percentage estimate of total interaction time.

5. In making the two-fold rating, only one of the statements describing the time spent in interaction is to be marked.

****

17. Amount of time spent in interaction with materials.

All of the time _____

A large proportion of the time (76-99%) _____

A considerable proportion of the time (50-75%) _____

Some of the time (26-49%) _____

A small proportion of the time (1-25%) _____

Deep involvement--A child evidences incidental interaction a <u>small</u> proportion of the time (1-10%)	Considerable involvement--A child evidences indidental interaction <u>some</u> of the time (11-25%)	Some involvement--A child evidences incidental interaction a <u>considerable</u> proportion of the time (26-39%)	Slight involvement--A child evidences incidental interaction a <u>large</u> proportion of the time (40-99%

None of the time _____

18. Amount of time spent in interaction with adults.

All of the time _____

A large proportion of time (76-99%) _____

A considerable proportion of the time (50-75%) _____

Some of the time (26-49%) _____

A small proportion of the time (1-25%) _____

Deep involvement--A child evidences incidental interaction a <u>small</u> proportion of the time (1-10%)	Considerable involvement--A child evidences incidental interaction <u>some</u> of the time (11-25%)	Some involvement--A child evidences incidental interaction a <u>considerable</u> proportion of the time (26-39%)	Slight involvement--A child evidences incidental interaction a <u>large</u> proportion of the time (40-99%)

None of the time _____

19. Amount of time spent in interaction with children.

All of the time _____

A large proportion of the time (76-99%) _____

A considerable proportion of the time (50-75%) _____

Some of the time (26-49%) _____

A small proportion of the time (1-25%) _____

Deep involvement--A child evidences incidental interaction a <u>small</u> proportion of the time (1-10%)	Considerable involvement--A child evidences incidental interaction <u>some</u> of the time (11-25%)	Some involvement--A child evidences incidental interaction a <u>considerable</u> proportion of the time (26-39%	Slight involvement--A child evidences incidental interaction a <u>large</u> proportion of the time (40-99%)

None of the time _____

V. PURPOSEFULNESS

As used here this variable is defined in terms of the persistence with which a child pursues his goals.

In the pursuit of his goals a child frequently encounters obstacles to the realization of these goals. When an obstacle is encountered, the behavioral response to the frustration may have about it elements of intensity, adaptive-non-adaptive qualities, and a persistence factor in striving to overcome or circumvent the barrier, etc. In rating the purposefulness of a child's behavior only the quality of persistence will be considered.

Persistence is defined in terms of the tenacity with which a child pursues his goals. Persistence has been defined into four levels with the basis of differentiation being the length of time which the child characteristically spends in response to a frustration.

1. Extreme persistence--The child characteristically spends five or more minutes in the pursuit of his frustration goals.

2. Persistence--the child characteristically spends three to four minutes in pursuit of his frustrated goals.

3. Some persistence--the child characteristically spends one to two minutes in pursuit of his frustrated goals.

4. Little or no persistence--the child characteristically spends less than one minute in pursuit of his frustrated goals.

Evidence for the rating of persistence comes from the response of a child in the face of frustration or in the blocking of goal oriented behaviors. From this statement it is evident that the basis for making the rating will come from the observation of behavior in response to the existence of or the imposition of frustrating events or circumstances.

In order that the rating of a child's persistence be meaningful, the rating must be accompanied by a knowledge of the strength of the frustrations he was encountering. The classification of a situation as to the extent to which it is frustrating has as its basis the extent to which the situation blocks goal-oriented behavior. Blocking may involve the existence of a barrier in any form, e.g. a verbal restriction, a physical barrier, or a task requiring a particular level of skill for its completion. The blocking may occur in any degree of completion.

The extent to which a situation is frustrating has been defined into four levels:

1. An extremely frustrating situation is one in which there are circumstances or events which completely block goal-oriented behavior or which make the continuance of goal-oriented behavior extremely difficult or risky. Examples of events or circumstances which would be considered as extremely frustrating are such physical barriers as the erection of a gate blocking off the stairs to an infant, closing and locking a door that the child cannot open; such verbal barriers as direct statements of restriction or forbidding; or barriers resulting from the level of task-difficulty which make it nearly impossible to complete the task within the child's developmental level.

2. A frustrating situation is one in which there are circumstances or events which considerably block goal-oriented behavior or which make the continuance of goal-oriented behavior considerably difficult or risky. These events or circumstances, however, can be overcome if a child exerts much effort. Examples of such situations are an infant attempting to climb up a high flight of stairs; a child asking to play with another group of children but is told, "Go away! You can't play with us!", or a child pulling a heavy wagon of blocks up an incline.

3. A somewhat frustrating situation is one in which circumstances or events block goal-oriented behavior to some degree or make the continuance of goal-oriented behavior somewhat difficult or risky. These events or circumstances can be overcome if the child exerts some exercise or effort. Examples of such situations are a child who falls off a tricycle and gets his foot caught in it; a child who is playing policeman says to another child, "You can't get by me!", and a child who tries to put on his own boots.

4. A situation in which there is little or no frustration is one in which there are few if any circumstances or events which block goal-oriented behavior or which make the continuance of goal-oriented behavior difficult. There are no barriers which the child cannot easily surmount, few if any adult or child restrictions or demands upon the child, and all tasks are within the child's developmental level.

The relationship between extent to which the situation is tension-producing and the extent to which the situation is frustrating must be pointed out. Both are closely related; however, the extent to which the situation is frustrating differs from the tension-producing qualities of a situation in that the extent to which the situation is tension-producing is a more global rating, i.e. frustration is only

one of the factors contributing to the extent to which the situation is tension-producing. For this reason, the extent to which a situation is frustrating may or may not be identical to the rating of the extent to which a situation is tension-producing.

At the end of each observation ratings will be made of the child's persistence in pursuing his frustrated goals at the various levels of frustration which occurred throughout the course of the observation. If a child were involved in nothing but situations which involved little or no frustration for the entire course of the observation, a rating of a child's persistence would be made only at the little or no frustration level. If the child were involved in situations with little or no frustration during the observation period with the exception of one somewhat frustrating situation and one extremely frustrating situation, then rating of the child's persistence at each of those separate levels would be made. If this latter description were altered to include three somewhat frustrating situations instead of just one, there would still be only three ratings, one rating each for the situations with little or no frustration, the situations which were somewhat frustrating, and the extremely frustrating situation. The difference would be that a child's persistence in the somewhat frustrating situations would be based upon a synthesis or an average of the way in which persistence was evidenced in all three somewhat frustrating situations. A major principle of scoring then is that the rating of level of frustration will represent a synthesis or an average of the child's persistence in pursuit of frustrated goals at each level of frustration during the course of the observation.

To facilitate the rating of the two variables, the scales have been combined into a two-fold table. In making this two-fold rating only the levels of frustration which occurred are scored. However, for every level of frustration that was evidenced, the persistence with which the child attempted to circumvent or overcome the barrier has to be rated.

In making this rating, it will be recalled that only the average or the characteristic amount of time spent in trying to circumvent or overcome the barriers at each of the levels of frustration will be considered.

In summary, the factors one needs to consider in making this rating are:

1. Both the level of frustration and the time spent in trying to overcome or circumvent the frustration at each of these level(s) need to be rated.

2. In judging the level of frustration, the rater is to consider the extent to which the situation blocks goal-oriented behavior.

3. In making the two-fold rating, only the levels of frustration which are evidenced are to be rated.

4. In making this rating, only the average or characteristic amount of time spent in trying to circumvent or overcome the barriers at each of the levels of frustration will be considered.

****

20. Purposefulness.

Extreme frustration				
Frustration				
Some frustration				
	Extreme persistence in trying to overcome barrier (5 minutes or more)	Persistence in trying to overcome barrier (3 to 4 minutes)	Some persistence in trying to overcome barrier (1 to 2 minutes)	Little or no persistence in trying to overcome barrier (less than 1 minute)
Little or no frustration				

APPENDIX "B"

OREGON STATE COLLEGE
School of Home Economics
Corvallis, Oregon

We are planning a long-range investigation of behavior patterns shown by preschool children in unfamiliar situations, and we hope to begin by observing a selected number of infants and younger preschool children in play with others.

We plan to have 16 children in our preliminary study, ranging in age from 10 months to 34 months. The child and his mother will be asked to come to the Orchard Street nursery school for an hour's free play with three other children of the same age.

We are writing to parents who have children registered in the application file of the Oregon State College nursery schools since we wish to observe the children again later. We would like to invite you to participate with us in this study, and hope that you may find it possible.

Two of our graduate assistants, Margaret Shea and Pat Walker, who are also assistants in the college nursery schools are working together on this study. One of them will telephone you in the next few days and will be ready to answer questions which you may have.

If you are interested, we hope that you and your child may be able to participate in the observation play period.

Sincerely yours,

(Mrs.) Katherine H. Read
Head, Department of Family Life

KHR:pw