JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1968, 1, 175-191 NUMBER 2 (SUMMER 1968)

A METHOD TO INTEGRATE DESCRIPTIVE AND EXPERIMENTAL FIELD STUDIES AT THE LEVEL OF DATA AND EMPIRICAL CONCEPTS[1]

SIDNEY W. BIJOU, ROBERT F. PETERSON, AND MARION H. AULT

UNIVERSITY OF ILLINOIS

It is the thesis of this paper that data from descriptive and experimental field studies can be interrelated at the level of data and empirical concepts if both sets are derived from frequency-of-occurrence measures. The methodology proposed for a descriptive field study is predicated on three assumptions: (1) The primary data of psychology are the observable interactions of a biological organism and environmental events, past and present. (2) Theoretical concepts and laws are derived from empirical concepts and laws, which in turn are derived from the raw data. (3) Descriptive field studies describe interactions between behavioral and environmental events; experimental field studies provide information on their functional relationships. The ingredients of a descriptive field investigation using frequency measures consist of: (1) specifying in objective terms the situation in which the study is conducted, (2) defining and recording behavioral and environmental events in observable terms, and (3) measuring observer reliability. Field descriptive studies following the procedures suggested here would reveal interesting new relationships in the usual ecological settings and would also provide provocative cues for experimental studies. On the other hand, field-experimental studies using frequency measures would probably yield findings that would suggest the need for describing new interactions in specific natural situations.

Psychology, like the other natural sciences, depends for its advancement upon both descriptive accounts and functional analyses of its primary data. Descriptive studies answer the question "How?". They may, for example, report the manner in which a Bantu mother nurses her child, or the way in which the Yellow Shafted Flicker mates. Experimental studies, on the other hand, provide the "Why?". They might discuss the conditions which establish and maintain the relationships between the mother and infant, between the male and female birds.

It has been claimed that progress in the behavioral sciences would be enhanced by more emphasis on descriptive studies. This may be true, but one may wish to speculate on why descriptive accounts of behavior have been deemphasized. One possibility is the difficulty of relating descriptive and experimental data. For example, a descriptive study of parent-child behavior in the home may have data in the form of ratings on a series of scales (Baldwin, Kalhorn, and Breese, 1949), while an experimental study on the same subject may have data in the form of frequencies of events (Hawkins, Peterson, Schweid, and Bijou, 1966). Findings from the first study cannot reasonably be integrated with the second at the level of data and empirical concepts. Anyone interested in relating the two must resort to imprecise theory or concepts like "permissive mother", "laissez-faire atmosphere", "controlling child", "negativism", *etc.* This practice is unacceptable to psychologists who believe that all concepts must be based on or linked to empirical events.

It is the thesis of this paper that descriptive field studies (which include cross-cultural, ecological, and normative investigations) and experimental field studies can be performed so that the data and empirical terms in each are continuous, interchangeable, and mutually interrelatable.

Barker and Wright (1955) state that one of the aims of their ecological investigations is to produce data that may be used by all investigators in child behavior and development.

[1]The formulation presented here was generated from the research conducted under grants from the U. S. Public Health Service, National Institute of Mental Health (M-2208, M-2232, and MH-12067), and from the U. S. Office of Education, Handicapped Children and Youth Branch (Grant No. 32-23-1020-6002, Proposal No. R-006). Reprints may be obtained from Sidney W. Bijou, Child Behavior Laboratory, University of Illinois, 403 East Healey, Champaign, Illinois 61820.

Their study of "Midwest" and its children (1955) is in part devoted to the development of a method which provides raw material (which they compared to objects stored in a museum) amenable to analyses from different theoretical points of view. There are two considerations which make this doubtful. First, their data consist of "running accounts of what a person is doing and his situation on the level of direct perception or immediate inference" with "minor interpretations in the form of statements *about* rather than descriptions *of* behavior or situations" (Wright, 1967). It would seem that the material they collect would be serviceable only to those who accept non-observables in the raw data defined according to their prescription. Investigators who prefer to define their hypothetical variables some other way or who wish to exclude non-observables will find it difficult to integrate their data with those in the Barker and Wright studies. Second, final data in the form of running narrations cannot readily be transformed into units describing interactions between behavioral and environmental events, such as duration, intensity, latency, or frequency. Any attempt to convert such verbal accounts into one or more of the interactional dimensions would require so many arbitrary decisions that it would be doubtful whether another investigator could even come close to producing the same operations and results.

If, however, frequency-of-occurrence measures of environmental and behavioral events were used in both descriptive and field experimental studies, data and empirical concepts could be made congruous. The measure of frequency is preferable to that of duration, intensity, and latency for several reasons (Skinner, 1953). First, this measure readily shows changes over short and long periods of observations. Second, it specifies the *amount* of behavior displayed (Honig, 1966). Finally, and perhaps most important, it is applicable to operant behaviors across species. Hence, a methodology based on frequency of events would be serviceable for both experimental and descriptive studies of both human and infra-human subjects. This versatility has been illustrated by Jensen and Bobbitt in a study on mother and infant relationships of the pigtailed macaques (Jensen and Bobbitt, 1967).

With the use of frequency measures, the work of the ecological psychologist and the experimental psychologist would both complement and supplement each other. Descriptive studies would reveal interesting relationships among the raw data that could provide provocative cues for experimental investigations. On the other hand, field experimental studies would probably yield worthwhile leads for descriptive investigations by pointing to the need for observing new combinations of behavioral classes in specified situations. Ecological psychologists would show in terms of frequency of events, the practices of a culture, subculture, or an institutional activity of a subculture; experimental investigators working with the same set of data terms and empirical concepts would attempt to demonstrate the conditions and processes which establish and maintain the interrelationships observed.

Before considering the procedures for conducting a descriptive study using frequency measures, it might be well to make explicit three basic assumptions. The first: for psychology as a natural science, the primary data are the observable interactions between a biological organism and environmental events, past and present. These interrelationships constitute the material to be recorded. This means that the method does not include accounts of behavior isolated from related stimulus events ("Jimmy is a rejected child." "Johnny is a highly *autistic* child." "First Henry moved about by making swimming movements, later he crawled, now he can walk with support.") Furthermore, it means that it excludes statements of generalizations about behavior and environmental interactions. ("This is an extremely aggressive child who is always getting into trouble.") Finally, it means that it excludes accounts of interactions between behavioral and environmental events intertwined with hypothetical constructs. ("The preschool child makes errors in describing the water line in a jar because of his undeveloped cognitive structure.")

The second assumption: concepts and laws in psychology are derived from raw data. Theoretical concepts evolve from empirical concepts and empirical concepts from raw data; theoretical interactional laws are derived from empirical laws and empirical laws from relationships in the raw data.

The third assumption: descriptive studies provide information only on events and their occurrence. They do not provide information

on the functional properties of the events or the functional relationships among the events. Experimental studies provide that kind of information.

We move on to consider the procedures involved in conducting a descriptive field investigation. They include: (1) specifications of the situation in which a study is conducted, (2) definitions of behavioral and environmental events in observable terms, (3) measurements of observer reliability, and (4) procedures for collecting, analyzing, and interpreting the data. We terminate the paper with a brief illustration of a study for the behavior of a 4-yr-old boy in a laboratory nursery school.

Specifying the Situation in which a Study is Conducted

We define the situation in which a study is conducted in terms of its physical and social setting and the *observable events* that occur within its bounds. The physical setting may be a part of the child's home, a hospital or residential institution, a store, or a playground in the city park. It may be a nursery school, a classroom in an elementary school, or a room in a child guidance clinic.

The specific part of the home selected as a setting may consist of the living room and kitchen if the design of the home precludes flexible observation (Hawkins, Peterson, Schweid, and Bijou, 1966). In a hospital it might be the child's bedroom, the dining room, or the day room (Wolf, Risley, and Mees, 1964). In a state school for the retarded, it may be a special academic classroom (Birnbrauer, Wolf, Kidder, and Tague, 1965); in a regular elementary school, a classroom (Becker, Madsen, Arnold, and Thomas, 1967); and in a nursery school, the schoolroom and the play yard (Harris, Wolf, and Baer, 1964).

During the course of a study, changes in the physical aspect of the situation may occur despite efforts to keep them constant. Some will be sufficiently drastic to prevent further study until restoration of the original conditions (*e.g.*, power failure for several days). Others will be within normal limits (*e.g.*, replacement of old chairs in the child's bedroom) and hence will not warrant disrupting the research.

The social aspect of the situation in a home might consist of the mother and the subject's younger sibling (Hawkins, *et al.*, 1966); in a child guidance clinic, the therapist and the other children in the therapy group. In a nursery school it might include the head teacher, the assistant teacher, and the children (Johnston, Kelley, Harris, and Wolf, 1966).

Sometimes the social situation changes according to routines and the investigator wishes to take records in the different situations created by the changes. For example, he may wish to describe the behavior of a preschool child as he engages in each of four activities in the morning hours of the nursery school: show and tell, music and games, snack, and pre-academic exercises. Each would be described as a field situation and data would be taken in each as if it were a separate situation. The events recorded could be the same for all the activities (*e.g.*, frequency of social contacts), or they could be specific to each depending upon the nature of the activity. They could also be a combination of both (*e.g.*, frequency of social contacts and sum total of prolonged productive activity in each pre-academic exercise).

Major variations in social composition in a home study that would be considered disruptive could include the presence of other members of the family, relatives, or friends. In a nursery school, it might be the absence of the head teacher, presence of the child's mother, or the absence of many of the children. These and other events like them would probably call a halt to data collection until the standard situation is returned.

Temporary social disruptions may take many forms. For example, in the home the phone may ring, a salesman may appear, a neighbor may visit; and in the nursery school it might be a holiday preparation, or a birthday party for a member of the group.

In summary, the physical and social conditions in which an ecological study is conducted is specified at the outset. Whether the variations occurring during the study are sufficient to disrupt data collection depends, in large measure on the interactions to be studied, practical considerations, and the investigator's experience in similar situations in the past. However, accounts of changes in physical and social conditions, whether major or minor, are described and noted on the data sheets.

Defining Behavioral and Stimulus Events in Observable Terms

In this method we derive definitions of behavioral and stimulus events from preliminary

investigations in the actual setting. Such pilot investigations are also used to provide preliminary information on the frequencies of occurrences of the events of interest and the feasibility of the situation for study.

A miniature episode in the life of a preschool boy, Timmy, will serve as an example. We start with having the observer make a running description of Timmy's behavior in the play yard in the style she would use if she were a reporter for a magazine.

> Timmy is playing by himself in a sandbox in a play yard in which other children are playing. A teacher stands nearby. Timmy tires of the sandbox and walks over to climb the monkeybars. Timmy shouts at the teacher, saying, "Mrs. Simpson, watch me." Timmy climbs to the top of the apparatus and shouts again to the teacher, "Look how high I am. I'm higher than anybody." The teacher comments on Timmy's climbing ability with approval. Timmy then climbs down and runs over to a tree, again demanding that the teacher watch him. The teacher, however, ignores Timmy and walks back into the classroom. Disappointed, Timmy walks toward the sandbox instead of climbing the tree. A little girl nearby cries out in pain as she stumbles and scrapes her knee. Timmy ignores her and continues to walk to the sandbox.

To obtain a clearer impression of the time relationships among antecedent stimulus events, responses, and consequent stimulus events, the objective aspects of the narrative account are transcribed into a three-column form and each behavioral and stimulus event is numbered in consecutive order.

Setting: Timmy (T.) is playing alone in a sandbox in a play yard in which there are other children playing. T. is scooping sand into a bucket with a shovel, then dumping the sand onto a pile. A teacher, Mrs. Simpson (S.), stands approximately six feet away but does not attend to T.

Time	*Antecedent Event*	*Response*	*Consequent Social Event*
9:14		1. T. throws bucket and shovel into corner of sandbox.	
		2. . . . stands up.	
		3. . . . walks over to monkeybars and stops.	
		4. . . . turns toward teacher.	
		5. . . . says, "Mrs. Simpson, watch me."	
			6. Mrs. S. turns toward Timmy.
	6. Mrs. S. turns toward Timmy	7. T. climbs to top of apparatus.	
		8. . . . looks toward teacher.	
		9. . . . says, "Look how high I am. I'm higher than anybody."	
9:16			10. Mrs. S. says, "That's good, Tim. You're getting quite good at that."

	10. Mrs. S. says, That's good, Tim. You're getting quite good at that."	11. T. climbs down	
		12. . . . runs over to tree.	
		13. . . . says, "Watch me climb the tree, Mrs. Simpson."	
			14. Mrs. S. turns and walks toward classroom.
	14. Mrs. S. turns and walks toward classroom.	15. T. stands, looking toward Mrs. S.	
9:18	16. Girl nearby trips and falls, bumping knee.		
	17. Girl cries.		
		18. T. proceeds to sandbox	
		19. . . . picks up bucket and shovel.	
		20. . . . resumes play with sand.	

Note that a response event (*e.g.*, 5. . . . says, "Mrs. Simpson, watch me.") may be followed by a consequent social event (*e.g.*, 6. Mrs. S. turns toward Timmy.) which may also be the antecedent event for the next response (*e.g.*, 7. T. climbs to top of apparatus.) Note, too, that the three-column form retains the temporal relationships in the narration. Note, finally, that only the child's responses are described. Inferences about feelings, motives, and other presumed internal states are omitted. Even words like "ignores" and "disappointed" do not appear in the table.

On the basis of several such running accounts and analyses a tentative set of stimulus and response definitives are derived and criteria for their occurrence are specified. This material serves as a basis for a provisional code consisting of symbols and definitions. Observers are trained to use the code and are tested in a series of trial runs in the actual situation.

Consider now the problems involved in defining behavioral and stimulus terms, devising codes, and recording events. But first let us comment briefly on the pros and cons of two recording methods.

When discussing the definitions of events and assessing reliability of observers, we refer to observers who record with paper and pencil. In each instance the same could be accomplished by electro-mechanical devices. The investigator must decide which procedure best suits his purpose. For example, Lovaas used instruments to record responses in studies on autistic behavior. He and his co-workers have developed apparatus and worked out procedures for recording as many as 12 responses in a setting. The following is a brief description of the apparatus and its operation (Lovaas, Freitag, Gold, and Kassorla, 1965*b*).

> The apparatus for quantifying behaviors involved two units: an Esterline-Angus 20-pen recorder and an operating panel with 12 buttons, each button mounted on a switch (Microswitch: "Typewriter pushbutton switch"). When depressed, these buttons activated a corresponding pen on the Esterline recorder. The buttons were arranged on a 7 by 14-in. panel in the configuration of the fingertips of an outstretched hand. Each button could be pressed independently of any of the others and with the amount of force similar to that required for an electric typewriter key (p. 109).

An electro-mechanical recording device has certain advantages over a paper-and-pencil

system. It requires less attention, thus allowing the observer to devote more of his effort to watching for critical events. Furthermore, instruments of this sort make it possible to assess more carefully the temporal relationships between stimulus and response events, as well as to record a large number of responses within a given period. On the other hand, paper-and-pencil recording methods are more flexible. They can be used in any setting since they do not require special facilities, such as a power supply.

Defining and Recording Behavioral Events

The main problem in defining behavioral events is establishing a criterion or criteria in a way that two or more observers can agree on their occurrences. For example, if it is desired to record the number of times a child hits other children, the criteria of a hitting response must be clearly given so that the observer can discriminate hitting from patting or shoving responses. Or if it is desired to count the number of times a child says, "No", the criteria for the occurrence of "No" must be specified to discriminate it from other words the child utters, and from non-verbal forms of negative expressions. Sometimes definitions must include criteria of loudness and duration. For example in a study of crying behavior (Hart, *et al.*, 1964), crying was defined to discriminate it from whining and screaming and it had to be (a) "loud enough to be heard at least 50 feet away, and (b) of 5-sec or more duration".

The definitions of complex behavioral events are treated the same way. Studies concerned with such intricate categories of behavior as isolate behavior, fantasy-play, aggressive behavior, and temper-tantrums must establish objective criteria for each class of responses included in the category. We shall elaborate on defining multiple response classes in the following discussion on recording behavioral events.

There are two styles of recording behavioral events in field situations: one consists of logging the incidences of responses (and in many situations, their durations); the other of registering the frequencies of occurrences and non-occurrences within a time interval. Sometimes frequencies and their durations are recorded (Lovaas, Freitag, Gold, and Kassorla, 1965*b*).

Recording the frequencies of occurrences and non-occurrences in a time interval requires the observer to make a mark (and only one mark) in each time interval in which the response occurred. It is apparent that in this procedure the maximum frequency of a response is determined by the size of the time unit selected. If a 5-sec interval were used, the maximum frequency would be 12 per min; if a 10-sec interval were employed, the maximal rate would be six responses per minute, and so on. Thus, in studies with a high frequency of behavioral episodes, small time intervals are employed to obtain high correspondence between the actual and recorded frequencies of occurrences.

There are several approaches to defining and recording single and multiple class responses. One method consists of developing a *specific observational code* for each problem studied. For example, in studies conducted at the Child Behavior Laboratory at the University of Illinois, codes were prepared for attending-to-work behavior, spontaneous speech, and tantruming. The attending-to-work or time-on-task code was employed with a distractible 7-yr-old boy. It included: (1) counting words, (2) looking at the words, and (3) writing numbers or letters. When any of these behaviors occurred at any time during a 20-sec interval, it was scored as an interval of work. In a second study involving a 6-yr-old boy with a similar problem, this code was used with one additional feature: in order for the observer to mark occurrence in the 20-sec interval, the child had to engage in relevant behavior for a minimum of 10 sec. The reliability on both codes averaged 90% for two observers over 12 sessions. (See Section 3 for our method of determining reliability.)

A code for spontaneous speech was developed for a 4-yr-old girl who rarely spoke. Incidences of speech were recorded whenever she uttered a word or words which were not preceded by a question or a prompt by a peer or teacher. Although this class of behavior was somewhat difficult to discriminate, reliability averaged 80% for two observers over 15 sessions.

Tantrum behaviors exhibited by a 6-yr-old boy were defined as including crying, whining, sobbing, and whimpering. The average reliability for this class of behavior was 80% for two observers over 11 sessions.

In contrast to this more or less vocal form of tantrum behavior, a code developed in another study on temper-tantrums centered around gross motor responses of an autistic child (Brawley, Harris, Allen, Fleming, and Peterson, 1968, in press.) Here a tantrum was recorded whenever the child engaged in self-hitting in combination with any one of the following forms of behavior: (1) loud crying, (2) kicking, or (3) throwing himself or objects about.

Another method of defining and recording responses is to develop a *general observational code,* one that is inclusive enough to study many behaviors in a given field situation. An example of such a code is the one prepared by the nursery school staff at the University of Washington. In essence, verbal and motor responses are recorded in relation to physical and social events using a three or four track system. Tables 1 and 2 show sample lines from data sheets. Each box represents an interval of 15 sec.

In Table 1, which is a segment of a data sheet for a nursery school girl who changed activities with high frequency, entries were made in the boxes in the top row to indicate occurrences of vocalizations (V). Entries were made in the middle row to shop proximity (P) or physical contact (T) with another person, and in the bottom row to indicate contact with physical objects (E) or with children and whether the interaction was parallel play (A) or shared play (C). Other marks and symbols are added in accordance with the problem studied. For example, each single bracket in Table 1 indicates leaving of one activity and embarking on another. During the 6-min period in which records were taken (24 15-sec intervals), the child changed her activity 12 times. During that time the teacher gave approval five times contingent upon her verbal or proximity behavior as indicated by X's above the top line (10, 11, 16, 17, and 18). A tally of the data indicated that she spent most of the 6-min period alone or in close proximity to another child, sometimes on the same piece of play equipment. During three intervals (16, 17, and 18) she talked (V), touched (T), and engaged in physical interaction with another child (C). Even though rate of activity change, and not peer interaction, was the subject of the study, the other data on social behavior provided interesting information: decline in rate of activity change was related to an increase in rate of appropriate peer behavior.

This code can be readily modified to handle more complex interactions. For example, it was used to record the behavior of a nursery school boy who shouted epithets, kicked, and hit other children. Ordinarily these aggressive acts would appear in the record sheets undifferentiated from a non-aggressive interaction. To differentiate them from other behaviors the symbol letter was circled if the behavior met the criteria of an aggressive act. As shown in Table 2, intervals 13, 22, and 23 contain a "V" with a circle, Ⓥ, which indicates aggres-

Table 1

Sample line from a data sheet of nursery school girl who changed activities with high frequency.

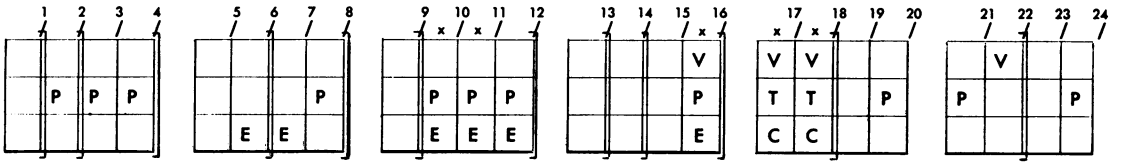
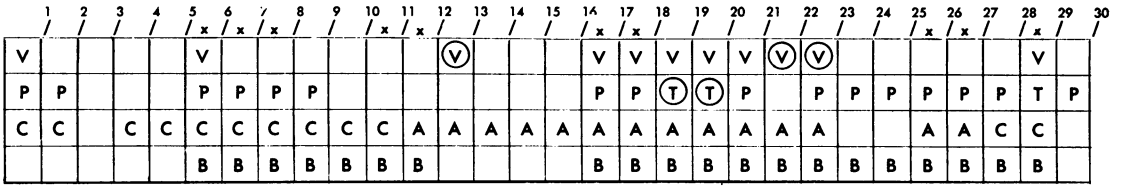


Table 2

Sample Line from a Data Sheet of Nursery School Boy Displaying Aggressive Behaviors



sive verbalizations, while intervals 19 and 20 contain a "T" with a circle, Ⓣ, which indicates physical "attack" (actual hitting, kicking, or pinching). Another bit of information was incorporated in the recording system. The letter "B" was entered in the fourth row to indicate that the child was playing with or being aggressive to a specific nursery school boy named Bill. This additional notation was made midway in the study when teachers observed that the subject and Bill usually behaved aggressively toward each other. Data collected before this change served as a baseline against which to judge the effects of changing social contingencies. Subsequently, teachers gave approval contingent on nonaggressive interactions between these boys as shown by the X's above intervals 6, 7, 8, 11, 12, 17, 18, 26, 27, and 29.

Another general observational code, tailored for analysis of pupils' behavior in the elementary school classroom, has been devised by Thomas and Becker (1967). Like the nursery school code, it consists of symbols and definitions designed to cover the range of interactions that may take place in the field situation defined by the classroom.

Defining and Recording Stimulus Events

The ease or difficulty of defining a stimulus class is related to its source. It has been pointed out (*e.g.*, Bijou and Baer, 1961) that some stimuli originate in natural and man-made things, some in the biological make-up of the subject himself, and some in the behavior of people and other living organisms. Consider briefly each source in turn.

Defining stimuli from physical things does not pose a difficult problem since physical objects are usually available for all to see. All that is required is that these stimuli be described in the usual physical dimensions of space, time, size, velocity, color, texture, and the like.

Defining stimuli which originate in the biological make-up of the subject is beset with difficulty mostly because of their obscurity under any circumstance and particularly under field conditions. Consider what must be available to an observer if he is to record in objective terms the duration, intensity, or frequency of stimuli involved in a toothache, "butterflies" in the stomach, general bodily weakness, dizziness, and hunger-pangs. Instruments would be needed to make visible all sorts of internal biological events; and for the most part, these are not yet available in practical forms. It seems clear that at present, field methods of research, especially with human beings, are not appropriate for describing biologically anchored variables. Research on these variables must be postponed until it is practical to monitor physiological actions through cleverly designed telemetric devices. But it should be stressed that the exact role of specific biological variables *must* be studied at some time for a thorough functional analysis of psychological behavior (defined here as the interaction of a total functioning biological individual with environmental events).

Defining social stimuli, or stimuli which evolve from the action of people, ranges in difficulty between physical and biological events. This is so because social events, like physical and biological events, must in many instances be described in terms of their physical dimensions, and as is well known, the components of social stimuli can be terribly subtle and complex. For the reader interested in a further analysis of social events within the framework of a natural science, Skinner's discussion is recommended (1953, pp. 298-304).

In field studies, the procedure for defining and recording social stimuli is the same as that for defining and recording response events, since social events are treated as the responses of people in antecedent or consequent relationships to the behavior of the subject. Therefore, the entire previous section on defining and recording behavioral events pertains to defining and recording social events.

Some social stimuli, like response stimuli, may consist of a single class of behavior on the part of an adult or a child and may be recorded on the basis of frequency or its occurrence or nonoccurrence within a time interval. Examples of single-class antecedent stimuli are simple commands and requests, *e.g.*, "Start now," "Gather around in a circle," "Come, let's ride the trikes." Examples of single-class consequent stimuli are confirmations ("Right"), disconfirmations ("Wrong"), approval ("Good") and disapproval ("You play too rough.")

Other social stimuli may be composed of several classes of behavior stemming from one person or several in concert. As in the case of defining multiple response classes, criteria for

each subclass in the group may constitute a code. A specific observational code may be developed to describe social events in a specific situation for a specific study. For example, in a study of autistic behavior, adult attention was defined as: "(1) Touching the child; (2) being within two feet of and facing the child; (3) talking to, touching, assisting or going to the child" (Brawley, *et al.* 1968). With such criteria the investigator catalogued the types of behaviors which constituted social interaction involving attention and excluded other stimuli originating in the behavior of an adult in contact with the subject.

General observational codes for social events, like those for response events, have also been devised to study many problems in a general type of field setting. For example, Becker and Thomas (1967) have developed a comprehensive code for recording the teacher's behavior in an elementary classroom situation.

Which classes of behavior-environmental interactions will be selected for study will depend on the purpose of the investigation; the maximum number, however, will be limited by the practical considerations. Studies requiring detailed analyses of many response classes may be planned as a series. The first dealing with grossly defined classes and the others with more and more progressively refined categories. For example, the first study may be concerned with the frequency of social contacts with adults and peers, and the second with specific verbal and motor responses directed to specific adults (teachers and parents) and peers (boys and girls).

Assessing Observer Reliability

Disagreements between observers may be related to inadequacies in (1) the observational code, (2) the training of the observers, or (3) the method of calculating reliability.

The observational code. Problems of defining and recording behavioral and stimulus events have been discussed in Section 2. Observer reliability is directly related to the comprehensiveness and specificity of the definitions in the observational code. Generally it is advisable to devise codes with mutually exclusive event categories, each definition having criteria that do not occur in any other definition.

Training of observers. Even when a code is completely serviceable, two observers may not necessarily record the occurrence of the same event at the same time unless each has been adequately trained in using the code and in controlling his behavior while observing and recording.

For example, training might begin by familiarizing the observer with the tools for recording, *e.g.*, the clipboard, stopwatch, and data sheets. This might be followed by an orientation to the code and exercises in recording behavioral events. A film or video tape of sequences similar to those in the actual situation might be used to provide supplementary experiences.

It is often helpful to have a second observer to record along with the first observer. During trial recordings the observers can indicate to each other the behaviors being scored and uncover misunderstandings regarding the nature of the code or ambiguities in the definition of particular responses. Such a procedure reduces interpretation on the part of the observer and can contribute to an improved code.

Since it is relatively easy for the observers to slip an interval in the course of a long recording session, they should be instructed to note the beginning of certain activities, *e.g.*, story time, snack, nap, *etc.* This allows them to determine easily when they are out of phase with one another. Slips may also result from inaccurate stopwatches. Watches should be periodically tested by starting them simultaneously and checking them a few hours later.

After training on the proficient use of the code the observer might then be given instruction on how to conduct himself while observing and recording. Thus, he might be told how to refrain from interacting with the subject, *e.g.*, ignore all questions, avoid eye-contact, and suppress reactions to the subject's activities as well as those associated with him. He might also be instructed in moving about to maintain a clear view of the subject yet not make it obvious that he is following him.

Method of calculating reliability. The reliability index is to some degree a function of how it is calculated. Suppose we have data from two observers showing the frequency of a class of events taken over 1 hr. Unless the sums obtained by each observer are equal, the smaller sum is divided by the larger to obtain a percentage of agreement. If the sums are identical the reliability index would be 100. This method is often used when the investiga-

tor is interested in frequencies *per se*, since the measure obtained gives only the amount of agreement over the total number of events observed. It does not indicate whether the two observers were recording the same event at exactly the same time. Thus, it might be possible that one observer was recording few behaviors during the first half hour and many during the second, while the second observer was doing just the opposite. To ascertain whether this is the case, one could divide the period of observation into small segments and calculate the reliability of each. Agreements over progressively smaller segments give confidence that the observers are scoring the same event at the same time. One may assess the agreement over brief intervals such as 5 or 10 sec. Reliability is calculated by scoring each interval as agree or disagree (match or mismatch) and dividing the total number of agreements by the number of agreements plus the number of disagreements. Note that one may score several agreements or disagreements in an interval if a number of events are being recorded simultaneously as shown in Tables 1 and 2. In this case the interval is broken down according to the number of different events recorded, with each event scored as a match or mismatch.

The reliability index may also be influenced by the frequency of response under study. When a behavior is displayed at a very low rate, the observer will record few instances of occurrence and many of nonoccurrence. In this situation the observers could disagree on the occurrence of the behavior yet still show high reliability due to their agreement on the large number of intervals where no behavior was recorded. A similar problem exists with regard to high-frequency behaviors. Here, however, the observers may disagree on the nonoccurrence of the behavior and agree on occurrence, because of the frequency of the latter. The problem may be resolved by computing not one but two reliability coefficients, one for occurrence and one for nonoccurrence.

In some cases the requirement of perfect matching of intervals may be relaxed slightly. Thus, behaviors recorded within one interval (especially if the interval is short) may also be considered as instances of agreement for reliability purposes. A technique of noncontinuous observing may also increase reliability (O'Leary, O'Leary, and Becker, 1967). In this procedure the observers record for shorter portions of time. For example, instead of taking continuous 10-sec observations, the observer might record for 10 out of every 15 sec, or for 20 out of every 30 sec. During the period in which the observer is not attending to the child, he should be recording the behaviors just observed.

The use of a second observer does not insure high reliability of recording; it is possible for both observers to agree on the scoring of certain events and at the same time be incorrect (Gewirtz and Gewirtz, 1964). Both observers might record some events which should not be noted and ignore others which should. Hence, a third observer might be used on occasion to determine if this possibility exists.

Collecting, Analyzing, and Interpreting Data

Data collection. Final data collection is begun as soon as it is evident that the observers are adequately trained, the field situation is feasible, and the subject has adapted to the presence of the observers.

Whether the investigator collects data during all of the time available for observation or takes time samples will depend upon many factors, including the purpose of the study, the nature of the data, and the practical considerations. Regardless of the frequency with which observations are made, it is recommended that the data be plotted at regular intervals to provide a kind of progress chart. A visual account of the fluctuations and trends can help the investigator make important decisions, *e.g.*, setting up the time for the next reliability evaluation or establishing the termination time for a phase of the study.

Data analysis. Up until now we discussed the investigators' activities in relation to the interactions between the observer and the field events. The investigator was viewed as a critic, watching the observer record the events in a natural ecology. Thus, in the data collection phase of a study the investigator's role is somewhat similar to that of a motion picture director evaluating what the camera is recording in relation to the scene as he sees it. In this section on data analysis and in the next on interpretation, we shall consider what the investigator does, not in relation to the recording equipment and field events, but in relation to the data collected.

Basically, in data analysis the investigator looks at the data collected to "see what is there". Usually he finds that making one or several transformations in the raw data helps him to see more clearly the relationships among the events observed. Transformational procedures might consist of converting the frequency counts into graphic, tabular, verbal, arithmetical, or statistical forms. Exactly which operations he performs on the data will depend on the purpose of the study, the nature of the data, and his theoretical assumptions about what can or cannot be demonstrated by a descriptive field study.

Usually, data analysis begins when data collection ends. However, as noted previously, an investigator might graph the data while the study is in progress. Under these circumstances data analysis might consist of revising and refining the graphs and making other transpositions to show the relationships among the subparts of the data.

Data collected in terms of rate are usually plotted in a graphic form with responses on the vertical axis and time of the horizontal axis. Points on the chart may represent either discrete or cumulative values. Discrete values are the sums or means for each successive session; cumulative values are the sums or means for all previous sessions. Therefore, curves with discrete values might go up, stay at the same level, or go down; cumulative curves might also go up or stay at the same level. Cumulative curves do not go down. A decrease in the frequency of a response is shown in the curve as a deceleration in rate (bends toward the horizontal axis); an increase in frequency as an acceleration (bends toward the vertical axis); a constant frequency as no change in rate; and a zero frequency as a horizontal line.

In most instances graphic presentations are made more meaningful when accompanied by percentage values. In addition, it is often advantageous to show percentages of occurrences in the different conditions and sub-conditions of the field situation.

Viewing the interactions in selected time periods (early morning, and late morning) or around certain events (before and after mealtime) as populations, statistical analyses may be made to assess the nature of and the reliability of differences observed.

Interpretation of findings. Essentially, interpretation of findings consists of the investigator's statements on what is "seen" in the data together with his conception of their generality. Such statements are the *raison-d'être* of an investigation.

Obviously, an investigator is free to interpret his findings in any way he chooses. The investigator who accepts the assumptions of a natural science approach to psychology seeks to limit his interpretations to empirical concepts and relationships consistent with his observations and the analytical operations made upon the products of his observations. Hence, in a descriptive field study his interpretations would usually consist of a discussion of what was found in the situation with comparisons to other findings obtained under functionally similar conditions. Conclusions on the similarities and differences between his findings and others would be incorporated in his argument for the generality of his findings. Interpretations in an experimental field study would depend on the number and type of manipulations employed and would usually be limited to describing the functional relationships obtained.

Illustrative Study

Using the procedures previously described, a study was undertaken to obtain a descriptive account of a boy in a laboratory nursery school at the University of Illinois. The nursery school curriculum and the practices of the teaching staff of this school were based on behavioral principles (Skinner, 1953, and Bijou and Baer, 1961).

Subject and field situation. The subject (Zachary) was typical of the children in the nursery school in the judgment of the teachers. He was 4.5 yr old, of high average intelligence (Peabody IQ 116) and from a middle socio-economic class family. On the Wide Range Achievement Test he scored kindergarten 3 in reading, pre-kindergarten 5 in spelling, and kindergarten 6 in arithmetic.

The nursery school consisted of a large room, approximately 21 by 40 ft. Evenly spaced along one wall were three doors which led to three adjacent smaller rooms. One of these rooms was a lavatory, the second contained paints, papers, and other equipment, and the third a variety of toys. Nearby was a large table and several chairs used for art activities and snack. Opposite these rooms along the other wall were several tables separated by

brightly colored, movable partitions. In these booths, the children worked on academic subjects.

The school was attended by 12 children, six boys and six girls, between 4 and 5 yr of age. The teaching staff consisted of a full-time teacher and an assistant teacher, and depending on the time of day, one to three undergraduates who assisted in administering new programs in reading, writing, and arithmetic.

In general, the morning program was as follows:

9:00-10:00 Art, academic, and pre-academic work
10:00-10:30 Free play
10:30-11:00 Snack
11:00-12:00 Academic work, show-and-tell, and storytime.

A typical morning might begin with art. At this time, 8 to 10 children sat around a large table working with various materials. During this activity each child in turn left the group for 10 to 20 min to work on writing or arithmetic. While engaged in writing or arithmetic the child worked with a teacher in one of the booths. After completing his assigned units of work he returned to his art activity and another child left the group to work on his units of writing or arithmetic. After all the children had participated in these academic subjects, the art period was terminated and was followed by play. During play, the children were free to move about, often spending much of the time in either of the smaller nursery school rooms playing with blocks or other toys. After approximately 30 min of play, the youngsters returned to the large table for a snack of juice and cookies. While eating and drinking they talked spontaneously and informally with their teachers and peers. Following snack time some of the children participated in reading while the others gathered for show-and-tell or storytime. During storytime the children sat on the floor in a group while the teacher read and discussed the story. In show-and-tell, instead of the teacher leading the group, each child had a chance to stand by the teacher in front of the group, and show an object he had brought from home and tell about it. As they did during the art period the children left the group one at a time for a period of reading. Because of variations in the amount of time a child spent on academic subjects, a child did not engage in all of these activities every day.

Behavioral and stimulus events recorded. The behaviors recorded were of two general categories: social contacts and sustained activities. Social contacts included verbal interchanges and physical contacts with children and teachers. Sustained activities involved behaviors in relation to the school tasks. The specific observational code developed for the study is presented in Table 3.

Observation began 3.5 weeks after the start of the school year and covered a 3-hr period in the morning. The observations were taken on 28 school days. The observer sat a few feet from the subject and discretely followed him as he moved from one activity to another in the nursery school room. Every 10 sec the teacher recorded the occurrence or nonoccurrence of events defined in the code. The data sheet was similar to that shown in Table 1; however, only the first and second rows were used.

Observer reliability. The reliability of observation and the adequacy of the behavioral code was evaluated several times throughout the study by having a second observer record stimulus and response events. Reliability was calculated by scoring each interval as a match or mismatch and dividing the total number of agreements by the number of agreements plus disagreements. Four checks on social contacts showed agreements of 75, 82, 85, and 87%. Three checks on sustained activity yielded agreements of 94, 95, and 97%. Thus, average agreement on social contacts exceeded 82% while average agreement of sustained activity exceeded 95%.

Analysis of Data

Social contacts. Data were gathered on Zachary's social behaviors in informal activities of art, play, snack, storytime, and show-and-tell. They will be described and samples of the detailed accounts in art and snack will be presented in graphic form. The youngster's most dominant behavior during the art period, shown in Fig. 1, was talking to others (14% of the time).

Teachers and peers talked with him about equally, an average of 8 and 7% respectively. Physical contacts between Zachary, teachers, and peers were low, around 1 to 2%.

The child's verbal behavior to peers during the play period was higher than in the art

Table 3

Observational Code for Describing the Behavior of a Boy in a Laboratory Nursery School

Symbol	*Definition*
	First Row (Social Contacts)
	*S* verbalizes to himself. Any verbalization during which he does not look at an adult or child or does not use an adult's or child's name. Does not apply to a group situation.
	*S* verbalizes to adult. *S* must look at adult while verbalizing or use adult's name.
	*S* verbalizes to child. *S* must look at child while verbalizing or use child's name. If in a group situation, any verbalization is recorded as verbalization to a child.
	Child verbalizes to *S*. Child must look at *S* while verbalizing or use *S*'s name.
	Adult verbalizes to *S*. Adult must look at *S* while verbalizing or use *S*'s name.
	Adult gives general instruction to class or asks question of class, or makes general statement. Includes storytelling.
	*S* touches adult. Physical contact with adult.
	*S* touches child with part of body or object. Physical contact with child.
	Adult touches *S*. Physical contact with adult.
	Child touches *S* with part of body or object. Physical contact with child.

Symbol	*Definition*
	Second Row (Sustained Activity)
	*Sustained activity in art.* *S* must be sitting in the chair, facing the material and responding to the material or teacher within the 10-sec interval. Responding to the material includes using pencil, paint brush, chalk, crayons, string, scissors or paste or any implement on paper, or working with clay with hands on clay or hands on implement which is used with clay, or folding or tearing paper. Responding to the teacher includes following a command made by an adult to make a specific response. The behavior must be completed (child sitting in his chair again) within two minutes.
	*Sustained activity in storytime.* *S* must be sitting, facing the material, or following a command given by the teacher or assistant. If the *S* initiates a verbalization to a peer, do not record sustained activity in the 10-sec interval.
	*Sustained activity in show-and-tell.* *S* must be sitting, facing the material, or following a command given by the teacher. If the *S* initiates a verbalization to a peer, do not record sustained activity in that 10-sec interval.
	*Sustained activity in reading.* *S* must be sitting in the chair, facing the material and responding to the material or the teacher within the 10-sec interval.
	*Sustained activity in writing.* *S* must be sitting in the chair, facing the material and responding to the material or the teacher within the 10-sec interval. Responding to the material includes using the pencil (making a mark), or holding the paper or folder. Responding to the teacher includes responding verbally to a cue given by the teacher.
	*Sustained activity in arithmetic.* *S* must be sitting in the chair, facing the material and responding to the material or the teacher within the 10-sec interval. Responding to the material or teacher includes using the pencil or eraser or holding the paper or folder or responding verbally to cue.
	Sustained activity did not occur in interval.

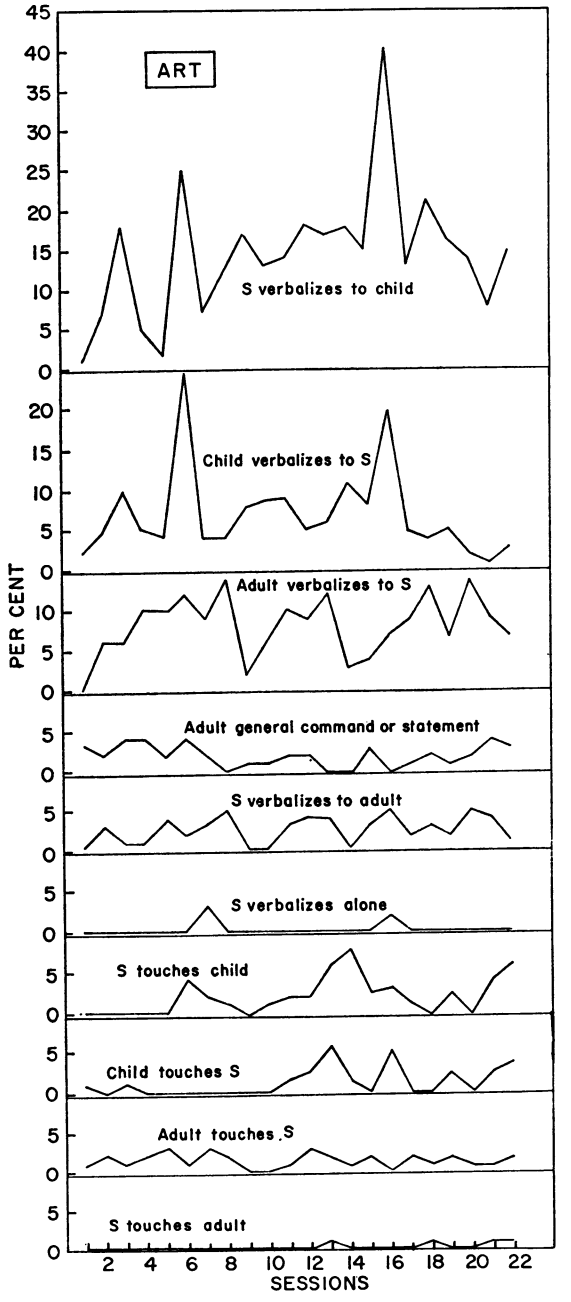


Fig. 1. Social contact during art.

period. He talked to his friends on an average of 38%; they talked to him on an average of only 10%. Verbal exchanges with teachers were low (an average of 2.5%). Zachary touched other children 7% of the time on the average and they reciprocated on an average of 3%. Physical contacts with teachers were relatively infrequent.

As in the art and play periods, Zachary's social interactions during snack time, shown in Fig. 2, consisted mostly of talking to his classmates, an average of 21%. They, in turn, talked to him only an average of 7%. During this period the teacher's general commands (instructions addressed to the group) were relatively high, averaging 7% in contrast to the 2% during art and play. Physical contacts with other children were low, as in art and play, about 3%.

Compared to the art, play, and snack peri-

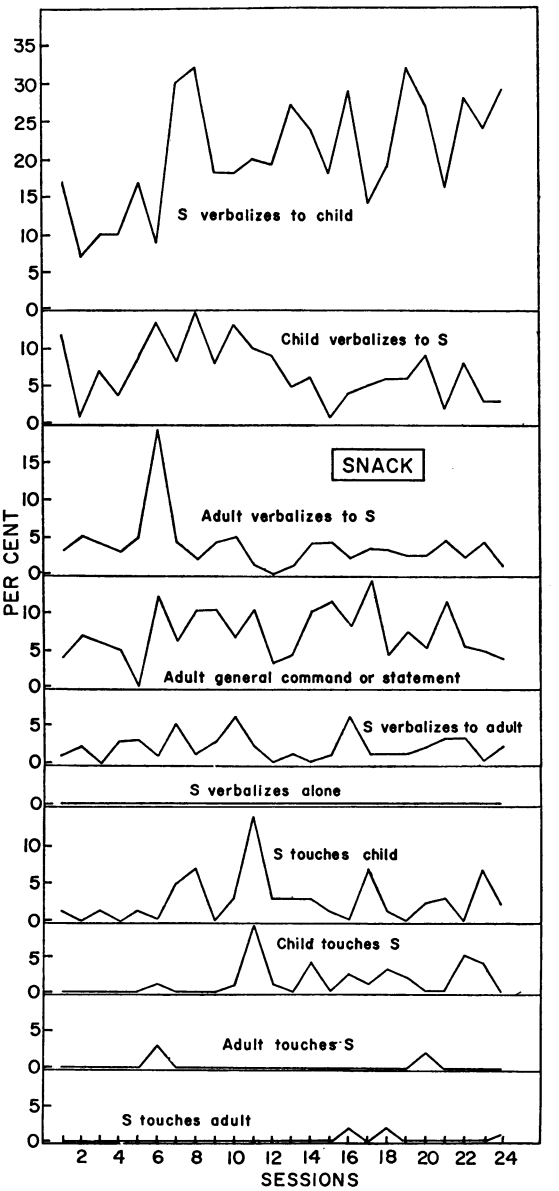


Fig. 2. Social interaction during snack.

ods, Zachary's verbalizations to peers and to teachers were low (8 and 4% respectively), and the number of times he touched children (10%), and children touched him were also relatively low (2%). Storytime had a high frequency of teacher's general commands and statements (average of 73%) since this category was scored when the teacher read and discussed the stories.

In show-and-tell, Zachary's social behavior was similar to that during storytime. He talked to other children 14% of the time and touched them 9% of the time. Zachary physically contacted teachers about 1% of the time and they reciprocated about 3% of the time.

In respect to Zachary's social behavior during the academic periods, these data clearly indicate that the teacher talked to Zachary a great deal during the reading (an average of 69%), writing (an average of 71%), and arithmetic periods (an average of 58%), and the child talked to the teacher with high frequency, particularly in reading (an average of 44%) and arithmetic (an average of 41%). In writing he talked to the teacher only 3% of the time. There were also a few instances in which the teacher touched Zachary and rare occasions in which Zachary interacted socially with other children. Figure 3 is a detailed graphic account of his social behavior during the writing period.

Sustained activity. For the observer to mark the occurrence of sustained activity, Zachary had to respond in a manner appropriate for a particular school activity. (See second part of Table 3). For example, during art, the child had to be sitting in his chair, facing the art materials and manipulating them during each 10-sec interval. Similar definitions were used for other situations and periods. Given these definitions, the results show a generally high level of sustained activity in all phases of the morning program. Daily rates of sustained activities in art, storytime, and show-and-tell range between 70 and 99% with an average of 89% for art, 95% for storytime, and 88% for show-and-tell. See Fig. 4 for variations from session to session in Zachary's sustained behavior during art. Sustained activity in reading, writing, and arithmetic, range from 90 to 100% over the days observed with an average of 97, 95, and 96% respectively. See Fig. 5 for variations in the child's sustained behavior in writing. Due to the limited availability of the observer, and the fact that not every activity occurred every day, the number of observations on each activity varied.

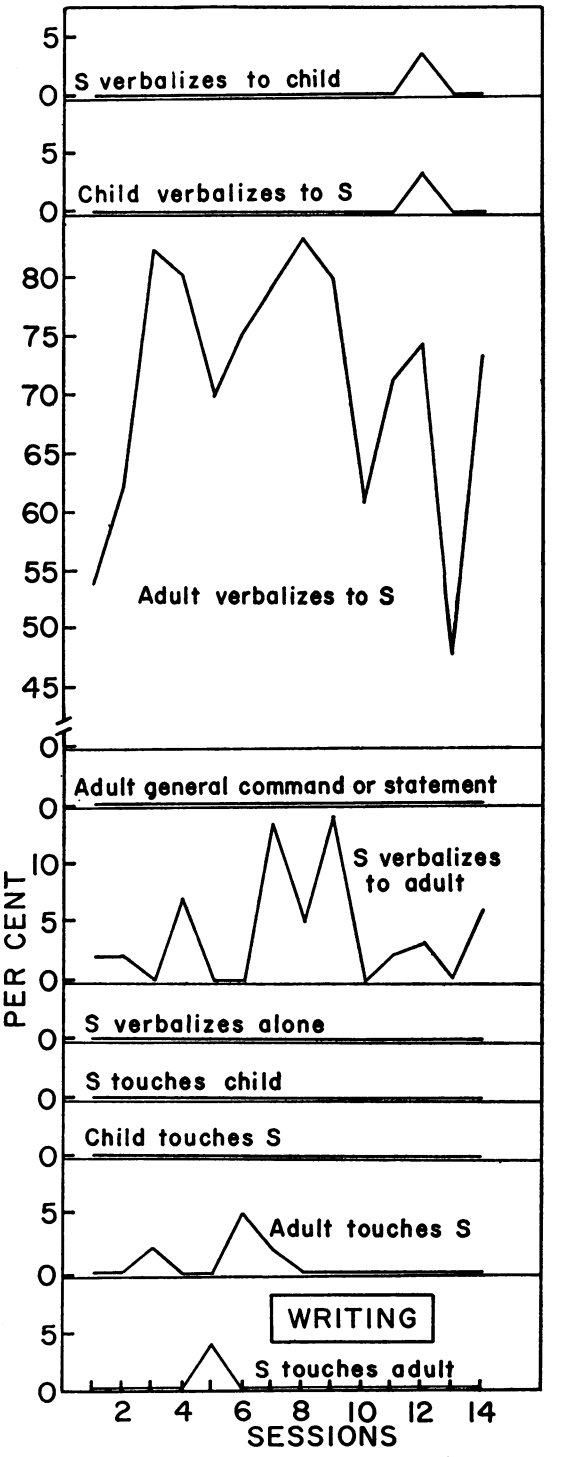


Fig. 3. Social behavior during writing.

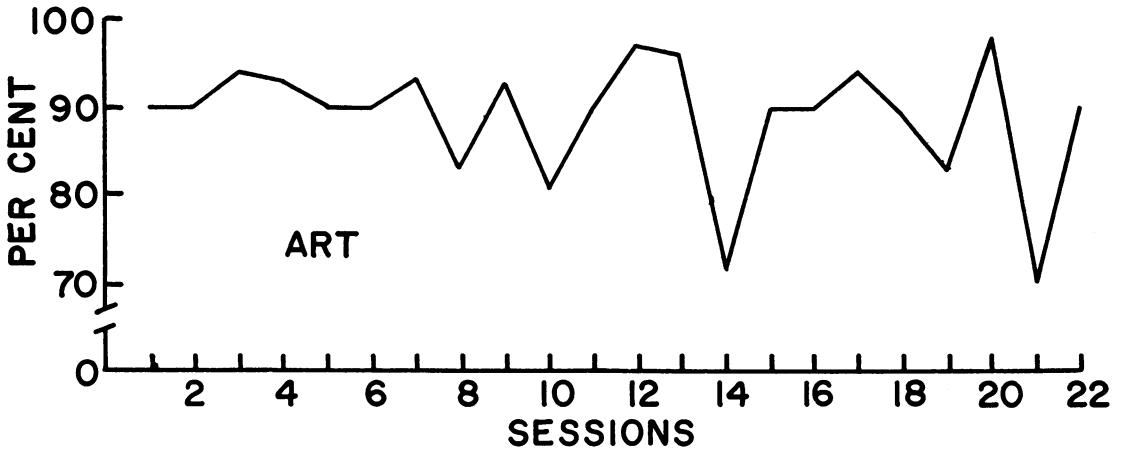


Fig. 4. Sustained activity during art.

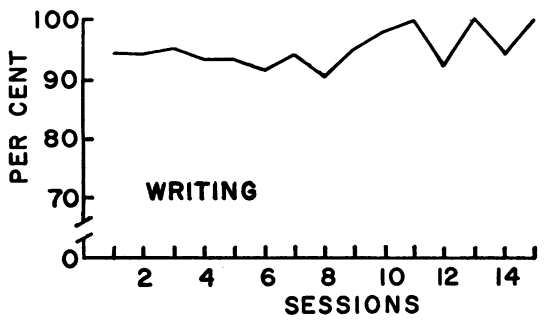


Fig. 5. Sustained activity during writing.

DISCUSSION

A descriptive account of the behaviors of a boy during the morning hours in a laboratory nursery school was obtained in terms of the frequency of occurrence of objectively defined stimulus and response events. The account shows rates of changes in social interactions (verbal and physical contacts) and sustained activities during eight periods of the school morning.

In the informal activities of the nursery school in which the youngster performed on an individual basis, as in art, free play, and snack time, the subject talked to his peers and teachers to a moderate degree. His peers and teachers responded to him verbally to a lesser extent. He talked more than he listened and over the period of the study, his verbal output increased. Physical interactions with peers and teachers in these situations were at a relatively low level. Finally, the youngster's sustained activity in the art period was high (between 70 and 98%) and became more variable, on a day-to-day basis, during the second half of the study. In the other two informal activities, storytime and show-and-tell, the child participated as a member of a group in which the teacher's verbal behavior was prominent, especially during storytime. In these two situations the child talked to others less, but as in art, free play, and snack time, he talked more than he listened. In storytime and show-and-tell he engaged in some body contacts with peers and teachers, yet his sustained activity on nursery school tasks was high, with a range of 90 to 99% for the former and 70 to 95% for the latter.

In the more structured activities of reading, writing, and arithmetic the teacher's verbal behavior to the child was high and his verbal behavior to her was correspondingly high, particularly in reading and arithmetic. During academic exercises all other social interactions were zero or near zero, and the child's sustained activities were consistently high over days, ranging from 90 to 100% of the time.

The data gathered in this study can serve two main purposes. First, they can provide normative information on behavior in a laboratory preschool. Thus, it might be interesting to compare this child's rates of response obtained in this study after 3.5 weeks of school with his rates during the last month of the school year. It might also be interesting to compare this child's behavior with another child's in the same nursery school. Such a comparison might be especially valuable if someone claimed that the second child's behavior was deviant. In addition, it might be informative to compare this child's behavior with a

comparable child in a community-operated nursery school. Second, the data suggest certain relationships between the behavior of the subject, the teacher, and other children. Thus, the investigator might use the data as a baseline for an experimental study in which conditions are manipulated to test for possible functional relationships.

REFERENCES

Baldwin, A. L., Kalhorn, J., and Breese, F. H. The appraisal of parent behavior. *Psychological Monographs*, 1949, **63**, No. 299.

Barker, R. G. and Wright, H. F. *Midwest and its children: the psychological ecology of an American town*. New York: Harper & Row, 1955.

Becker, W. C., Madsen, C. H., Jr., Arnold, Carole R., and Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. *Journal of Special Education*, 1967, **1**, 287-307.

Becker, W. C. and Thomas, D. R. A revision of the code for the analysis of a teacher's behavior in the classroom. Unpublished manuscript, 1967.

Bijou, S. W. and Baer, D. M. *Child development: a systematic and empirical theory*. Vol. 1, New York: Appleton-Century-Crofts, 1961.

Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., and Tague, Cecilia. Classroom behavior of retarded pupils with token reinforcement. *Journal of Experimental Child Psychology*, 1965, **2**, 219-235.

Brawley, Eleanor R., Harris, Florence R., Allen, K. Eileen; Fleming, R. S., and Peterson, R. F. Behavior modification of an autistic child. *Behavioral Science*, 1968, in press.

Gewirtz, Hava and Gewirtz, J. L. A method for assessing stimulation behaviors and caretaker-child interaction. Unpublished manuscript, 1964.

Harris, Florence R., Wolf, M. M., and Baer, D. M. Effects of adult social reinforcement on child behavior. *Young Children*, 1964, **20**, 8-17.

Hawkins, R. P., Peterson, R. F., Schweid, Edda, and Bijou, S. W. Behavior therapy in the home: Amelioration of problem parent-child relations with the parent in a therapeutic role. *Journal of Experimental Child Psychology*, 1966, **4**, 99-107.

Honig, W. K. Introductory remarks. In W. K. Honig (Ed.) *Operant behavior: areas of research and application*. New York: Appleton-Century-Crofts, 1966.

Jensen, G. D. and Bobbitt, Ruth A. Implications of primate research for understanding infant development. *The exceptional child*, Vol. 1, J. Hellmouth (Ed.), Special Child Publications, Seattle, Washington, 1967.

Johnston, Margaret S., Kelley, C. Susan, Harris, Florence R., and Wolf, M. M. An application of reinforcement principles to development of motor skills of a young child. *Child Development*, 1966, **37**, 379-387.

Lovaas, O. I., Freitag, G., Gold, Vivian J., and Kassorla, Irene C. Recording apparatus and procedure for observation of behaviors of children in free play settings. *Journal of Experimental Child Psychology*, 1965, **2**, 108-120. (*b*)

Lovaas, O. I., Freitag, G., Gold, Vivian J., and Kassorla, Irene C. Experimental studies in childhood schizophrenia: analysis of self-destructive behavior. *Journal of Experimental Child Psychology*, 1965, **2**, 67-84. (*a*)

O'Leary, K. D., O'Leary, Susan G., and Becker, W. C. Modification of a deviant sibling interaction pattern in the home. *Behaviour Research and Therapy*, 1967, **5**, 113-120.

Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.

Wolf, M. M., Risley, T. R., and Mees, H. L. Application of operant conditioning procedures to the behavior problems of an autistic child. *Behaviour Research and Therapy*, 1964, **1**, 305-312.

Wright, H. F. *Recording and analyzing child behavior*. New York: Harper & Row, 1967.

Received 22 April 1968.