Ann. Rev. Sociol. 1984. 10:263–82

SYSTEMATIC FIELD OBSERVATION

George J. McCall

Department of Sociology, University of Missouri, St. Louis, Missouri 63121

Abstract

This chapter reviews, from a sociological standpoint, the methodological and empirical literature on systematic field observation in several fields. Three historical waves of interest in and refinement of systematic field observation techniques are identified, and sociological contributions to each are noted. These techniques have been applied in studies that cover a variety of research purposes and topical subject matters. There are three distinct conceptual and methodological paradigms advocated by researchers who do systematic field observation: the behavior variables approach, the behavior element approach, and the constructivist approach. Debates over how to construct any observation system reflect clashes among these alternative paradigms. I propose a role-expectations view of observation systems, which requires a reconceptualization of (*a*) the process of developing an observation system and (*b*) the nature, sources, and management of observational error. In the final section, I review empirical investigations of observer cheating, bias, drift, and reactive effects, as well as psychometric advances in assessing the reliability and validity of observational data. Due to recent developments, systematic field observation is better suited than ever before to sociologists' needs and interests.

INTRODUCTION

The systematic and naturalistic observation of human social interaction is a well established but comparatively little used tool in social research. Routine application of this method has been largely confined to certain of the less experimental branches of applied psychology—developmental, clinical, educational, and industrial. Wider interest in systematic field observation has been quite cyclical, with waves of increased popularity and the attendant methodological advance occurring roughly every 20 years.

0360-0572/84/0815-0263$02.00

The behaviorist ideas popularized by John B. Watson in the 1920s inspired extensive observational study of the social behavior of children, so that by the early 1930s the fundamental methodology of systematic field observation had been developed in its essentially modern form. The role that sociologists such as Dorothy Swaine Thomas, F. Stuart Chapin, and Stuart A. Rice played in these early developments is almost forgotten today. During the 1950s, the rediscovery of the primary group directed much attention to the systematic observational analysis of group processes and social interaction; the contributions made by numerous sociologists, particularly Robert F. Bales, are more widely known.

In this essay, I will review developments within the third and most recent wave of activity, which peaked during the 1970s. Several factors account for the renewal of a widespread interest in systematic field observation during that decade. Social research had been seriously criticized for its excessive reliance on interview and questionnaire methods (Webb et al 1966), and psychology was experiencing a crisis of confidence in laboratory experimentation that engendered greater interest in field settings and more naturalistic methods. The clinical professions were losing faith in the "talking" therapies and turning increasingly to Skinnerian forms of behavior therapy and behavioral assessment. More positively, an interest in social interaction was being pursued more aggressively across all of the social and behavioral sciences, resulting in conceptual ferment as to the very nature of interaction process.

The entry of new actors with different perspectives reinvigorated the methodology of systematic field observation. Sociologists naturally linked observation to their tradition of survey research (Reiss 1968b, 1971a, 1976). The neobehaviorist clinical psychologists contributed strong psychometric concerns for reliability, validity, and generalizability of measurements (Cone 1977, Berk 1979, Mitchell 1979); the resulting theory of generalizability (Cronbach et al 1972) became a key tool in observational methodology. Yet, it was the emergence of human ethology that spurred the greatest methodological advances (Hutt & Hutt 1970, Blurton Jones 1972). By viewing human social behavior within the framework of evolutionary biology, ethologists brought with them the separate but parallel observational methodology developed by biologists to study animal behavior and social organization—a tradition far more sophisticated in its techniques for quantitative analysis (Lehner 1979, Colgan 1978, Hazlett 1977).

Observation, in all its varieties, is a process of "planned, methodical watching that involves constraints to improve accuracy" (Weick 1968:358). Thus, while nearly everyone who goes to a zoo *sees* the animals there, and many even *watch* some of those animals, very few can be said to *observe* their behavior. What distinguishes observation from mere watching is the use of careful, methodical plans for "the selection, provocation, recording, and encoding of

that set of behaviors and settings concerning relevant organisms 'in situ' which is consistent with empirical aims" (Weick 1968:360). What distinguishes the various overlapping and intergrading observational methods from each other is the character of these plans.

Observation is systematic to the degree that plans for selection, provocation, recording, and encoding are both explicit and preset, rather than either implicit or emergent. Observation qualifies more fully as *field* observation the more completely each of these plans is executed within a field setting. Observation is more naturalistic, the less extensive and less potent the plan for provoking behavior.

The bounds of this review of systematic field observation are thus necessarily relative and imprecise. Most clearly excluded are "participant observation" in the classical ethnographic style (Emerson 1981) and observational studies conducted in laboratory or other contrived settings (Weick 1968), even though both share numerous problems and techniques with systematic field observation. Studies in which field-recorded data are primarily postcoded in the researcher's office or laboratory are only covered peripherally, on the grounds that using precoded instruments favors greater explicitness in observation and is more directly and completely a field operation. Accordingly, I do not deal here in any detail with research that relies on the postcoding of either narrative field notes (Barker & Wright 1955, Yarrow & Waxler 1979, Melbin 1972) or tape recordings and films (Kendon 1979, West & Zimmerman 1982).

The focus of this review, then, is naturalistic field observation in which the plans for selection, provocation, recording, and encoding are explicit, preset, and executed more or less contemporaneously.

SUBSTANTIVE APPLICATIONS

Although systematic field observation is advantageously employed to train practitioners (Simon & Boyer 1974) and to validate other methods (Haynes 1978), I will only review its use in primary data collection. The empirical literature based on systematic field observation is quite extensive, encompassing a variety of research purposes and a considerable range of subject matters.

Functional Applications

In conventional academic research, investigators strive to develop a coherent body of knowledge for its own sake. Systematic field observation may be used (Kelly 1967) to establish empirical regularities, to anchor the development of theories, to test specific hypotheses, and to evaluate and revise theories. Our richly theoretical knowledge of children's acquisition of language, for example, owes much of its current shape to naturalistic field observation (Dale 1978).

Yet, like survey research, systematic field observation is more often a tool of applied than of academic research. To sociologists, applied research usually means policy research. Excepting the British mass-observation program (Madge & Harrison 1939), systematic observational studies to assess large-scale social problems have been uncommon. Reiss's (1968a, 1971b) study of police brutality and injustice stands as a modern example of what can be learned about the incidence, nature, and causes of a social problem by using this method. More common policy research based on systematic field observation includes studies evaluating the implementation and effectiveness of preventive or remedial programs (Kelling et al 1974, Fienberg et al 1976, Paul 1984, Alevisos et al 1978, Turner et al 1983).

The predominant use of this method in applied research, however, has been in "action research" in the clinical and practice fields. Systematic field observation is used there in (*a*) the determination of normative behavioral standards, such as developmental norms for child behavior or work efficiency norms (established through time and motion studies); (*b*) diagnostic assessment of individual or group performance to determine what, if any, intervention may be warranted; and (*c*) posttreatment assessment to evaluate need for further intervention. During the third wave behavioral assessment through systematic field observation became a vital and highly developed technology, particularly within the clinical fields (Jones et al 1975, Kent & Foster 1977, Wildman & Erickson 1977, Cone & Foster 1982, Hartmann & Wood 1982).

Topical Applications

Direct observation can be particularly useful in the study of (*a*) verbally deficient or unsophisticated individuals, such as young children; (*b*) behaviors that most people cannot accurately describe, such as interaction rates or fleeting expressive movements; and (*c*) events that subjects may be motivated to distort or omit in reporting. Under favorable circumstances, however, virtually all natural social phenomena are amenable to systematic field observation (Reiss 1971a). The range of phenomena examined through this method did indeed widen appreciably during the third wave, although the traditional topics of child behavior and organizational role performance still dominate the literature.

Researchers studying child behavior pioneered and continue to champion the development of systematic field observation (Hutt & Hutt 1970, Cairns 1979). Much of their observational work has focused on developmental patterns and child-rearing practices, and it is these studies that have been most affected by the innovative concepts and techniques of human ethology (McGrew 1972, Blurton Jones 1972). Not surprisingly, then, the ethological approach has also been extended to the study of developmental disorders (Sackett 1978b). Chil-

dren's social behavior has also been a major locus of development in the observational techniques of behavioral assessment and of classroom behavior research (Medley & Mitzel 1963, Chanan & Delamont 1975, Stubbs & Delamont 1976).

The second classical topic of systematic field observation has been organizational role performance—principally by classroom teachers (Flanders 1970, Rosenshine & Furst 1973), industrial workers (Barnes 1980), and managers (Chapple & Sayles 1961). The extension of this line of observational study to other institutional sectors has been one of the most productive innovations to emerge during the third wave. Sociologists in particular opened the criminal justice system to systematic observational study (McCall 1978), examining the role performances of lawyers, judges (Mileski 1971), and especially the police (reviewed in Reiss 1979, Black 1980, Sykes & Brent 1983). The medical system has also been explored in considerable detail, through systematic observational study of the role performance of health-care workers in a variety of institutional settings (Paul 1984, Alevisos et al 1978, Brookshire et al 1978).

Apart from these two primary topics, a considerable array of social phenomena have been considered. Social deviance has been a significant observational interest among both clinical researchers and sociologists, principally in the areas of predelinquent acts (Jones et al 1975), drinking behavior (Plant et al 1977, Kessler & Gomberg 1974, Wiseman 1970), gambling (Blasiovich et al 1976), sexual deviance (Humphreys 1970), and overeating (Stunkard & Kaplan 1977). Family interaction, too, has been subject for observation among both groups (Riskin & Faunce 1972), although the studies often focus more narrowly on marital (Filsinger & Lewis 1981) or parent-child interaction (Lytton 1971).

Social interaction may also be viewed more abstractly, as in systematic field observation of (*a*) group dynamics (Conger & McLeod 1977, Omark et al 1980); (*b*) collective behavior (McPhail 1972, Hare 1980); (*c*) communication, both verbal (Davis 1978) and nonverbal (Scherer & Ekman 1982a); and (*d*) interaction process (Cairns 1979, Duncan & Fiske 1977, Lamb et al 1979). In fact, the proliferation of complex theoretical frameworks for analyzing interaction process—all owing much to Erving Goffman's seminal ideas—has been a major stimulus of renewed interest in systematic observation. Some of these new frameworks focus on the role of those shared communication codes that make the intricate coordination of lines of action possible, while others stress the processes through which such codes are employed to produce coordinated action. The varieties of new frameworks—and their sources in microsociology, ethnomethodology, sociolinguistics, cognitive anthropology, structural linguistics, kinesics, and cybernetics—are reviewed more systematically by Kendon (1982) and McDermott & Roth (1978).

OBSERVATIONAL PARADIGMS

From a methodological perspective, however, there are only three fundamental conceptualizations of interaction (Kendon 1982); a distinctive paradigm for systematic observation has evolved around each of them (Weick 1968, Medley & Mitzel 1963).

The Behavior Variables Approach

In this approach (Kendon 1982), one selects in advance what appear to be—on practical or theoretical grounds—the important parameters of behavior. One then sets out to measure and perhaps to interrelate them with each other or with other individual or social variables. This approach is especially characteristic of applied research, where interest may be largely confined to, say, how often a child insults his parents or how much of the day a worker is idle. Scores or profiles are established for individuals or groups, typically through use of a *sign-code system* of observation (Medley & Mitzel 1963) in which the observer watches for and notes the occurrence of select, prelisted events. Records most often take the form of a checklist or tally sheet (Hinde 1973) where the observer notes which of these events did occur, perhaps how frequently, and, less commonly, with what duration. The observation period is often divided into smaller time segments, permitting the use of time-sampling techniques (Thomson et al 1974).

The Behavior Element Approach

Alternatively, social interaction can be regarded as made up of discrete elements, i.e. a finite set of "repeatable patterns of behavior that are, as it were, the building blocks whose concatenations and clusterings create the more elaborate sequences and structures" (Kendon 1982:474–75). This approach, modeled on the chemical table of elements, is particularly congenial to the ethologist's notion of genetically and physiologically "fixed action patterns," but it is also quite familiar to sociologists in the form of Bales's (1950) system of interpersonal acts. According to this approach, interaction is most appropriately studied by means of *category systems* of observation (Medley & Mitzel 1963). Given the table of elements, the observer attends sequentially to every act, noting its classification and, perhaps, its time of onset and offset. Observers typically employ portable electronic devices to record, clock, and store digitally encoded data (Sidowski 1977, Sykes & Whitney 1969). The objective is to discern the synchronic and diachronic structure of the stream of behavior by statistically analyzing these data (van Hooff 1982). Synchronic analysis clusters either behaviors or actors through either factor or hierarchical cluster analysis (Gottman 1978, Fagen & Young 1978). Diachronic or sequential analysis (Bakeman 1978, Gottman & Bakeman 1979) initially relied on Mar-

kov chain techniques, exemplified in the sociological work of Brent & Sykes (1979, Sykes & Brent 1983). Sackett (1979) and Oden (1977) describe more powerful techniques for sequential and time series analysis that are now available.

The Constructivist Approach

Finally, interaction can be viewed as a social accomplishment, with an emphasis on analyzing how the actions of multiple participants are jointly organized to produce or to manage some socially recognizable episode or event (Kendon 1982). This approach has selective affinity with *field-format systems* of observation (Weick 1968), which promote comprehensiveness of description through ensuring that the observer systematically attends to various aspects of a natural event. An innovative example of such an observation system is the observer-questionnaire format that Black & Reiss (1967) developed for their study of contingencies in the unfolding of police-citizen interactions. This format [partially reprinted in McCall (1978) and refined by Tifft (1970)] requires the observer to attend to different critical features at each phase of the encounter. Because advocates of the constructivist approach are interested in discerning principles of organization, they are increasingly using systems diagramming techniques (Sustare 1978)—such as network analysis—and even mathematical grammars (Westman 1977, van Hooff 1982) in analyzing observational data.

OBSERVATION SYSTEMS

Observation systems should not be confused with formats for encoding and recording data, such as the off-the-shelf instruments Simon & Boyer (1974) collected. Rather, an observation system is

> a more-or-less formalized set of rules for extracting information from the stream of behavior. These rules specify the target events or behaviors, the observation settings, and the observers; they also specify how the events are sampled, the dimensions of the events that are assessed, and how the data are recorded. Furthermore, observation rules may specify how the resulting data are combined to form scores. Thus, the unique set of rules defining a specific observation system also determines the cost, the detail, and the generality of the resulting information, as well as the questions to which the information is relevant (Hartmann & Wood 1982:110).

Hawkins (1982), Reed (1978), and Blurton Jones (1972) discuss the art of evolving such a set of rules, i.e. of developing an observation system. Herbert & Attridge (1975) provide a detailed specification of the descriptive, psychometric, and practical information necessary to document an observation system. Here I only review selected issues surrounding some general alternatives in constructing an observation system.

Plans for Selection

Observation is always selective and purposive and involves decisions about what should be noticed and what should be ignored. One must decide both which episodes of interaction will be observed and how observer attention will be allocated within a given episode.

In view of the rigorous theory of generalizability, sampling theory for the selection of episodes remains surprisingly underdeveloped (Mitchell 1979, Linehan 1980). Generalizability theory, however, has at least shown researchers that in selecting episodes for study they should simultaneously take account of various facets of interaction—i.e. actors, actions, settings, and occasions. How these facets are considered in selecting episodes (Longabaugh 1980) and how many observations of what length are needed depend on the kinds of generalizations one wishes to make. Possible subjects for generalization include:

1. The behavior profiles of specific individuals (Cronbach et al 1972): The facets of the actors and actions to be observed are more or less fixed by the terms of the specific research question. Sampling frames for the setting and occasion facets (Wahler et al 1976) can be constructed after reconnaissance observation (Lehner 1979) to identify where and when the behaviors of interest are likely to occur.
2. The behavior repertoire of a specific class of individuals (Fagen 1978, Melbin 1972): None of the four facets is fixed, and a comparatively large number of observations is needed to ensure that all potential behavior types are seen.
3. Some specific interaction form (Scherer & Ekman 1982b): Lists of social episodes of any type are seldom found. Therefore, to construct the sampling frames one must resort to a variety of multistage sampling procedures (Reiss 1971a). The optimal sample size depends on the estimated variability of the interaction form (Barnes 1980).

Most of the sampling literature, however, takes episodes as given and deals only with the allocation of observer attention within an episode. Target features may include interactive units (Patterson & Moore 1979), events, environmental states, or even compound ecobehavioral units (Barker & Wright 1955), but most often are the behaviors of individuals. The observer is usually asked to note one of the following features of a behavior (Cone & Foster 1982): (*a*) its frequency, or a derivative such as its rate, relative frequency, or conditional probability; (*b*) its duration, or a derivative such as the interval or latency; or (*c*) its quality, e.g. its intensity, accuracy, or acceptability. Ideally, an observation system specifies which dimensions of which actions by which actors shall be noted at which times. Such a plan need not be uniform across actions, actors, and times within an episode; it may be differential, e.g. the selection of target

behaviors may vary for different types of actors or for different time periods for the same target actor.

The most ambitious strategy, advocated mainly by proponents of the behavior element paradigm, calls for describing every defined behavior of each actor over the entire course of the observation period (Sanson-Fisher et al 1979). Most within-episode selection strategies are less comprehensive and pivot on one of four facets of interaction: (*a*) actors, as in focal-animal and sociometric matrix sampling; (*b*) actions, as in event sampling and scan sampling; (*c*) settings, as in spatial grid sampling (Sykes et al 1976); or (*d*) occasions, as in time-sampling (Thomson et al 1974). The objective of time-sampling is to make observation discontinuous—conducted only at preselected instants or during preselected intervals, often with rotation or randomization of target actors between successive time samples.

The applicability of any of these within-episode strategies (Altmann 1974, Slater 1978, Sackett 1978a) depends not only on the nature of the episode—i.e. the number of actors, the rate and duration of behaviors, and the structure of the setting—but also on the purpose of the investigation, since each strategy tends to yield a somewhat different estimate of frequency and duration parameters (Dunbar 1976, Powell et al 1977, Repp et al 1976, Murphy & Goodall 1980).

Plans for Encoding

Selective attention does not automatically result in selective noticing; recognition of perceptual objects requires conceptual representation of received information—that is, it must be encoded within some semantic structure. Such a code system or observational language is the most visible component of any observation system. Major differences in the form and features of observational languages—e.g. sign-code type, category type, field-format type—reflect the interests and purposes of the investigators. For example, the advantages of an exhaustive and mutually exclusive set of codes for analyzing behavior structures are clearcut, whereas such logical closure may be quite irrelevant for assessing whether a patient is now smoking less often. Thus, many of the issues concerning code systems essentially represent irresolvable paradigm clashes.

An issue that affects all researchers, regardless of the paradigm they employ, is choosing the level of analysis, i.e. of using molecular vs molar units of observation (Hollenbeck 1978); analysis at multiple levels has also been increasingly urged (Lamb et al 1979). The criteria to be used and the techniques for aggregating molecular units into molar units have been discussed extensively (Cone & Foster 1982, van Hooff 1982), but comparatively little attention has been paid to the logical problems of aggregation and disaggregation long familiar to sociologists (Hannan 1971). Hierarchical coding (Sykes 1977) and multiple-level code systems (Bales & Cohen 1979, Lamb et al 1979) provide promising alternatives to statistical aggregation.

Whether a code system should be extensive, covering many features of interaction, or intensive, focusing more narrowly on only a few, depends not only on the purpose of observation but also on considerations of observer capacity. How many distinctions can an observer be expected to make in any finite time? The complexity of a code system has a significant influence on the extent of observational error (Mash & McElwee 1974, Jones et al 1975, Frame 1979, House & House 1979). Yet Sykes (1977) and Stephenson (1979) have shown how careful design of the syntactical structure of observational languages permits very complex coding if one takes advantage of the redundancy and hierarchical organization inherent in any useful grammar. [The Sykes code is reprinted in McCall 1978.]

Plans for Recording

Encoded information becomes observational data only when it is recorded. Debates over how and when precoded recording should occur are also, in good part, paradigm clashes.

The use of portable electronic digital recording devices is strongly favored by advocates of the behavior element paradigm because of their need for comprehensive recording of large numbers of very brief but frequently occurring events. Although these electronic keyboard devices have been engineered to minimize mechanical and human errors in their use, they do impose some practical constraints (Sykes 1977, Simpson 1979) that seem unwarrantable to adherents of other paradigms. Moreover, visible recording devices may increase the reactivity of measurement (McPhail et al 1972).

Similarly, the needs of the behavior element paradigm render contemporaneous recording on the spot virtually mandatory, even though overt recording—by whatever technological means—may be more reactive than delayed recording out of view. Some constructivist systems, on the other hand, may at times require some retrospective use of information in order for an observer to discern fully what had been occurring earlier. Contemporaneous recording thus would be unworkable, but retrospective recording—even while in the field—does raise the possibility of introducing memory distortions (Weick 1968).

THE HUMAN FACTOR: ERROR, RELIABILITY, AND VALIDITY

From a more sociological perspective, the plans and rules that comprise an observation system can be viewed alternatively as role expectations, governing not only the job of the observer but also his role relationships with the persons being observed (Reiss 1971a). Developing an observation system, then, amounts to a process of role definition or job design. Some of these role expectations may themselves be foolish or misguided, so that observer com-

pliance with them eventuates in misleading data. Errors of observation encompass not only "system errors" of this type but also "observer errors," here redefined as the violation or compromise of role expectations.

Observer Errors

Reiss (1976) and Fassnacht (1982) examine both system and observer errors within the comprehensive framework of error sources that Webb et al (1966) developed. Here I will only review empirical investigations of observer cheating, bias, drift, and reactive effects. As other reviewers have noted (Kent & Foster 1977, Wildman & Erickson 1977, Cone & Foster 1982, Hartmann & Wood 1982), these studies' mixed findings suggest that such observer errors (*a*) influence data not uniformly but rather in a scattered and unpredictable fashion and (*b*) are not harmful unless they go undetected and unmeasured.

ERRORS IN JOB PERFORMANCE

Observer cheating (Kent & Foster 1977) has been considered less often than interviewer cheating and is less readily detected, though perhaps more easily prevented.

Observer bias, or systematic error stemming from the cognitive operations of the observer himself, can be monitored statistically (House 1980). Unless proper methodological remedies (Weick 1968) are employed, the information-processing limitations of the human observer do tend to impose regularity and orderliness on otherwise unruly data (Hollenbeck 1978, Mash & Makohoniuk 1975). Observer prejudices (Reiss 1971a) and expectancies (Kent & Foster 1977) about subjects, hypotheses, and data may also systematically distort observational data, although the most stringently designed investigations (O'Leary et al 1975, Redfield & Paul 1976) suggest that such effects may only be appreciable when expectancies are reinforced by the principal investigator.

Observer drift, or measurement decay in observer performance, is a well documented but manageable threat to standardized data collection (Taplin & Reid 1973, Romanczyk et al 1973, Kent et al 1974). Particularly troublesome is "consensual observer drift," where interobserver consistency remains high yet observer accuracy declines over time (Johnson & Bolstad 1973).

ERRORS STEMMING FROM ROLE RELATIONS

Reactive effects, or changes in the behavior of the observed as a consequence of the observer's presence, are inherently difficult to document. Several detection strategies have been employed (Johnson & Bolstad 1973, Baum et al 1979), and it appears that reactivity is a common but far from universal effect of observation (Haynes & Horn 1982, Kazdin 1982). Sykes (1978) suggests that reactivity is a function of the social context of observation and of the situational norms governing the social relationship between observer and observed. This relationship is centrally defined by the purpose and rationale of observation, as for instance, in some

type of evaluation effort (Hagen et al 1975). Other factors shaping the relationship, such as relative social positions, the degree of previous acquaintance, and demographic characteristics, may also condition the occurrence of reactivity (Baum et al 1979, Hartmann & Wood 1982). The inconsistent findings on the effects of the degree of intrusiveness or conspicuousness of observation (Baum et al 1979) may indicate that intrusiveness can only be defined in terms of the situational norms governing the observer-observed relationship. Ultimately, of course, reactivity depends on the social sensitivity of those being observed: infants, the mentally disabled, the intoxicated, the very confident and open, and those wholly engrossed in role performance may be less responsive to the presence of observers (Baum et al 1979, Sykes 1978).

Although many studies demonstrate reactive effects, only a small proportion of the behaviors recorded in these investigations were actually affected (Kent & Foster 1977). If social desirability is the motivating factor in reactivity (Kazdin 1982), modifications of conduct should be limited to those aspects that subjects assume may be under evaluation. Moreover, Reiss (1971a) shows that social desirability concerns often cannot override the intrinsic constraints of situated action. Finally, the persistence of any reactive effects of observation is widely questioned. Habituation to observation is a well documented phenomenon (Johnson & Bolstad 1973, Baum et al 1979), so that reactive effects may mainly threaten data obtained early in an investigation. Guidelines for the length of time or the number of observation sessions required for habituation remain empirically questionable, however (Hartmann & Wood 1982).

CONTROL OF OBSERVER ERRORS Viewing an observation system in terms of role expectations affords two advantages in managing observer error. First, such a view encourages more serious use (Reiss 1968b, 1971a) of the extensive lore on field roles and field relations (McCall 1978, Emerson 1981) that has developed within participant observation research. Second, reinterpreting observer error as the violation or compromise of role expectations suggests that the sources of observer error are to be found in the social organization of the research project. They include:

1. Faulty job design: Insufficient or inappropriate specification of the observer role lays the ground for observer error by augmenting the role strain, ambiguity, and conflict (Secord & Backman 1974) found in all social roles. Developers of observation systems should pay close heed to the principles of empirical job design (Porter et al 1975).

2. Selection, training, and supervision of observers: Given the historic association of systematic field observation with applied psychology, one might suppose that these arts would be highly developed. In fact, however, observer training (Spool 1978, Hartmann & Wood 1982) has not changed significantly

in the last 50 years. In addition, the selection of observers has only rarely involved any direct testing of aptitudes and skills (Hartmann & Wood 1982)—despite clear demonstrations of individual differences in observational skill (Mazanec & McCall 1975, Smith 1983). Supervision and continuing socialization (Reiss 1971a) of observers, in contrast, has become considerably more sophisticated (Johnson & Bolstad 1973, Taplin & Reid 1973, Romanczyk et al 1973, Kent et al 1974) as researchers attempt to control observer drift, bias, and cheating.

3. Influences of informal organization: Both role theory (Secord & Backman 1974) and organizational research (Porter et al 1975) suggest that role bargaining and some patterned evasion of norms by observers can be expected, due to the constant tug of informal social pressures from colleagues and from the observed. Job design and training must anticipate the shape of such pressures, and supervision must make provision for their containment.

Reliability

A notable third-wave trend has been the marked shift from a reliance on measures of interobserver agreement toward the use of the traditional psychometric concept of reliability that emphasizes the accuracy and stability of measurement. This change has occurred for a variety of reasons. First, although a plethora of sophisticated agreement statistics has been developed (Berk 1979), each ordinarily yields a different result when applied to the same data (Frick & Semmel 1978, Hartmann 1977). Second, interobserver agreement on sequential features of the data has proved extraordinarily difficult to conceptualize and measure (Hollenbeck 1978, Wampold & Holloway 1983). Third, it has become apparent that elements other than differences among observers contribute to disagreement (Yarrow & Waxler 1979). Fourth, it has also been increasingly recognized that agreement, accuracy, and stability can vary independently (Medley & Mitzel 1963, Johnson & Bolstad 1973, DeMaster et al 1977).

The key development in assessing the reliability of observational data has been the theory of generalizability, propounded by Medley & Mitzel (1963) and refined by Cronbach et al (1972). [Subsequent developments are reviewed by Shavelson & Webb (1981).] Within this explicitly psychometric framework, the components of an observational score are considered to be determined by the specific conditions under which it was obtained, including the occasion, setting, scoring system, and observer. Using analysis of variance procedures, the contributions of these conditions or facets are identified and summarized by means of variance components, intraclass correlation coefficients, and measurement error statistics.

The advantages of this approach are numerous (Berk 1979, Mitchell 1979).

For example, large variance components associated with a facet or an interaction term among facets indicate limited generalizability along those dimensions. Intraclass correlations can be used to estimate the number of additional occasions, settings, or observers needed to achieve satisfactory generalizability. Still, with comparatively few exceptions (Alevisos et al 1978), generalizability theory has had more of a conceptual than a technical impact (Cone 1977, Jones 1977, Coates & Thoresen 1978, Wieder & Weiss 1980). The statistical requirements of generalizability theory are not wholly compatible with the logic of the single-subject designs so important to clinical researchers (Jones 1977, Strossen et al 1979), and observational data frequently violate the psychometric assumptions of concurrent and serial independence of observations (Hollenbeck 1978, Shavelson & Webb 1981).

Validity

Validity, or the extent to which a score measures what it is intended to measure, can be addressed directly within the framework of generalizability theory (Cone 1977). For the most part, however, discussions of observational validity have relied on more traditional conceptions. Since observation is always conducted for a purpose, in the end the question of validity comes down to whether or not the observational data are adequate to achieve it. Observations with impressive validity for one purpose may be only moderately valid or even invalid for others (Cone 1982).

Content validity, or the adequacy with which an instrument samples the behavioral domain of interest, is important in the initial development of any behavior-coding schema (Linehan 1980), but it is especially critical when the behavior element paradigm is used. Criterion-related validity—i.e. the usefulness or accuracy of observation scores in predicting some performance criterion (predictive validity) or in substituting for some other established measure (concurrent validity)—is particularly important in the action research (Paul 1984, Kazdin 1979) to which the behavior variables paradigm is so commonly applied.

Of greatest interest, however, is construct validity, i.e. the degree to which observations accurately measure some theoretical concept. Extensive and imaginative efforts to establish the construct validity of observational measures are exemplified in Johnson & Bolstad's (1973) and Jones et al's (1975) work. The utility of multitrait-multimethod techniques of convergent and discriminant validation, aimed at removing the distortions produced by method variance, has not been overlooked by systematic field observation researchers (Blunden et al 1974, Borich et al 1978, Weinrott et al 1981). Such validity investigations may warrant reanalysis, however, in light of the radically new view that methods dilute trait relationships rather than add irrelevant systematic variance (Campbell & O'Connell 1982).

CONCLUSION

Due in part to sociologists' direct and indirect contributions, systematic field observation is a sophisticated and powerful tool of social research, better suited than ever before to the study of social interaction and social organization. Compared to interviews and questionnaires (Reiss 1976), this method provides: (*a*) more reliable information about events; (*b*) greater precision regarding their timing, duration, and frequency; (*c*) greater accuracy in the time ordering of variables; and (*d*) more accurate and economical reconstructions of large-scale social episodes. Still, most sociologists turn to observation only when verbal methods fail—that is, when respondents cannot or will not report accurately. Even then, observation is more likely to take the form of ethnographic field work. There is a strong tradition of participant observation research within sociology that has a special affinity to the study of social organization, lends itself to solo research, and importantly, demands a close interplay between direct observation and verbal methods.

No comparable sociological tradition of systematic field observation exists yet. Sociologists have been important but numerically minor contributors, and not many read widely in the methodological literature of applied psychology. Textbook coverage of systematic field observation is therefore weak at best. It certainly takes no account of either the constructivist paradigm so relevant to studying social organization or those multimethod designs that conjoin systematic field observation with interview surveys (Tifft 1970, McCall 1978) in order to explore the fit between words and deeds. Institutionalization of the recent sociological interest in systematic field observation therefore depends on timely development of better instructional materials and training programs. Several developments suggest that it is not yet too late:

1. The very appearance of a chapter such as this one may in itself be taken to indicate continuing interest—or at least curiosity.
2. The comforting fact that systematic field observation is essentially a species of survey research design is now more widely appreciated among sociologists.
3. Psychometric advances within systematic field observation are highly congenial to the sophisticated measurement concerns inherent in the advanced causal modeling techniques so popular among sociologists today.
4. Sociological interest in the study of interaction processes remains high, so that the recently developed constructivist paradigm of systematic field observation may yet sink lasting roots.
5. Interest in policy research continues to grow, despite funding cutbacks, so that more and more applied sociologists may find themselves attracted to systematic field observation.

Literature Cited

Alevisos, P. N., DeRisi, W., Liberman, R., Eckman, T., Callahan, E. 1978. The Behavior Observation Instrument: A method of direct observation for program evaluation. *J. Appl. Behav. Anal.* 11:243–57

Altmann, J. 1974. Observational study of behavior: Sampling methods. *Behaviour* 49: 227–67

Bakeman, R. 1978. Untangling streams of behavior: Sequential analysis of observational data. See Sackett 1978c, pp. 63–78

Bales, R. F. 1950. *Interaction Process Analysis*. Reading, Mass: Addison-Wesley

Bales, R. F., Cohen, S. P. 1979. *SYMLOG: A System for the Multiple Level Observation of Groups*. New York: Free

Barker, R. G., Wright, H. F. 1955. *Midwest and Its Children: The Psychological Ecology of an American Town*. New York: Harper & Row

Barnes, R. M. 1980. *Motion and Time Study: Design and Measurement of Work*. New York: Wiley. 7th ed.

Baum, C. G., Forehand, R., Zegiob, L. E. 1979. A review of observer reactivity in adult-child interactions. *J. Behav. Assess.* 1:167–78

Berk, R. A. 1979. Generalizability of behavioral observations: A clarification of interobserver agreement and interobserver reliability. *Am. J. Ment. Defic.* 83:460–72

Black, D. 1980. *The Manners and Customs of the Police*. New York: Academic

Black, D., Reiss, A. J. Jr. 1967. Patterns of behavior in police and citizen transactions. In *Studies in Crime and Law Enforcement in Major Metropolitan Areas. Field Surveys III*, US President's Commission on Law Enforcement and Administration of Justice, 2:1–139. Washington DC: USGPO

Blasiovich, J., Ginsburg, G. P., Howe, R. C. 1976. Blackjack: Choice shifts in the field. *Sociometry* 39:274–76

Blunden, D., Spring, C., Greenberg, L. M. 1974. Validation of the classroom behavior inventory. *J. Consult. Clin. Psychol.* 42:84–88

Blurton Jones, N., ed. 1972. *Ethological Studies of Child Behavior*. Cambridge: Cambridge Univ. Press

Borich, G. D., Malitz, D., Kugle, C. L. 1978. Convergent and discriminant validity of five classroom observation systems: Testing a model. *J. Educ. Psychol.* 70:119–28

Brent, E. E. Jr., Sykes, R. E. 1979. A mathematical model of symbolic interaction between police and suspects. *Behav. Sci.* 24:388–402

Brookshire, R. H., Nicholas, L. S., Krueger, K. M., Redmond, K. J. 1978. The clinical interaction analysis system: A system for observational recording of aphasia treatment. *J. Speech Hear. Disord.* 43:437–47

Cairns, R. B., ed. 1979. *The Analysis of Social Interactions: Methods, Issues, and Illustrations*. Hillsdale, NJ: Erlbaum

Campbell, D. T., O'Connell, E. J. 1982. Methods as diluting trait relationships rather than adding irrelevant systematic variance. In *Forms of Validity in Research*, ed. D. Brinberg, L. H. Kidder, pp. 93–111. San Francisco: Jossey-Bass

Chanan, G., Delamont, S., eds. 1975. *Frontiers of Classroom Research*. Windsor, England: Natl. Found. Educ. Res.

Chapple, E. D., Sayles, L. R. 1961. *The Measure of Management*. New York: Macmillan

Coates, T. J., Thoresen, C. E. 1978. Using generalizability theory in behavioral observation. *Behav. Ther.* 9:605–13

Colgan, P. W., ed. 1978. *Quantitative Ethology*. New York: Wiley

Cone, J. D. 1977. The relevance of reliability and validity for behavior assessment. *Behav. Ther.* 8:411–26

Cone, J. D. 1982. Validity of direct observation assessment procedures. See Hartmann 1982, pp. 67–79

Cone, J. D., Foster, S. L. 1982. Direct observation in clinical psychology. In *Handbook of Research Methods in Clinical Psychology*, ed. P. C. Kendall, J. N. Butcher, pp. 311–54. New York: Wiley

Cone, J. D., Hawkins, R. P., eds. 1977. *Behavioral Assessment: New Directions in Clinical Psychology*. New York: Brunner/Mazel

Conger, R. D., McLeod, D. 1977. Describing behavior in small groups with the Datamyte event recorder. *Behav. Res. Methods Instrum.* 9:418–24

Cronbach, L. J., Gleser, G. C., Nanda, H., Rajaratnam, N. 1972. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley

Dale, P. S. 1978. What does observing language mean? See Sackett 1978b, pp. 219–37

Davis, K. 1978. Methods for studying informal communication. *J. Commun.* 28:112–16

DeMaster, B., Reid, J., Twentyman, C. 1977. The effects of different amounts of feedback on observer's reliability. *Behav. Ther.* 8:317–29

Dunbar, R. I. M. 1976. Some aspects of research design and their implications in the observational study of behaviour. *Behaviour* 58:78–98

Duncan, S. Jr., Fiske, D. W. 1977. *Face-to-*

Face Interaction: Research, Methods, and Theory. Hillsdale, NJ: Erlbaum
Emerson, R. M. 1981. Observational field work. *Ann. Rev. Sociol.* 7:351–78
Fagen, R. M. 1978. Repertoire analysis. See Colgan 1978, pp. 25–42
Fagen, R. M., Young, D. Y. 1978. Temporal patterns of behavior: Durations, intervals, latencies, and sequences. See Colgan 1978, pp. 79–114
Fassnacht, G. 1982. *Theory and Practice of Observing Behavior*. New York: Academic
Fienberg, S. E., Larntz, K., Reiss, A. J. Jr. 1976. Redesigning the Kansas City preventive patrol experiment. *Evaluation* 3:124–31
Filsinger, E. E., Lewis, R. A., eds. 1981. *Assessing Marriage: New Behavioral Approaches*. Beverly Hills, Calif: Sage
Flanders, N. A. 1970. *Analyzing Teaching Behavior*. Reading, Mass: Addison-Wesley
Frame, R. 1979. Interobserver agreement as a function of the number of behaviors recorded simultaneously. *Psychol. Rec.* 29:287–96
Frick, T., Semmel, M. I. 1978. Observer agreement and reliabilities of classroom observational measures. *Rev. Educ. Res.* 48:157–84
Gottman, J. M. 1978. Nonsequential data analysis techniques in observational research. See Sackett 1978c, pp. 45–61
Gottman, J. M., Bakeman, R. 1979. The sequential analysis of observational data. See Lamb et al 1979, pp. 185–206
Hagen, R. L., Craighead, W. E., Paul, G. L. 1975. Staff reactivity to evaluative behavioral observations. *Behav. Ther.* 6:201–5
Hannan, M. T. 1971. *Aggregation and Disaggregation in Sociology*. Lexington, Mass: Lexington
Hare, A. P. 1980. A dramaturgical analysis of street demonstrations. *Group Psychother. Psychodr. Sociom.* 33:99–120
Hartmann, D. P. 1977. Considerations in the choice of interobserver reliability estimates. *J. Appl. Behav. Anal.* 10:103–16
Hartmann, D. P., ed. 1982. *Using Observers to Study Behavior*. San Francisco: Jossey-Bass
Hartmann, D. P., Wood, D. D. 1982. Observational methods. In *International Handbook of Behavior Modification and Therapy*, ed. A. S. Bellack, M. Hersen, A. E. Kazdin, pp. 109–38. New York: Plenum
Hawkins, R. P. 1979. The functions of assessment: Implications for selection and development of devices for assessing repertoires in clinical, educational, and other settings. *J. Behav. Assess.* 1:501–16
Hawkins, R. P. 1982. Developing a behavior code. See Hartmann 1982, pp. 21–35
Haynes, S. N. 1978. *Principles of Behavioral Assessment*. New York: Gardner
Haynes, S. N., Horn, W. F. 1982. Reactivity in behavioral observations: A methodological and conceptual critique. *Behav. Assess.* 4: 369–85
Hazlett, B. A., ed. 1977. *Quantitative Methods in the Study of Animal Behavior*. New York: Academic
Herbert, J., Attridge, C. 1975. A guide for developers and users of observation systems and manuals. *Am. Educ. Res. J.* 12:1–20
Hinde, R. A. 1973. On the design of check-sheets. *Primates* 14:393–406
Hollenbeck, A. R. 1978. Problems of reliability in observational research. See Sackett 1978c, pp. 79–98
House, A. E. 1980. Detecting bias in observational data. *Behav. Assess.* 2:29–31
House, B. J., House, A. E. 1979. Frequency, complexity, and clarity as covariates of observer reliability. *J. Behav. Assess.* 1: 149–66
Humphreys, L. 1970. *Tearoom Trade: Impersonal Sex in Public Places*. Chicago: Aldine
Hutt, S. J., Hutt, C. 1970. *Direct Observation and Measurement of Behavior*. Springfield, Ill: Thomas
Johnson, S. M., Bolstad, O. D. 1973. Methodological issues in naturalistic observation: Some problems and solutions for field research. In *Behavior Change: Methodology, Concepts, and Practice*, ed. L. A. Hamerlynck, L. C. Handy, E. J. Mash, pp. 7–67. Champaign, Ill: Research
Jones, R. R. 1977. Conceptual vs. analytic uses of generalizability theory in behavioral assessment. See Cone & Hawkins 1977, pp. 330–43
Jones, R. R., Reid, J. B., Patterson, G. R. 1975. Naturalistic observation in clinical assessment. In *Advances in Psychological Assessment*, ed. P. McReynolds, 3:42–95. San Francisco: Jossey-Bass
Kazdin, A. E. 1979. Situational specificity: The two-edged sword of behavioral assessment. *Behav. Assess.* 1:57–75
Kazdin, A. E. 1982. Observer effects: Reactivity of direct observation. See Hartmann 1982, pp. 5–19
Kelling, G., Pate, T., Dieckman, D., Brown, C. E. 1974. *The Kansas City Preventive Patrol Experiment: A Technical Report*. Washington DC: Police Found.
Kelly, J. G. 1967. Naturalistic observations and theory confirmation: An example. *Hum. Dev.* 10:212–22
Kendon, A. 1979. Some methodological and theoretical aspects of the use of film in the study of social interaction. In *Emerging Strategies in Social Psychological Research*, ed. G. P. Ginsburg, pp. 67–91. New York: Wiley
Kendon, A. 1982. The organization of behavior in face-to-face interaction: Observa-

tions on the development of a methodology. See Scherer & Ekman 1982a, pp. 440–505
Kent, R. N., Foster, S. L. 1977. Direct observational procedures: Methodological issues in naturalistic settings. In *Handbook of Behavioral Assessment,* ed. A. R. Ciminero, K. S. Calhoun, H. E. Adams, pp. 279–328. New York: Wiley
Kent, R. N., O'Leary, K. D., Diament, C., Dietz, A. 1974. Expectation biases in observational evaluation of therapeutic change. *J. Consult. Clin. Psychol.* 42:774–80
Kessler, M., Gomberg, C. 1974. Observations of barroom drinking: Methodology and preliminary results. *Q. J. Stud. Alcohol* 35: 1392–96
Lamb, M. E., Suomi, S. J., Stephenson, G. R., eds. 1979. *Social Interaction Analysis: Methodological Issues:* Madison: Univ. Wis. Press
Lehner, P. N. 1979. *Handbook of Ethological Methods.* New York: Garland STPM
Linehan, M. M. 1980. Content validity: Its relevance to behavioral assessment. *Behav. Assess.* 2:147–59
Longabaugh, R. 1980. The systematic observation of behavior in naturalistic settings. In *Handbook of Cross-Cultural Psychology,* ed. H. C. Triandis, J. W. Berry, 2:57–126. Boston: Allyn & Bacon
Lytton, H. 1971. Observation studies of parent-child interaction: A methodological review. *Child Dev.* 42:651–84
Madge, C., Harrison, T. 1939. *Britain by Mass-Observation.* Harmondsworth, England: Penguin
Mash, E. J., Makohoniuk, G. 1975. The effects of prior information and behavioral predictability on observer accuracy. *Child Dev.* 46:513–19
Mash, E. J., McElwee, J. D. 1974. Situational effects on observer accuracy: Behavioral predictability, prior experience, and complexity of coding categories. *Child Dev.* 45:367–77
Mazanec, N., McCall, G. J. 1975. Sex, cognitive categories, and observational accuracy. *Psychol. Rep.* 37:987–90
McCall, G. J. 1978. *Observing the Law: Field Methods in the Study of Crime and the Criminal Justice System.* New York: Free
McDermott, R. P., Roth, D. R. 1978. The social organization of behavior: Interactional approaches. *Ann. Rev. Anthropol.* 7:321–46
McGrew, W. C. 1972. *An Ethological Study of Children's Behavior.* New York: Academic
McPhail, C. 1972. *Theoretical and methodological strategies for the study of crowd phenomena.* Presented at 67th Ann. Meet. Am. Sociol. Assoc., New Orleans
McPhail, C., Pickens, R. B., Smith, R. L. 1972. *Obstrusiveness in systematic observation: A field experiment with two observation-recording procedures.* Presented at 67th Ann. Meet. Am. Sociol. Assoc., New Orleans
Medley, D. M., Mitzel, H. E. 1963. Measuring classroom behavior by systematic observation. In *Handbook of Research on Teaching,* ed. N. L. Gage, pp. 247–328. Chicago: Rand McNally
Melbin, M. 1972. *Alone and With Others: A Grammar of Interpersonal Behavior.* New York: Harper & Row
Mileski, M. 1971. Courtroom encounters: An observation study of a lower criminal court. *Law Soc. Rev.* 5:473–538
Mitchell, S. K. 1979. Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychol. Bull.* 86:376–90
Murphy, G., Goodall, E. 1980. Measurement error in direct observations: A comparison of common recording methods. *Behav. Res. Ther.* 18:147–50
Oden, N. 1977. Partitioning dependence in nonstationary behavioral sequences. See Hazlett 1977, pp. 203–20
O'Leary, K. D., Kent, R. N., Kanowitz, J. 1975. Shaping data collection congruent with experimental hypotheses. *J. Appl. Behav. Anal.* 8:43–51
Omark, D. R., Strayer, F. F., Freedman, D. G., eds. 1980. *Dominance Relations: An Ethological View of Human Conflict and Interaction.* New York: Garland STPM
Patterson, G. R., Moore, D. 1979. Interactive patterns as units of behavior. See Lamb et al 1979, pp. 77–96
Paul, G. L., ed. 1984. *Observational Assessment Instrumentation for Institutional Research and Treatment.* Cambridge, Mass: Harvard Univ. Press. In press
Plant, M. A., Kreitman, N., Miller, T., Duffy, J. 1977. Observing public drinking. *J. Stud. Alcohol* 38:867–80
Porter, L. W., Lawler, E. E., Hackman, J. R. 1975. *Behavior in Organizations.* New York: McGraw-Hill
Powell, J., Martindale, B., Kulp, S., Martindale, A., Bauman, R. 1977. Taking a closer look: Time sampling and measurement error. *J. Appl. Behav. Anal.* 10:325–32
Redfield, J. P., Paul, G. L. 1976. Bias in behavioral observation as a function of observer familiarity with subjects and typicality of behavior. *J. Consult. Clin. Psychol.* 44:156
Reid, J. B. 1978. The development of specialized observation systems. In *A Social Learning Approach to Family Intervention,* ed. J. B. Reid, 2:43–49. Eugene, Oreg: Castalia
Reiss, A. J. Jr. 1968a. Police brutality—Answers to key questions. *Trans-Action* 5(8):10–19

Reiss, A. J. Jr. 1968b. Stuff and nonsense about social surveys and observation. In *Institutions and the Person,* ed. H. S. Becker, B. Geer, D. Riesman, R. S. Weiss, pp. 357–67. Chicago: Aldine

Reiss, A. J. Jr. 1971a. Systematic observation of natural social phenomena. In *Sociological Methodology 1971,* ed. H. Costner, pp. 3–33. San Francisco: Jossey-Bass

Reiss, A. J. Jr. 1971b. *The Police and the Public*. New Haven, Conn: Yale Univ. Press

Reiss, A. J. Jr. 1976. Systematic observation surveys of natural social phenomena. In *Perspectives on Attitude Assessment: Surveys and Their Alternatives,* ed. H. W. Sinaiko, L. A. Broedling, pp. 123–41. Champaign, Ill: Pendleton

Reiss, A. J. Jr. 1979. Systematic social observation in police research. In *Police and the Social Order: Contemporary Research Perspectives,* ed. J. Knutsson, E. Kuhlhorn, A. J. Reiss Jr., pp. 271–305. Stockholm: Natl. Swed. Counc. Crime Prev.

Repp, A. C., Roberts, D. M., Slack, D. L., Repp, C. F., Berkler, M. S. 1976. A comparison of frequency, interval, and time-sampling methods for data collection. *J. Appl. Behav. Anal.* 9:501–8

Riskin, J. E., Faunce, E. E. 1972. An evaluative review of family interaction research. *Fam. Process* 11:365–456

Romanczyk, R. G., Kent, R. N., Diament, C., O'Leary, K. D. 1973. Measuring the reliability of observational data: A reactive process. *J. Appl. Behav. Anal.* 6:175–84

Rosenshine, B., Furst, N. 1973. The use of direct observation to study teaching. In *Second Handbook of Research on Teaching,* ed. R. W. Travers, pp. 122–83. Chicago: Rand McNally

Sackett, G. P. 1978a. Measurement in observational research. See Sackett 1978c, pp. 25–43

Sackett, G. P., ed. 1978b. *Observing Behavior I: Theory and Applications in Mental Retardation*. Baltimore: Univ. Park

Sackett, G. P., ed. 1978c. *Observing Behavior II: Data Collection and Analysis Methods*. Baltimore: Univ. Park

Sackett, G. P. 1979. The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In *Handbook of Infant Development,* ed. J. D. Osofsky, pp. 623–49. New York: Wiley

Sanson-Fisher, R. W., Poole, A. D., Small, G. A., Fleming, I. R. 1979. Data acquisition in real time—An improved system for naturalistic observations. *Behav. Ther.* 10:543–54

Scherer, K. R., Ekman, P., eds. 1982a. *Handbook of Methods in Nonverbal Behavior Research*. Cambridge: Cambridge Univ. Press

Scherer, K. R., Ekman, P. 1982b. Methodological issues in studying nonverbal behavior. See Scherer & Ekman 1982a, pp. 1–44

Secord, P. F., Backman, C. W. 1974. *Social Psychology*. New York: McGraw-Hill. 2nd ed.

Shavelson, R. J., Webb, N. M. 1981. Generalizability theory: 1973–1980. *Br. J. Math. Stat. Psychol.* 34:133–66

Sidowski, J. B. 1977. Observational research: Some instrumental systems for scoring and storing behavioral data. *Behav. Res. Methods Instrum.* 9:403–4

Simon, A., Boyer, E. G. 1974. *Mirrors for Behavior III: An Anthology of Observation Instruments*. Wyncote, Pa: Commun. Mater. Cent.

Simpson, M. J. A. 1979. Problems of recording behavioral data by keyboard. See Lamb et al 1979, pp. 137–56

Slater, P. J. B. 1978. Data collection. See Colgan 1978, pp. 7–24

Smith, H. W. 1983. Same-sex versus cross-sex observational recognition: An effect of attention or recall? *Percept. Mot. Skills* 57:380–82

Spool, M. D. 1978. Training programs for observers of behavior: A review. *Pers. Psychol.* 31:853–88

Stephenson, G. R. 1979. PLEXYN: A computer-compatible grammar for coding complex social interactions. See Lamb et al 1979, pp. 157–84

Strossen, R. J., Coates, T. J., Thoresen, C. E. 1979. Extending generalizability theory to single-subject designs. *Behav. Ther.* 10: 606–14

Stubbs, M., Delamont, S., eds. 1976. *Explorations in Classroom Observation*. New York: Wiley

Stunkard, A., Kaplan, D. 1977. Eating in public places: A review of reports of the direct observation of eating behavior. *Int. J. Obes.* 1:89–101

Sustare, D. 1978. Systems diagrams. See Colgan 1978, pp. 275–311

Sykes, R. E. 1977. Techniques of data collection and reduction in systematic field observation. *Behav. Res. Methods Instrum.* 9:404–17

Sykes, R. E. 1978. Toward a theory of observer effect in systematic field observation. *Hum. Organ.* 37:148–56

Sykes, R. E., Brent, E. E. Jr. 1983. *Policing: A Social Behaviorist Perspective*. Rutgers, NJ: Rutgers Univ. Press

Sykes, R. E., Larntz, K., Fox, J. C. 1976. Proximity and similarity effects on frequency of interaction in a class of naval recruits. *Sociometry* 39:263–69

Sykes, R. E., Whitney, F. 1969. Systematic observation utilizing the Minnesota interac-

tion data coding and reduction system. *Behav. Sci.* 14:167–69

Taplin, P. S., Reid, J. B. 1973. Effects of instructional set and experimental influences on observer reliability. *Child Dev.* 44:547–54

Thomson, C., Holmberg, M., Baer, D. M. 1974. A brief report on a comparison of time-sampling procedures. *J. Appl. Behav. Anal.* 7:623–26

Tifft, L. L. 1970. *Comparative police supervision systems: An organizational analysis.* PhD thesis. Univ. Ill., Urbana

Turner, S. D., Nielsen, L., Murray, M. 1983. Observation systems in program evaluation: Three examples. *Eval. Rev.* 7:373–83

van Hooff, J. A. R. A. M. 1982. Categories and sequences of behavior: Methods of description and analysis. See Scherer & Ekman 1982a, pp. 362–439

Wahler, R. G., House, A. E., Stambaugh, E. E. 1976. *Ecological Assessment of Child Problem Behavior: A Clinical Package for Home, School, and Institutional Settings.* New York: Pergamon

Wampold, B. E., Holloway, E. L. 1983. A note on interobserver reliability for sequential data. *J. Behav. Assess.* 5:217–25

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences.* Chicago: Rand McNally

Weick, K. E. 1968. Systematic observational methods. In *Handbook of Social Psychology,* ed. G. Lindzey, E. Aronson, 2:357–451. Reading, Mass: Addison-Wesley. 2nd ed.

Weinrott, M. R., Jones, R. R., Boler, G. R. 1981. Convergent and discriminant validity of five classroom observation systems: A secondary analysis. *J. Educ. Psychol.* 73: 671–80

West, C., Zimmerman, D. H. 1982. Conversation analysis. See Scherer & Ekman 1982a, pp. 506–41

Westman, R. S. 1977. Environmental languages and the functional bases of animal behavior. See Hazlett 1977, pp. 145–201

Wieder, G. B., Weiss, R. L. 1980. Generalizability theory and the coding of marital interaction. *J. Consult. Clin. Psychol.* 48:469–77

Wildman, B. G., Erickson, M. T. 1977. Methodological problems in behavioral observation. See Cone & Hawkins 1977, pp. 255–73

Wiseman, J. P. 1970. *Stations of the Lost: The Treatment of Skid Row Alcoholics.* Englewood Cliffs, NJ: Prentice-Hall

Yarrow, M. R., Waxler, C. Z. 1979. Observing interaction: A confrontation with methodology. See Cairns 1979, pp. 37–65