UNOBTRUSIVE MEASURES IN BEHAVIORAL ASSESSMENT

ALAN E. KAZDIN

THE PENNSYLVANIA STATE UNIVERSITY

A major distinguishing characteristic of behavioral assessment is the direct assessment of overt behavior. Direct assessment is assumed to provide a sample of behavior that reflects client performance in the situation in which behavior is assessed, even if the assessment procedures were not implemented. Yet, in the majority of investigations, behavioral assessment procedures are obtrusive, *i.e.*, subjects are aware that their behavior is being assessed. The potential problem with obtrusive assessment is that it may be reactive, *i.e.*, affect how subjects perform. Recent research has demonstrated that obtrusive observations often are reactive and that behaviors assessed under obtrusive and unobtrusive conditions bear little relation. From methodological and applied perspectives, additional attention needs to be given to unobtrusive measures of behavior change. The present paper illustrates unobtrusive measures in behavior modification including direct observations, archival records, and physical traces of performance. In addition, validation and assessment problems, questions about the obtrusiveness of the measures, and ethical issues are discussed.

DESCRIPTORS: assessment, methodology, observational technology, reactivity

Behavioral assessment has many distinguishing characteristics. A few of the significant features include direct assessment of overt behavior, interest in behavior as samples of performance rather than as signs of underlying personality, use of multiple modalities of assessment, and evaluation of such influences on performance as antecedents, consequences, and factors associated with the assessment procedures themselves (see Ciminero, Calhoun, and Adams, 1977; Cone and Hawkins, 1977; Hersen and Bellack, 1976).

Perhaps one of the most salient features of behavioral assessment is the direct observation of overt behavior. Direct observations may be one of the few central characteristics that unites many different areas within behavior therapy. Different methods are used to assess behavior directly depending upon the target problem,

subjects, and setting. For example, assessment of such problems as anxiety or social skills often has relied upon contrived laboratory conditions to present actual or simulated cues to evoke the target behavior (*e.g.*, behavioral avoidance tests for fear of snakes or behavioral role-playing performance to measure dating skills). Investigations in applied settings, such as schools, the home, hospitals, and the community, have directly assessed overt behavior to measure the problems, skills, and deficits focused upon in training.

Notwithstanding many benefits, behavioral assessment has some limitations. In most research in behavior modification, assessment procedures are obtrusive. *Obtrusiveness of assessment refers to the fact that subjects may, in varying degrees, be aware that assessment is going on.* Different circumstances may contribute to the obtrusiveness of the measure. For example, with direct observation the subjects may be informed that specific behaviors are of interest, be placed in a simulated or contrived situation where their performance is directly ob-

Preparation of this paper was facilitated by a grant (MH31047) from the National Institute of Mental Health. Requests for reprints should be sent to the author, Department of Psychology, The Pennsylvania State University, University Park, Pennsylvania 16802.

served, see an observer who records their behavior, or indeed serve as an observer of their own behavior (self-monitoring).

Obtrusiveness of an assessment procedure is a matter of degree. Some measures (*e.g.*, questionnaires) may be more salient to the subject than others and may sustain continuous awareness that behavior is being evaluated. Other measures may be less salient (*e.g.*, observers at the back of a classroom), and subject awareness of assessment may decrease over time (as observers become part of the classroom environment and their activities are ignored).

The potential problem with obtrusive assessment is that it may be reactive. *Reactivity refers to the influence that the assessment procedure exerts on the subject's performance.* Obtrusive measures are not necessarily reactive. Just because subjects are aware that assessment is going on does not necessarily mean that their behavior will be affected. However, obtrusive assessment may influence subject behavior and lead to conclusions that would differ from those obtained if behavior were assessed without the subject's awareness.

As with obtrusiveness, reactivity of assessment is not an all or none matter. Subject behavior can be affected in varying degrees by the assessment procedures. For a given measurement strategy, reactivity may be a function of various conditions of assessment. For example, if subjects are informed that the results of assessment will be used to determine the reinforcers that they earn, assessment may be much more reactive than if subjects are told that the results have absolutely no bearing upon their performance.

REACTIVITY OF BEHAVIORAL ASSESSMENT

Reactivity of obtrusive assessment has been recognized as a problem both in traditional and behavioral assessment (Kent and Foster, 1977; Webb, Campbell, Schwartz, and Sechrest, 1966). Different lines of research have demonstrated

the reactive effects of direct behavioral observations. For example, observations of client-therapist behavior in therapy sessions (Roberts and Renzaglia, 1965) and parent-child behavior in the home (Johnson, Christensen, and Bellamy, 1976) yield different results if gathered unobtrusively and obtrusively.

Research has demonstrated that the presence of observers in applied settings influences subject performance (Mercatoris and Craighead, 1974; Surratt, Ulrich, and Hawkins, 1969; White, 1977), although there are exceptions (Hagen, Craighead, and Paul, 1975). The reactivity of obtrusive assessment even has been evident in research on observer behavior (*e.g.*, Kent, O'Leary, Diament, and Dietz, 1974; Romanczyk, Kent, Diament, and O'Leary, 1973). When observers are monitored obtrusively by an experimenter, their interobserver agreement and use of particular behavioral categories differ from when they are monitored unobtrusively.

Other areas of behavioral research have suggested reactive influences of direct observations. Research on social skills training has recently examined interpersonal behaviors in obtrusive and unobtrusive assessment conditions where subjects provide role-play responses and are placed in a contrived situation while unaware that they are being observed (Bellack, Hersen, and Lamparski, 1979; Bellack, Hersen, and Turner, 1978). The results suggest very little relationship in social skill performance across different assessment conditions. Additional studies have shown that behaviors differ when assessed under obtrusive and unobtrusive conditions (*e.g.*, Carmody, 1978; Kazdin, 1974a; Hersen, Eisler, and Miller, 1974). The different results obtained in these latter studies cannot be attributed unambiguously to the obtrusiveness of the assessment conditions because other dimensions varied across conditions, such as when the measures were administered (posttest *vs.* follow-up) and the specific tasks presented (*e.g.*, self-report *vs.* *in vivo* test).

Perhaps the clearest demonstration that behavioral assessment can be reactive derives from

research on self-monitoring (Kazdin, 1974b; Nelson, 1977). Observation of one's own behavior often influences the behavior that is recorded. The reactivity of self-assessment has been actively researched to identify variables that contribute to behavior change (see Ciminero, Nelson, and Lipinski, 1977). In general, research from several sources has suggested that obtrusive assessment conditions may influence overt performance.

For behavior modification, the reactivity of obtrusive assessment raises both methodological and clinical issues. From a methodological standpoint, obtrusive assessment conditions constitute a method factor associated with particular assessment procedures (Campbell and Fiske, 1959). Method factors may contribute in varying degrees to the results that are obtained. Ideally, behavior should be assessed with different methods so that the contribution of unique method factors of the particular assessment procedures can be separated from the behavior of interest. In fact, behavioral research has taken considerable pains to assess behavior with many different methods in various areas of treatment. Yet, even when different assessment procedures are used, the measures usually share one particular methodological characteristic, namely, obtrusiveness. In light of the reactivity often associated with obtrusive assessment, as noted above, it would seem important to evaluate behavior with both obtrusive and unobtrusive methods.

For a clinical perspective, there may be even greater reasons to be concerned with the reactive effects of assessment. Applied research is designed to produce behavior change in a person's everyday life, usually in situations when the person does not believe that his or her behavior is assessed. For example, one would like assurances that a parent is responding appropriately to a child when observers are not in the home or when tape recordings are not made. Showing appropriate behavior only when the observer is present runs the risk that changes in behavior are restricted to reactive assessment periods. As-

essment would be desirable under conditions in which subjects were unaware that their behavior was being observed. Data obtained from unobtrusive measures provide important information about the generality of the results. In this context, *generality refers to the assessment of behavior under conditions which differ from obtrusive measures usually used for treatment evaluation.*

Unobtrusive methods of assessment need not replace any existing methods. Their major role is to supplement direct observations currently in use. Because of the specificity of behavior, the reactive effects of obtrusive observations, and the importance of assessing behavior free from observer influences, unobtrusive measures provide important information about behavior change and the assessment conditions of which behavior may be a function.

Unobtrusive measures have been used relatively infrequently in behavioral assessment. The purpose of the present paper is to illustrate various alternatives for unobtrusive measures in behavior modification research and to encourage the use of such measures in evaluating behavior-change programs. Unobtrusive assessment methods and considerations pertaining to their implementation and interpretation are discussed.

UNOBTRUSIVE ASSESSMENT TECHNIQUES

Different techniques have been used in behavior modification to assess behavior unobtrusively. Direct methods of observing behavior have been the most commonly used, but archival records and physical traces of behavior have been examined as well. The assessment methods themselves are likely to be quite familiar to most behavioral researchers because they are commonly used as obtrusive methods of assessment. Collecting data unobtrusively does not necessarily mean using entirely different assessment methods but rather ensuring that the subjects are unaware of assessment.

Direct Observation

With direct observation, behavior is observed by persons who categorize the subjects' performance according to various behavioral codes that reflect the target focus. Direct observations have been obtained unobtrusively in contrived as well as naturalistic situations.

Contrived situations. Many behaviors of interest in behavioral research might not be easily obtained in naturalistic situations because the responses are of low frequency, require special precipitating conditions, or would be prohibitive to observe in terms of available resources. Situations often are contrived to evoke responses so that the target behavior can be assessed without arousing the subject's awareness.

Contrived situations that assess behavior unobtrusively can be illustrated by programs evaluating social interaction. For example, Bellack *et al.* (1979) evaluated the social behaviors of college students by placing them in a situation with an opposite-sexed confederate. Subjects were told that a "slight scheduling mix-up" required them to wait. While waiting, the confederate precipitated social interaction which was videotaped unobtrusively and later evaluated for such measures as eye contact, duration of responding, smiles, and other measures. Similarly, social interaction among psychiatric patients has been evaluated by requiring patients to remain in a waiting room with another patient (actually an accomplice) who engaged in a prearranged sequence of behaviors to prompt social interaction (Gutride, Goldstein, and Hunter, 1973).

Hersen *et al.* (1974) developed assertive behavior among psychiatric patients and used a contrived task to evaluate treatment. Patients were short-changed, *i.e.*, presented with less money in canteen booklets after training than they were originally promised. Each patient's response to being short-changed was videotaped and later evaluated to assess overall assertiveness and latency to respond. Frederiksen, Jenkins, Foy, and Eisler (1976) evaluated the effects of

treatment designed to train patients to avoid abusive verbal outbursts. Situations on the ward that previously had precipitated these outbursts were arranged to occur after treatment. When the contrived situations were implemented, the patient's responses (*e.g.*, hostile comments, inappropriate requests) were assessed unobtrusively by staff normally present on the ward.

Occasionally, contrived situations have included phone calls that are designed to appear completely unrelated to treatment. For example, McFall and Marston (1970) called subjects who completed an assertion training program. The caller posed as a magazine salesperson and completed a prearranged sequence of requests. The client's responses were evaluated unobtrusively to study refusal behavior. Phone call measures have been used frequently to evaluate how clients respond to seemingly unreasonable requests. The calls provide unobtrusive measures because the association between the treatment program and the phone call is concealed, at least until the client is debriefed after assessment.

Naturalistic observation. Direct observation of overt behavior in the setting in which clients normally function has been widely used in behavioral research, particularly applied behavior analysis. Assessment has been conducted in classrooms, day-care centers, homes, institutions, and many other settings in the natural environment. Although many studies have assessed behaviors directly, it is not always clear whether the observations were obtained unobtrusively.

In most programs, direct observations appear to have been obtrusive. For example, in many classroom programs, observers who are stationed at the back or side of the room may make the observations obtrusive. On the other hand, the observations may be unobtrusive if observers are not present, if the teacher observes behavior, if observers are placed in a booth behind a one-way mirror, or if videotaped records are obtained from a concealed camera. Even under these latter conditions, assessment may be reactive if it is part of an intervention program in which

clients are informed that their behavior is being evaluated or that specific behaviors will be followed by various consequences.

Occasionally, the intervention program and observation procedures may facilitate unobtrusive assessment. A particularly interesting classroom example was provided by Hauserman, Walen, and Behling (1973) who increased interracial interaction among first graders in an integrated classroom. The teacher rewarded interaction with a "new friend" without specifying the need for interracial interaction during a lunch period and unobtrusively recorded whether interracial partners shared a lunch table. Observations during a free-play period were more obviously unobtrusive because of the absence of specific contingencies. Unobtrusive measures both in the lunchroom and during free-play periods afterwards tended to reflect the efficacy of the program.

Direct observations have been obtained unobtrusively in other settings. For example, Hollandsworth, Glazeski, and Dressel (1978) treated a male college graduate who was extremely anxious and deficient in verbal skills. Training improved speaking in an organized and fluent fashion, using self-corrective statements, and generating questions. An unobtrusive measure of verbal skills was obtained by having observers pose as shoppers in the store in which the client worked. Unobtrusive observations before and after treatment were made as the client interacted with customers. Observations of behaviors related to those altered in treatment corroborated the efficacy of training.

Products of Behavior

Occasionally, direct observations of overt behavior can be supplemented or even replaced by measuring various products of behavior such as administrative records or physical traces of performance. Products of behavior often are unobtrusive measures because subjects are not likely to be aware that these traces will be used for research purposes. Indeed, records and traces

of behavior often are examined years after the behaviors leading to these products have been performed (Webb *et al.*, 1966).

Archival records. Records of performance consist of information usually recorded by various administrative, governmental, and institutional agencies. Records of birth, crime, death, and marital status are examples of commonly used archival records in the social sciences. Archival records have been used in many applied behavioral programs because they can often be conveniently obtained. In addition, records of performance frequently serve as socially important criteria for evaluating intervention effects. For example, recidivism and records of arrest or contact with police have been used to evaluate behavioral interventions with prisoners (Jenkins, Witherspoon, DeVine, deValera, Muller, Barton, and McKee, Note 1), and days out of the hospital, discharge, and readmission rates have often been used to evaluate programs with psychiatric patients (*e.g.*, Paul and Lentz, 1977).

Achievement Place has used archival records as unobtrusive outcome measures, in addition to the extensive use of direct measures of overt behavior (*e.g.*, Phillips, Phillips, Fixsen, and Wolf, 1971) and self-report inventories (Eitzen, 1975). Fixsen, Phillips, Phillips, and Wolf (1976) compared the effectiveness of Achievement Place, institutional care, and probation in the rehabilitation of predelinquents. Several archival measures were used including recidivism, police and court contacts, grades and attendance at school, and records of school dropouts.

Social and community extensions of behavioral programs have frequently used archival records to evaluate performance. Programs designed to curb energy consumption in the home have obtained records from meters that register oil, gas, or electrical consumption (*e.g.*, Hayes and Cone, 1977; Seaver and Patterson, 1976). Conservation of automobile fuel has been evaluated unobtrusively by recording mileage from the odometers of cars (Foxy and Hake, 1977).

Applications in business and industry have evaluated punctuality by looking at daily time cards (Hermann, de Montes, Dominguez, Montes, and Hopkins, 1973). Cash shortages associated with particular employees who work at a cash register of a small business were evaluated by monitoring the internal record of the register that cumulates the sum of each transaction (Marholin and Gray, 1976). Research on the reduction of crime has used daily police records of burglaries as archival data (Schnelle, Kirchner, Macrae, McNees, Eck, Snodgrass, Casey, and Uelton, 1978; Schnelle, Kirchner, McNees, and Lawler, 1975). Each of these assessment methods provides the opportunity to obtain data unobtrusively.

Physical traces. Traces refer to material deposited in or removed from the environment, or to permanent physical products. Traces have been used relatively infrequently in behavioral research, presumably because many of the behaviors of interest in treatment (*e.g.*, anxiety, hyperactive behavior, depression, language, social skills) do not often leave physical traces in the environment. On the other hand, some behaviors are readily amenable to observation of trace measures, such as the number of job tasks that individuals complete or the number of items individuals may steal from each other (*e.g.*, Azrin and Wesolowski, 1974; Pierce and Risley, 1974). Yet, rarely are such measures obtained unobtrusively.

Physical traces have been used in a few areas of behavioral research. For example, research on littering has evaluated various interventions unobtrusively by assessing the number of pieces of litter or weight of accumulated litter across such settings as a zoo, athletic stadium, forest ground, movie theater, and urban community settings (*e.g.*, Burgess, Clark, and Hendee, 1971; Chapman and Risley, 1974). Hayes, Johnson, and Cone (1975) evaluated littering by marking and planting various pieces of litter among existing litter in a federal youth correctional facility. Pieces of litter were coded by making small tears, ink marks, folds and creases, and

other procedures. Deposit of the marked items among other pieces of litter permitted evaluation of pickup of existing trash without the fear that individuals would generate new trash merely to receive reinforcers.

A procedure that also involved marking items was used to provide a trace measure of theft in a retail department store (McNees, Egli, Marshall, Schnelle, and Risley, 1976). Different types of merchandise were marked with coded tags that were removed and retained when the items were purchased. Accumulation of the number of tagged items remaining in the store plus those tags removed from purchased items provided a physical trace measure used to infer missing items, *i.e.*, shoplifting.

Physical trace measures of performance have been observed extensively in classroom research. Written products of behavior have been used to assess classroom interventions for reading, arithmetic, and creative writing, and constitute important traces of behavior.¹ In most programs, however, academic responses have not been observed unobtrusively.

CONSIDERATIONS IN USING UNOBTUSIVE MEASURES

Unobtrusive measures represent an important addition to behavioral assessment. Behavioral assessment has extended traditional assessment in such ways as observing overt behavior and examining performance in the environment in which target behaviors normally occur. However, in both traditional and behavioral assessment, treatment effects have been evaluated primarily with obtrusive measures. Relatively little evidence is available attesting to the efficacy of treatment where subjects were not aware

¹Insofar as physical traces are written records, they might be classified as an archival measure. However, archival records usually refer to governmental and institutional records. The distinction is not critical to the present discussion. For an elaboration of different categories of unobtrusive measures, the reader may wish to refer to other sources (Bouchard, 1976; Kazdin, 1980; Webb *et al.*, 1966).

of the assessment procedures. For programs that attempt to alter behaviors in the natural environment, direct unobtrusive assessment may provide important corroborative data assessing whether treatment effects generalize beyond possibly reactive assessment conditions.

The advantage of unobtrusive measures in behavioral research is not that they are free from potential problems of their own. Rather, the methodological problems of unobtrusive measures, where they exist, *differ* from those of other measures that are currently in wider use (Webb *et al.*, 1966). The strength of unobtrusive assessment methods is that they can corroborate other sources of data independently of reactivity. Yet, a number of issues need to be considered in evaluating the results of unobtrusive measures.

Validation of Unobtrusive Measures

In behavioral assessment in general, relatively little attention has been given to various types of validity discussed in traditional psychometric assessment (Goldfried and Linehan, 1977). As might be expected, the attention accorded validation of unobtrusive behavioral measures is even less. Part of the reason for the lack of validation research is that unobtrusive measures are usually improvised for a single study. Because the measures are not standardized or routinely included in an assessment battery, their assessment characteristics receive less attention than might otherwise be the case. The general lack of attention to various types of validity introduces problems in interpreting the results obtained from unobtrusive measures.

The interpretation of many measures may not seem to require comment because the measures have "face validity," *i.e.*, seem of obvious relevance. Yet, the validity of many unobtrusive measures can be questioned. For example, in one study, clients who completed assertion training were phoned and asked to volunteer to serve for a few hours at a new hospital (Kazdin, 1974a). The request was designed to elicit refusal behavior because of the unreasonable re-

quests that were made (*e.g.*, volunteering at an inconvenient time). Although the measure was unobtrusive, one can seriously question whether the phone call request was an appropriate measure of assertive behavior. Requesting volunteers for hospital work may more readily assess altruism rather than assertion skills.

Investigations that employ unobtrusive measures, particularly in contrived situations, rarely provide supporting information to validate whether the construct of interest is assessed (construct validity). If the results on the unobtrusive measure reflect intervention effects and corroborate data from other measures, the interpretation of the measure is less ambiguous than if no effects are demonstrated. The latter results make it unclear whether treatment effects were weak or the assessment device was inappropriate or insensitive.

Archival records occasionally present problems in deciding what is measured. Measures of recidivism, discharge, and readmission, for example, may not reflect the behaviors of interest. For instance, contact with the police and courts after release may not necessarily reflect how well ex-prisoners are doing in terms of appropriate social behavior. Reduced contact with the law might reflect superior criminal skills or inferior police surveillance and detection. Similarly, discharge and readmission rates of psychiatric patients may reflect changes in hospital philosophy rather than treatment effects (Gripp and Magaro, 1974). Indeed, discharge and readmission may bear little relationship to the psychological status of the patient (Fairweather, Sanders, Crissler, and Maynard, 1969). Although archival data may be of interest in their own right, the information they convey may be determined by many factors other than the behavior of direct interest.

Another problem in interpreting unobtrusive measures results from the narrow range of stimulus conditions that are assessed. The difficulties in obtaining unobtrusive measures have caused investigators to sample few components of the responses of interest. For example, in contrived situations clients may be presented with a

restricted sample of stimulus conditions, such as interacting in a room with a confederate or being short-changed. The measure presents a single task from which inferences are drawn about larger segments of behavior. Given the specificity of behavior (Mischel, 1968), it may be important to draw from a larger sample of stimulus conditions to make inferences about performance. Investigations using unobtrusive measures usually provide little information to suggest that the content of the measures represents the larger domain of interest (content validity).

Unobtrusive measures do not always present interpretive problems. For example, observations of family interaction or child play at school may directly assess the target focus. If assessment is unobtrusive, interpretive problems about the data would be reduced, if not eliminated completely. Additional issues may be relevant, such as whether behaviors extend to other than the assessed situations, but this is not a unique issue to the measurement device.

Assessment Problems

Many unobtrusive methods involve direct observations of overt behavior where observers classify behavior according to specific behavioral codes. Interobserver agreement is assessed to evaluate whether observations are obtained consistently over time and among different observers. Several factors may influence the data that observers collect including changes in how the codes are applied over time (observer drift), the number of behaviors or subjects that are observed simultaneously, feedback and expectancies conveyed by the investigator, and the kind of response sequence that observers received during training (Kazdin, 1977; Kent and Foster, 1977). These problems are not unique to unobtrusive measures of overt behavior and, hence, need not be elaborated here. However, it is important to mention the assessment problems of direct observation because they present special obstacles for the investigator using unobtrusive observations.

Direct and unobtrusive measures of client behavior may be difficult to obtain because the presence of an observer, as observer, should not be detected. Yet, it may be difficult to assess interobserver agreement and to keep the observations unobtrusive if two or more observers must simultaneously record behavior. Possibly, interobserver agreement might be assessed by comparing data from one observer with tape recordings or videotaped records that could be obtained unobtrusively. Indeed, if the primary data could be obtained through audio or videotaped records, observers would not be able to influence subject performance and agreement could be obtained when the records were played.

Aside from maintaining unobtrusive assessment conditions for the subjects, research has suggested the desirability of keeping reliability checks unobtrusive for observers as well. Reactive effects of checking reliability are difficult to control in ordinary circumstances, even when no attempt is made to observe subject behavior unobtrusively. Conducting unobtrusive reliability checks on unobtrusive measures of overt subject behavior presents a formidable task for the investigator.

Obtrusiveness of the Measures

Whether many measures that are designed to be unobtrusive are in fact is a question that warrants comment. Measures occasionally are referred to as "unobtrusive" to reflect the intent of the investigator. Whether the client was unaware of the assessment procedure is a matter of surmise. For example, a confederate and experimental ruse are often utilized in contrived situations. Yet, subjects might easily see through the ruse when they are "surprisingly" interrupted on a task (Friedman, 1971) or seated in a waiting room where an accomplice coincidentally prompts treatment-relevant behavior (Gutridge *et al.*, 1973). In some cases, unobtrusive measurement takes place in the same room where the subjects have been assessed under obtrusive conditions which may further raise suspicions (Bellack *et al.*, 1979).

Investigators rarely assess whether subjects were aware of the "unobtrusive" measurement procedure. As an exception, Frederiksen *et al.* (1976) asked patients on the ward whether they had realized that situations to which they were previously exposed were contrived. The patients did not identify the situations as related to assessment. Part of the reason for this may have been that the situations were modeled after actual events that had previously precipitated abusive behavior prior to treatment. In general, if it is possible for subjects to see through the ruse, it might be useful to gather information to evaluate whether the measure was unobtrusive.

Ethical Issues

Although ethical issues arise in all human experimentation, special sensitivity to these issues may be required with the use of unobtrusive measures. Attempting to obtain measures that circumvent the subject's awareness might violate the requirements of informed consent. Obviously, from the standpoint of assessment, it would be of unclear value to request permission from the subject to obtain unobtrusive observations without sensitizing subjects to the assessment procedures that follow. Perhaps subjects could provide consent for several different types of assessment opportunities, only some of which would actually be utilized by the investigator. Occasionally, subjects have not been informed about the unobtrusive measure until after it has been administered. After assessment clients may be debriefed and provided with the option of having the information remain confidential and not be used.

Subjecting clients to assessment procedures prior to obtaining permission may not be adequately compensated by providing them with the option of not using the information. The mere act of assessment may be objectionable to many clients and would not be agreed to in advance. Also, once the information is obtained, it may provide a threat to privacy. The ethical issues of using unobtrusive measures may be attenuated or exacerbated by the specific conditions

under which assessment is conducted, the extent to which the information normally would be available, the event to which the assessment procedures place the subject under stress or risk, and similar factors.

For many unobtrusive measures, ethical issues associated with deception may not arise. For example, many archival records and physical traces would not be deceptive or present special threats to individual privacy because the information is ordinarily available and the identity of individual persons may not be used (*e.g.*, crime records in a police precinct or theft of merchandise in a store). The purposes of research may not raise ethical issues with such measures over and above those issues raised in collecting the information in the first place. Of course, how the information is used may alter the ethical requirements from the subjects' perspective.

For present purposes, it is important merely to note that ethical issues raise considerations for using unobtrusive measures. Although the issues cannot be elaborated here, their importance should not be minimized. However, ethical issues and guidelines related to circumstances, such as those associated with unobtrusive measurement have been outlined elsewhere (Ad hoc Committee on Ethical Standards in Psychological Research, 1973).

CONCLUSIONS

Behavioral assessment has made remarkable gains. Assessment methods have been exemplary in looking at different channels or modalities of behavior. Many complex clinical responses (*e.g.*, anxiety) have been carefully elaborated in the course of treatment evaluation. In applied programs, intervention effects are routinely evaluated on those behaviors that served as the impetus for treatment.

Most assessment procedures in behavior modification research are obtrusive. Because subjects are aware that their behavior is assessed, they may perform differently than they would with

unobtrusive assessment methods. The reactive effects of assessment in behavioral research need not be surmised. Several studies have shown that the "same" behavior may differ depending upon whether subjects are aware that their behavior is assessed. Interestingly, even behavior of observers has been shown to differ as a function of whether they believe they are being monitored.

The reactive effects of assessment are not necessarily artifacts or sources of bias. Indeed, many researchers have creatively researched reactivity of assessment to help produce changes in clinically relevant behavior (Ciminero, Nelson, and Lipinski, 1977; Nelson, 1977). Yet, from the standpoint of evaluating treatment gains, obtrusive assessment provides information that needs to be supplemented whenever possible with unobtrusive measures. Research suggesting little relationship across obtrusive and unobtrusive assessment conditions and the concerns over the specificity of behavior across situations argue for use of unobtrusive measures to supplement existing assessment strategies.

The use of unobtrusive measures is not the only solution that can resolve the problems created by reactive assessment. An alternative solution might be to elaborate the conditions of obtrusive assessment that contribute or lead to reactive effects. Research might elaborate the cues of the assessment situation, format of assessment, instructions to the client, and other factors that may make direct observation of behavior less reactive (see Kazdin, 1980). For example, influences of the testing situation have been researched for the behavioral assessment of fear. The level of fear that subjects evince on behavioral avoidance tests varies as a function of whether subjects are told their behavior is assessed in a clinic versus a research or laboratory setting (Bernstein, 1973), whether the measure is said to assess anxiety or physiological arousal (Bernstein and Nietzel, 1974), or whether subjects are encouraged to evince little fear (Kazdin, 1973).

Once the specific cues of the testing situation that influence subject performance are known,

they can be minimized to generate performance that might closely approximate the information obtained with unobtrusive assessment. Obtrusive measures offer many advantages, such as convenience, relative standardization across experiments, and lack of the need for deception. Hence, research devoted to the conditions that may make obtrusive measures less reactive is quite worthwhile.

Whether improvements are made in understanding and minimizing the reactivity of obtrusive measures, the use of unobtrusive measures needs to increase as well. Unobtrusive measures provide a test of the generality of experimental results and answer an extremely important treatment outcome question, namely, how do clients perform when they do not know their performance is monitored? For many intervention techniques and client populations within behavior modification, the answer to the question remains to be provided.

REFERENCE NOTE

1. Jenkins, W. O., Witherspoon, A. D., DeVine, M. D., deValera, E. K., Muller, J. B., Barton, M. C., and McKee, J. M. *The post-prison analysis of criminal behavior and longitudinal follow-up evaluation of institutional treatment*. A report on the Experimental Manpower Laboratory for Corrections, February, 1974.

REFERENCES

- Ad hoc Committee on Ethical Standards in Psychological Research, *Ethical principles in the conduct of research with human participants*. Washington, D.C.: American Psychological Association, 1973.
- Azrin, N. H. and Wesolowski, M. D. Theft reversal: An overcorrection procedure for eliminating stealing by retarded persons. *Journal of Applied Behavior Analysis*, 1974, 7, 577-581.
- Bellack, A. S., Hersen, M., and Lamparski, D. Role-play tests for assessing social skills: Are they valid? Are they useful? *Journal of Consulting and Clinical Psychology*, 1979, 47, 335-342.
- Bellack, A. S., Hersen, M., and Turner, S. M. Role-play tests for assessing social skills: Are they valid? *Behavior Therapy*, 1978, 9, 448-461.
- Bernstein, D. A. Behavioral fear assessment: Anxiety or artifact? In H. Adams and I. P. Unikel

- (Eds), *Issues and trends in behavior therapy*. Springfield, Illinois: Charles C Thomas, 1973.
- Bernstein, D. A. and Nietzel, M. T. Behavioral avoidance tests: The effects of demand characteristics and repeated measures of two types of subjects. *Behavior Therapy*, 1974, **5**, 183-192.
- Bouchard, T. J. Field research methods: Interviewing, questionnaires, participant observation, systematic observation, unobtrusive measures. In M. D. Dunnette (Ed), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976.
- Burgess, R. L., Clark, R. N., and Hendee, J. C. An experimental analysis of anti-litter procedures. *Journal of Applied Behavior Analysis*, 1971, **4**, 1-5.
- Campbell, D. T. and Fiske, D. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, **56**, 81-105.
- Carmody, T. P. Rational-emotive, self-instructional, and behavioral assertion training: Facilitating maintenance. *Cognitive Therapy and Research*, 1978, **2**, 241-253.
- Chapman, C. and Risley, T. R. Anti-litter procedures in an urban high-density area. *Journal of Applied Behavior Analysis*, 1974, **7**, 377-383.
- Ciminero, A. R., Calhoun, K. S., and Adams, H. E. (Eds) *Handbook of behavioral assessment*. New York: Wiley, 1977.
- Ciminero, A. R., Nelson, R. O., and Lipinski, D. P. Self-monitoring procedures. In A. R. Ciminero, K. S. Calhoun, and H. E. Adams (Eds), *Handbook of behavioral assessment*. New York: Wiley, 1977.
- Cone, J. D. and Hawkins, R. P. (Eds) *Behavioral assessment: New directions in clinical psychology*. New York: Brunner/Mazel, 1977.
- Eitzen, D. S. The effects of behavior modification on the attitudes of delinquents. *Behaviour Research and Therapy*, 1975, **13**, 195-299.
- Fairweather, G. W., Sanders, D. H., Crissler, D. L., and Maynard, A. *Community life for the mentally ill*. Chicago: Aldine, 1969.
- Fixsen, D. L., Phillips, E. L., Phillips, E. A., and Wolf, M. M. The teaching-family model of group home treatment. In W. E. Craighead, A. E. Kazdin, and M. J. Mahoney (Eds), *Behavior modification: Principles, issues, and applications*. Boston: Houghton Mifflin, 1976.
- Fox, R. M. and Hake, D. F. Gasoline conservation: A procedure for measuring and reducing the driving of college students. *Journal of Applied Behavior Analysis*, 1977, **10**, 61-74.
- Frederiksen, L. W., Jenkins, J. O., Foy, D. W., and Eisler, R. M. Social skills training to modify abusive verbal outbursts in adults. *Journal of Applied Behavior Analysis*, 1976, **9**, 117-125.
- Friedman, P. H. The effects of modeling and role-playing on assertive behavior. In R. D. Rubin, H. Fensterheim, A. A. Lazarus, and C. M. Franks (Eds), *Advances in behavior therapy*. New York: Academic Press, 1971.
- Goldfried, M. R. and Linehan, M. M. Basic issues in behavioral assessment. In A. R. Ciminero, K. S. Calhoun, and H. E. Adams (Eds), *Handbook of behavioral assessment*. New York: Wiley, 1977.
- Gripp, R. F. and Magaro, P. A. Token economy program in the psychiatric hospital: Review and analysis. *Behaviour Research and Therapy*, 1974, **12**, 205-228.
- Gutridge, M. E., Goldstein, A. P., and Hunter, G. F. The use of modeling and role playing to increase social interaction among asocial psychiatric patients. *Journal of Consulting and Clinical Psychology*, 1973, **40**, 408-415.
- Hagen, R. L., Craighead, W. E., and Paul, G. L. Staff reactivity to evaluative behavioral observations. *Behavior Therapy*, 1975, **6**, 201-205.
- Hauserman, N., Walen, S. R., and Behling, M. Reinforced racial integration in the first grade: A study in generalization. *Journal of Applied Behavior Analysis*, 1973, **6**, 193-200.
- Hayes, S. C. and Cone, J. D. Reducing residential electrical energy use: Payments, information, and feedback. *Journal of Applied Behavior Analysis*, 1977, **10**, 425-435.
- Hayes, S. C., Johnson, V. S., and Cone, J. D. The marked item technique: A practical procedure for litter control. *Journal of Applied Behavior Analysis*, 1975, **8**, 381-386.
- Hermann, J. A., de Montes, A. I., Dominguez, B., Montes, F., and Hopkins, B. L. Effects of bonuses for punctuality on the tardiness of industrial workers. *Journal of Applied Behavior Analysis*, 1973, **6**, 563-570.
- Hersen, M. and Bellack, A. S. (Eds) *Behavioral assessment: A practical handbook*. New York: Pergamon, 1976.
- Hersen, M., Eisler, R. M., and Miller, P. M. An experimental analysis of generalization in assertive training. *Behaviour Research and Therapy*, 1974, **12**, 295-310.
- Hollandsworth, J. G., Glazeski, R. C., and Dressel, M. E. Use of social-skills training in the treatment of extreme anxiety and deficient verbal skills in the job-interview setting. *Journal of Applied Behavior Analysis*, 1978, **11**, 259-269.
- Johnson, S. M., Christensen, A., and Bellamy, G. T. Evaluation of family intervention through unobtrusive audio recordings: Experiences in "bugging" children. *Journal of Applied Behavior Analysis*, 1976, **9**, 213-219.
- Kazdin, A. E. The effect of suggestion and pretesting on avoidance reduction in fearful college students. *Journal of Behavior Therapy and Experimental Psychiatry*, 1973, **4**, 213-221.
- Kazdin, A. E. Effects of covert modeling and model reinforcement on assertive behavior. *Journal of Abnormal Psychology*, 1974, **83**, 240-252. (a)

- Kazdin, A. E. Self-monitoring and behavior change. In M. J. Mahoney and C. E. Thoresen (Eds), *Self-control: Power to the person*. Monterey, California: Brooks/Cole, 1974. (b)
- Kazdin, A. E. Artifact, bias, and complexity of assessment: The ABC's of reliability. *Journal of Applied Behavior Analysis*, 1977, **10**, 141-150.
- Kazdin, A. E. *Research design in clinical psychology*. New York: Harper & Row, 1980.
- Kent, R. N. and Foster, S. L. Direct observational procedures: Methodological issues in naturalistic settings. In A. R. Ciminero, K. S. Calhoun, and H. E. Adams (Eds), *Handbook of behavioral assessment*. New York: Wiley, 1977.
- Kent, R. N., O'Leary, K. D., Diament, C., and Dietz, A. Expectation biases in observational evaluation of therapeutic change. *Journal of Consulting and Clinical Psychology*, 1974, **42**, 774-780.
- Marholin, D., II and Gray, D. Effects of group response cost procedures on cash shortages in a small business. *Journal of Applied Behavior Analysis*, 1976, **9**, 25-30.
- McFall, R. M. and Marston, A. R. An experimental investigation of behavior rehearsal in assertive training. *Journal of Abnormal Psychology*, 1970, **76**, 295-303.
- McNees, M. P., Egli, D. S., Marshall, R. S., Schnell, J. F., and Risley, T. R. Shoplifting prevention: Providing information through signs. *Journal of Applied Behavior Analysis*, 1976, **9**, 399-405.
- Mercatoris, M. and Craighead, W. E. Effects of non-participant observation on teacher and pupil classroom behavior. *Journal of Educational Psychology*, 1974, **66**, 512-519.
- Mischel, W. *Personality and assessment*. New York: Wiley, 1968.
- Nelson, R. O. Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler, and P. M. Miller (Eds), *Progress in behavior modification, Volume 5*. New York: Academic Press, 1977.
- Paul, G. L. and Lentz, R. J. *Psychological treatment of chronic mental patients: Milieu versus social-learning programs*. Cambridge, Massachusetts: Harvard University Press, 1977.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L., and Wolf, M. M. Achievement Place: Modification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis*, 1971, **4**, 45-59.
- Pierce, C. H. and Risley, T. R. Recreation as a reinforcer: Increasing membership and decreasing disruptions in an urban recreation center. *Journal of Applied Behavior Analysis*, 1974, **7**, 403-411.
- Roberts, R. R. and Renzaglia, G. A. The influence of tape recording on counseling. *Journal of Counseling Psychology*, 1965, **12**, 10-16.
- Romanczyk, R. G., Kent, R. N., Diament, C., and O'Leary, K. D. Measuring the reliability of observational data: A reactive process. *Journal of Applied Behavior Analysis*, 1973, **6**, 175-184.
- Schnelle, J. F., Kirchner, R. E., Macrae, J. W., McNees, M. P., Eck, R. H., Snodgrass, S., Casey, J. D., and Uselton, P. H. Police evaluation research: An experimental and cost-benefit analysis of a helicopter patrol in a high crime area. *Journal of Applied Behavior Analysis*, 1978, **11**, 11-21.
- Schnelle, J. F., Kirchner, R. E., McNees, P. M., and Lawler, J. M. Social evaluation research: The evaluation of two police patrolling strategies. *Journal of Applied Behavior Analysis*, 1975, **8**, 353-365.
- Seaver, W. B. and Patterson, A. H. Decreasing fuel oil consumption through feedback and social commendation. *Journal of Applied Behavior Analysis*, 1976, **9**, 147-152.
- Surratt, P. R., Ulrich, R. E., and Hawkins, R. P. An elementary student as a behavioral engineer. *Journal of Applied Behavior Analysis*, 1969, **2**, 85-92.
- Webb, E. J., Campbell, D. T., Schwartz, R. C., and Sechrest, L. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.
- White, G. D. The effects of observer presence on the activity level of families. *Journal of Applied Behavior Analysis*, 1977, **10**, 734.

Received 6 November 1978.

(Final Acceptance 6 April 1979.)