

NONREACTIVE MEASURES IN PSYCHOTHERAPY OUTCOME RESEARCH

David R. Patterson

University of Southern California School of Medicine

and

Lee Sechrest

Institute for Social Research, University of Michigan

ABSTRACT. *Attempts to assess outcomes of psychotherapy have relied too heavily on measures likely to be highly reactive, by which is meant that the processes of measurement affect what is being measured. Commonly employed measures of outcome such as therapist ratings or self-report are especially susceptible to effects attributable to knowledge of what is being measured. There are advantages in developing and using measures that, even though imperfect in other ways, are minimally reactive. Nonreactive measures should be considered as supplementary and not as substitutes for other measures. Problems with nonreactive measures involving ethics, validity, and other psychometric limitations are troublesome but not insurmountable. Promising categories of nonreactive measures include physical traces of past behaviors, archival records, nonverbal behaviors, and observations carried out in contrived situations. Measures with dependably low levels of reactivity, including physiological measures, should also be considered for use.*

Recently, Smith, Glass, and Miller (1980) completed their meta-analysis of over 50 psychotherapy outcome studies with results generally encouraging to proponents of psychotherapy. Although acknowledging the limitations that critics have found in their methods (e.g., Eysenck, 1978; Rosenthal, 1976), the authors were adamant in their conclusion that the research supports the efficacy of psychotherapy: "Psychotherapy is beneficial, consistently so and in many different ways" (p. 103). The impact of that conclusion, however, diminished in light of an additional finding

Requests for reprints may be addressed to David R. Patterson, Ph.D., now at the Department of Medical Psychology, Oregon Health Sciences University, Portland, OR 97201.

from their study: the experimental effect size of the outcome studies is positively related to the reactivity of the outcome measure employed. In fact, of the several variables that proved to be correlated with effect size (e.g., client IQ, therapist experience, amount of client hospitalization), reactivity accounted for the most variance. Certainly, this finding suggests that the reactive nature of psychotherapy outcome measurements ought to be examined and alternatives proposed.

Simply stated, reactivity refers to the phenomenon of the measurement process producing change in what is measured (Campbell & Stanley, 1963). Almost without exception, the measures currently used in psychotherapy outcome studies are potentially quite reactive. Reactivity can present a methodological problem in two ways (Campbell & Stanley, 1963). First, a reactive pretest may account directly for behavioral change or may enhance the effects of a particular treatment. Clients, for example, may detect that the items on a self report measure are goals for treatment and change their behavior in response to the pretest rather than the treatment. Or, a pretest may sensitize clients to a treatment so that it produces more change than would otherwise have occurred. Second, and more likely, clients may distort their responses because of their awareness of being measured, not necessarily in a predictable direction. Clients may report themselves as improved on post-treatment questionnaire items, independently of whether treatment had an effect. This reactivity could occur for any number of reasons, among them the client's desires to please the therapist or to persuade themselves they got their money's worth out of therapy. On the other hand, some clients might also report themselves as unimproved (e.g., to prolong treatment or to justify continuing financial support).

In many types of social science research, reactivity can be controlled through the use of adequate experimental designs (Campbell & Stanley, 1963; Cook & Campbell, 1979). Unfortunately, carrying out an optimally designed psychotherapy outcome study appears unfeasible (Goldstein, Heller, & Sechrest, 1966). The cost and ethical problems involved in having enough subjects and control groups to rule out the effects of reactivity are beyond what any researcher's time and resources would allow. Analogue studies may rely on experimental design to eliminate reactive effects, but standard psychotherapy studies must depend on other means (Goldstein et al., 1966).

The use of nonreactive, or less reactive, outcome measures is clearly an option worth considering (Goldstein et al., 1966; Webb, Campbell, Schwartz Sechrest, & Grove, 1981). Such measures do not require a change in the experimental design nor, in many instances, any special or extra cooperation of the subject. Although some nonreactive measures may be quite complex, burdensome to the investigator, or ethically questionable, others are clearcut and simple. Hospital records, for example, contain a wealth of potential therapy outcome measures that are often overlooked by investigators.

Optimally, it would be desirable to employ unobtrusive measures which do not require subjects to know they are being studied. Nevertheless, if such options are not available, the investigator can still employ measures that are at least low in reactivity. For some measures, the subjects' awareness of being studied is likely under most circumstances to have little influence on responses to the instrument. Ability tests (e.g., academic achievement, social skills training tests), in which the subject functions at maximal performance *because* of awareness of what is at stake, are an example of tests of relatively low reactivity—in the sense that if all subjects

try to do their best, there will be little variance to be accounted for by differential responses between treated and untreated groups.

This paper will examine the prospects for using minimally reactive measures in psychotherapy outcome research. The reactive nature of currently used measures will be examined, and suggestions will be made as to how less reactive measures might be devised and employed. Our discussion will criticize standard outcome measures for their reactive nature, but it will by no means recommend that these measures be discarded. As will be seen, nonreactive measures have their share of problems. It will be argued instead that measures low in reactivity be used to supplement currently used measures such as questionnaires and rating scales.

THE REACTIVITY OF CURRENTLY USED OUTCOME MEASURES

Several years ago NIMH published a guideline to measures that can be used to assess change in psychotherapy (Waskow & Parloff, 1975). This publication represented the culmination of an impressive effort to further the state of the art of psychotherapy research. Not only were a wide variety of change measures presented, but in each area covered experts were chosen to make recommendations as to the best instruments and evaluate them methodically. While the drawbacks to each measure were discussed by the authors, little attention was paid to the issue of reactivity, and there were no specific recommendations for nonreactive measures. The subsequent discussion will focus on the reactivity of the measures discussed by the NIMH reviewers, since critiques of other relevant characteristics can be found elsewhere (see Fiske, 1977; Fiske, Hunt, Luborsky, Orne, Parloff, Reiser, & Tuma, 1970; Strupp, 1977; Waskow & Parloff, 1975).

Before discussing these outcome measures, it will be useful to describe Webb et al.'s (1981) classification for the bias that can result from reactive measurement, especially from subject awareness of being measured. The authors discuss the four sources of respondent reactivity. The first involves the subject's awareness of being tested or the "guinea pig effect." If subjects are cognizant that they are in a study (or being evaluated as a psychotherapy client), they may alter their characteristic responses so that a misleading impression of them is obtained. It is possible that reactivity might be greater under some experimental conditions than others and be mistaken for a specific treatment effect. If all groups are equally reactive, treatment effects might be washed out. The second source of respondent bias is role selection (i.e., subjects in an experiment may only be exposing a particular self, selected from a broad range of behavior repertoires). Jourard (1971) argued that this role selection is particularly true of therapy clients irrespective of whether they are also experimental subjects. Thus, in a psychotherapy study, clients might respond from the narrow perspective of therapy clients and produce responses not characteristic of their extratherapy identities. The third way reactivity can affect the respondent is when measurement acts as a change agent. That is, the measurement process rather than the actual experimental intervention may produce behavioral changes, at least temporarily (Campbell & Stanley, 1963). Clients in weight loss programs, for instance, may lose weight in response to being weighed each day rather than because of the specific therapy to which they are exposed. Response sets constitute the fourth and final type of subject bias based on reactivity. There is a body of literature that indicates that subjects often respond to questionnaires in a set, biased manner (Cronbach, 1946; Wiggins, 1973). A common

example of this is the acquiescence response set in which subjects tend to endorse any statements, even those of opposite meaning (Cronbach, 1946; Sletto, 1937).

The investigator, as well as the respondent, contributes to reactive biases in two ways. First, interviewer characteristics may contribute substantial variance to the data. A partial list of such characteristics include: race (Cantril, 1944; Katz, 1942), age (Riesman & Ehrlich, 1961), religion (Robinson & Rohde, 1946), and social class (Riesman, 1956). Such variance will limit the generalizability of the study. Cialdini and Baumann (1981), for example, found that expressed attitudes toward ERA when questioners were "feminist" young females were considerably more favorable than would have been suggested by a nonreactive measure. The second source of investigator-based reactive bias involves change in the research instrument. To the degree that interviewers change as research instruments, and they do (i.e., Campbell, 1961; Holmes, 1958; Rosenthal, 1976), additional bias may be introduced into the results.

These four categories of bias serve as useful guidelines for examining the sources of reactivity in each of the psychotherapy change measures discussed by Waskow and Parloff (1975). The authors include the following types of outcome assessment techniques: patient inventories, projective tests, direct self reports, behavioral measures, physiological measures, interview and speech samples, and cognitive and perceptual tests. In addition, they suggest that therapist-based measures of outcome, "significant other" ratings, community informant ratings, and the evaluations of independent clinicians be included in a battery.

Patient self reports, objective tests, patient inventories, and projective tests are all susceptible to at least three types of reactive bias, although to varying degrees. It should first be noted that for all these measures, subjects are aware that they are being evaluated and/or studied; thus, we have no way of determining the degree to which these measures reflect their characteristic behavior. Second, the frequent desirability of a pretest-posttest design in order to use these measures means that reactive measurement effects may often be confounded with treatments. For example, clients given a self-esteem scale at the beginning of therapy might be especially likely to focus on related problems during therapy and perhaps show greater change, whether genuine or artifactual. Finally, the behavior reflected by these instruments may be influenced by response sets. Although self-report type measures may be especially reactive, there is ample evidence that projective tests are susceptible to bias from awareness of being measured and the desire of most persons to appear in a favorable light (Sechrest, 1968).

Despite their other appealing methodological qualities, some of the measures used in behavioral studies are among the most reactive in current outcome research (Smith, Glass, & Miller, 1980). Examples of behavioral measures are the Fear Survey Schedule (Wolpe, 1964), problem checklists, and daily life charts (Kanfer, 1975), all of which are self-report type measures. Awareness of being tested, the selection of role behaviors, and the measurement acting as a change agent can all contribute to the reactivity of the self-report measures used in behavioral studies. Many of the measures that will be recommended later in this paper involve direct behavioral observations, which are rich sources of nonreactive measures so long as subjects are not much aware that their behavior is being assessed or are not aware of which aspects of their behavior are of interest.

Two less reactive forms of measurement than those thus far discussed include the analysis of psychophysiological responses and perceptual and cognitive ten-

dencies. The former type of outcome instrument is appealing because, while clients will certainly be aware of being measured, the degree of volitional control they can exercise over their physiological responses without specific training will be minimal. Despite their appeal, Lang (1975) states that there currently "does not yet exist a standard set of physiological measures or measurement contexts that have general use in psychotherapy research" (p. 90). The measurement of perceptual and cognitive tendencies is another promising source of low reactivity data that also have not received enough attention to enable their use in outcome measures (Scott, 1975). The Weschler Adult Intelligence Test, for example, is largely immune to reactive effects (except for the malingerer) because it is an ability measure eliciting optimal performance, but its use in psychotherapy research has been minimal (Scott, 1975).

With respect to nonreactivity, the most promising of those client change assessment techniques suggested by Waskow and Parloff (1975) are speech samples from client interviews (see Kiesler, 1975). Although within psychotherapy sessions clients may be playing a relatively limited role, their speech samples can be studied with the other types of reactive bias described in this paper being kept at a minimum. Some studies that have used this technique (e.g., Gottschalk & Gleser, 1969; Matarazzo & Wiens, 1972) will be discussed later, along with other nonreactive measures of potential value.

Waskow and Parloff also suggest that outcome can be assessed by therapists, independent clinical evaluators, significant-other ratings, and community informant ratings. All of these measures may, however, be highly reactive. The potential reactivity of therapists and significant others is obvious, but even independent clinical evaluators might be a reactive form of measurement. For example, an independent clinical evaluator might employ different standards of judgment in a psychotherapy study than would be employed in a different context. Moreover, it is usually desirable to use a pretest-posttest design with these measures, thus exposing the results to the possibility of instrument change. The investigator must be especially cautious with the data obtained from human raters because of the degree to which their judgments may alter over the time between two measurement points.

It thus appears that most of the measures proposed by Waskow and Parloff (1975) are potentially influenced by at least one of the forms of reactivity bias discussed. The exceptions, physiological, cognitive-perceptual, and speech sampling measures, do not seem to have been developed sufficiently to be commonly implemented into outcome research (Holtzman, 1975; Lang, 1975; Scott, 1975). In this respect, an otherwise thorough measurement battery is left incomplete.

NONREACTIVE MEASURES AS SUPPLEMENTS

In any study in which the only measures used are reactive, the construct validity or the external validity of the study—or both—will be threatened. For example, if subjects detect from a pre-experimental self-report questionnaire that a reduction in apologetic behavior is a desirable result of therapy, then it will be difficult if not impossible to determine whether a subsequent change in frequency of such behavior can be attributed to the specific treatment or only to the pretest plus treatment. By confounding, the pretest will have become part of the treatment, thus obscuring the "treatment" construct. Reactivity would then limit the generalizability

of the study. If the clients' outcome responses are influenced by the characteristics of the interviewer or other aspects of measurement, we are restricted in the inferences we can make about other client populations.

Although this common methodological problem is discouraging, we do not advocate that potentially reactive measures be abandoned. On the contrary, the battery suggested by Waskow and Parloff (1975) represents highly informative, psychometrically sound sources of quantifiable data. What is crucial, however, is that the same source of bias, reactivity, exists in all of them.

When a consistent source of error can be found in a measurement battery, one way to remedy the situation is to add additional measures that share the theoretically relevant components of the other measures but not the same weaknesses (Webb et al., 1981). In other words, the flaw in one measure should be compensated by the strength of another. The critical shortcoming of the Waskow and Parloff battery is that there is no measure that compensates for the reactivity that runs throughout the others.

Any score can, obviously, be biased in only one of two directions: the score can be misleadingly high or misleadingly low. Consequently, if biases differ in either direction or degree, the average of two of them will be less than the one that is most extreme. If we knew the degree of bias, we would certainly choose the less biased measure. The problem is that we can rarely estimate bias in any measure except in the context provided by other measures. Barring independent estimates of bias, our naive assumption must be that by combining different measures, and up to a point the more, the better, the biases will cancel each other out. Measures can be considered to be different to the degree that they do not share common sources of bias (Sechrest, 1975). It is bias, i.e., constant error, rather than random error that is at issue in recommending nonreactive measures (Sechrest & Phillips, 1979). It should be evident that a single reliable measure, even if biased, may be preferable to a conglomeration of measures containing large components of random error.

Reactivity bias in measures may be troublesome in several different ways that have different implications for psychotherapy outcome research. First, if reactive bias can be considered to be a constant, or nearly so, it will affect interpretations of an absolute sort about outcomes but not interpretations of differences between groups. For example, if everyone tends to downplay anxiety by about the same amount, one might conclude that treated clients are better off than they actually are, but that bias would not result in exaggeration of experimental-control group differences. A constant reactivity bias would also not jeopardize comparisons of individuals. Thus, if it is not important to be able to estimate absolute levels of performance or of a disposition, a constant reactivity bias would not be troublesome. Second, reactivity might be greater in some groups than in others, e.g., there might be a reactivity treatment interaction. Differential levels of reactivity would make interpretations of either differences between individuals or between treatment groups hazardous at best. If males are more reactive on some measure than females, then scores of an individual man and a woman could not legitimately be compared. And if treatment results in greater reactivity on some outcome measure, comparison to a less reactive control group would be misleading. A third way in which a reactivity bias might be troublesome, which was mentioned earlier, is if it reduces the sensitivity of a measure, e.g., if people are generally motivated to give highly approvable answers, any effects of treatment could be obscured by a ceiling effect. Of course,

if on a pretest everyone gave highly favorable responses, a researcher would not likely consider using the measure as a posttest. In a posttest only design, however, the results could appear that both a treatment and, say, a placebo had produced quite good results. None of these problems would exist, obviously if degree of reactivity could be well estimated, for then one could simply allow for it. The problem arises because measures are reactive in some degree that is usually unknown unless it can be assumed to be vanishingly small, as with unobtrusive measures. The only way to deal with that inestimable reactivity is to use measures that are less reactive or nonreactive in combination with reactive measures in order that bias might be, by averaging, reduced.

It is, therefore, recommended that rather than dispensing with the traditional measures that have been discussed, the investigator supplement them with one or more measures low in reactivity. Nonreactive measures should not be used to replace other measures, for as will be seen, they have problems of their own. What is important is that the weaknesses of reactive and nonreactive measures do not overlap. By compensating for the weaknesses of reactive measures through the use of nonreactive instruments, investigators can greatly enhance the methodological integrity of psychotherapy outcome studies without altering the experimental design.

ISSUES IN THE USE OF NONREACTIVE MEASURES

It has been argued that the use of nonreactive measures can compensate for a common deficiency in psychotherapy outcome research that cannot be remedied through experimental design or other types of measures. If nonreactive measures are to be implemented, it is important for the researcher to be cognizant of their weaknesses so that other instruments that are not fallible in the same respects can be included in the outcome battery. The limitations of nonreactive measures fall into four categories: ethical problems, construct validity, psychometric weaknesses, and the lack of systematization (Webb et al., 1981).

Ethical Considerations

The use of nonreactive or unobtrusive measures is likely to stimulate more ethical considerations than will their reactive equivalents. The primary reason for this is that the nature of these measures often requires that subjects not be aware of being tested; thus, their private concerns are being threatened. Many nonreactive measures applicable to psychotherapy require that the subject be oblivious to the measurement process. The ethical concerns associated with most of these measures need not become an issue if the researcher attends to two of the guidelines set forth by the APA ethics committee (1973).

The first of these principles requires that the investigator inform subjects of all aspects of the research that within reason would discourage them from participating. In addition, the investigator must explain to the subjects those aspects of the study about which they inquire. The researcher must strike a delicate balance in this respect. Clients must be informed that they may be observed unobtrusively by the therapist, family, or other informants. The description of the nature of this observation must be explicit enough to satisfy subjects, yet not enough so that they are able to predict when they will be measured. Naturally, the subject must feel

free to refuse being measured in this manner. The subtle coercion the subject may experience if the therapist is soliciting the agreement must be given particular attention.

The second applicable principle is that the subject must be protected from physical and mental discomfort, harm, and danger. Fiske (1977) has pointed out that psychotherapy researchers may become caught up in the noble function that treatment is purported to serve. Because psychotherapy is, for the most part, directed at alleviating human suffering, investigators may come to believe that any procedural decision they choose to evoke is justified. It is important that this line of thinking does not result in the investigator producing undue distress in the client.

The ethical dilemmas that may arise in nonreactive measurement are brought to light when one considers the clever methods used by McFall and his colleagues (McFall & Lillesand, 1971; McFall & Marston, 1970) to test the efficacy of assertiveness training programs. These investigators had confederates, posing as salespeople, phone the subjects and attempt to convince them to buy magazines or volunteer to help "stuff envelopes" for a local campaign. Subjects were kept on the phone by the persistent confederates until they had acquiesced or had refused five times. The subjects were subsequently phoned again, debriefed, and asked if the tape recording of the phone conversation could be used in the study. These studies were clearly experimental in nature, and treatment was a laboratory analog of behavior rehearsal therapy. Thus, there were no problems as far as the clients feeling deceived by the therapist and/or experiencing anxiety associated with evaluation concerning the outcome of the phone calls. The same could not be assumed were the subjects actual therapy clients. The ethical issues associated with applying this method to a standard therapy outcome study are important to consider, since McFall's research demonstrates a highly compelling use of nonreactive measures.

Ethical issues place an annoying but necessary restraint on the methodologist's creativity. For example, examining people's garbage has proven to be a rich source of social-psychological data (Rathje, 1979; Rathje & Hughes, 1975), but when applied to therapy outcome studies would be an unwarranted invasion of privacy, since individual clients would be identifiable. Some of the measures that will later be proposed may, if used with sufficient concern for ethics, lose some of their potential for nonreactivity.

Construct Validity

One probable reason nonreactive measures are infrequently used in psychotherapy research is that little is known regarding their construct validity. As interesting as it may be, for example, that some people have many more speech dysfluencies than others, that observation is of no value unless we know what it means, i.e., what construct can be presumed to underlie it. Fiske (1977) defines construct validity as the degree to which "the measure yield(s) empirical relationships (both null and non-null) consistent with those derived from the conceptual framework within which the target construct is imbedded" (p. 33). At this point the literature is not only lacking in demonstrating empirical relationships between nonreactive measures and the constructs presumed to account for them, but there has also been little attention paid to the advantages of deriving such measures from a theoretical framework.

Some exceptions to the latter case are the anxiety and hostility scales developed by Gottschalk and Gleser (Gottschalk & Gleser, 1969; Gottschalk, Winget, & Gleser, 1969). These scales were evolved from a psychoanalytic framework. They involve taking grammatical clauses from 5-minute samples of experimental interviews in which the subject is instructed to talk about an interesting or dramatic personal experience. The clauses are rated for the degree to which they reflect certain constructs. The Hostility Directed Outward Scale, for example, ranges on a continuum from denial of hostility, through reference to anger without an object, to hostility toward a situation or intrahuman objects, and finally to varying degrees of hostility toward humans. Although the difficulties with reducing psychoanalytic constructs to testable observations are well known (Fisher & Greenberg, 1977), the work of Gottschalk and Gleser remains a good example of how theory can lead to a relatively nonreactive observation.

The research done on human figure drawings as an indicator of a person's well-being provides a good example of the extent to which intervening studies have to be done in order to develop an empirically-based construct validity for non-reactive measures.

If clients could be required to draw human figures before and after treatment in the context of their being something other than psychological tests, then the drawings could potentially be used as a nonreactive addition to an outcome battery. Before this could be the case, however, it would be necessary to demonstrate that figure drawings are related to a construct of interest. The early research indicated that, indeed, human figure drawings reflected a characteristic of patients that ideally would change with treatment. Rapaport, Gill, and Shafer (1946) and Machover (1949) suggested that persons in a depressed state are likely to draw smaller human figures than those who are not in this condition. Lewinsohn (1964) lent further support to this conjecture by demonstrating a relationship between height of figure drawings and depression in psychiatric patients. Similarly, Pepitone and Gray (1964) found that the size of figure drawing is affected by the drawer's self-esteem. This evidence in itself would suggest that figure drawings could provide a low or non-reactive measure with empirically based construct validity. However, later research was less encouraging. Roback and Webersinn (1966) reported that the difference in figure size between depressed and non-depressed patients only approached statistical significance. Moreover, Salzman and Harway (1967) failed to find a difference in height or area of figure drawings between depressed patients and normal controls. Further, they failed to demonstrate a change in drawing size when the depressed patients were retested after showing improvement on other measures. The work that has been done on human figure drawings suggests that considerable additional work may have to be done on some nonreactive measures in order for one to be confident of their construct validity.

The process of narrowing down a psychotherapy construct to a measurable concept is a tricky one. Strupp (1977) points out that what may be regarded as improvement in one case may be deterioration in another. As an example, an increase in sexually-oriented interactions with members of the opposite sex may be seen as desirable for a shy client, while a decrease in the equivalent behavior might behoove a Don Juan type. Or, a change toward more stylish clothing might seem positive for inhibited clients, while a more conservative trend would seem to reflect positive change in exhibitionistic clients. These examples point to the necessity for research that will define the theory supporting each measure one uses.

Research needs to be directed to different aspects of the nonreactive measurement problem. Whether a given measurement is, or may get to be, reactive must first be determined. Will clients become aware that observers are taking note of their dress? Since measures are reactive to varying degrees, it is the degree of reactivity that needs to be determined, and then the implication of that level of reactivity for outcome measurement must be assessed. Next, it must be empirically established that change in the outcome measure is actually related to improvement in therapy. For example, an empirical relationship must be established between change in dress and other measures of psychological well being. Further, the direction of the change that is considered an improvement should be established *a priori*. What type of change in dress would be expected to represent improvement for what type of clients? Little research on these interim steps now exists. Until it does, use of nonreactive measures in psychotherapy outcome studies will languish.

Psychometric Properties and the Lack of Systematization

Some measures low in reactivity may not have any more psychometric problems than those typical of standard questionnaires. However, unobtrusive measures, which comprise a large proportion of nonreactive measures, are often a different story. Webb and his colleagues (1981) point out that unobtrusive measures often do not lend themselves to the statistical analysis necessary to substantiate their psychometric properties. As opposed to objective tests and questionnaires, it is difficult to obtain variability estimates for unobtrusive measures across different versions of the same measure, across occasions and across time. In other words, the investigator is hard put to obtain reliability estimates for these types of measures. A second problem recognized by the authors is that unobtrusive measures may lack sensitivity, especially in situations that involve binary choices. Often, unobtrusive measures will reflect only "either/or" choices without considering the range of options in between.

An additional difficulty is that there is currently no systematic list or catalog of unobtrusive measures. Certainly, there is no such list for the psychotherapy outcome researcher, for whom very few examples can be found. At present it appears that when such measures are used, they are developed by researchers to meet their idiosyncratic needs, and the same measure is seldomly used in more than one study.

The first two limitations are ones that the investigator will have to tolerate. With respect to the third problem, if unobtrusive measures become more frequently used and better understood, a generalizable taxonomy of measures may be developed. The initial attempt in Webb et al. (1981) is a start.

INDIVIDUALIZED MEASUREMENTS

Interestingly, the difficulty found in systematizing nonreactive measures also suggests a potential advantage of their use. The face and content validity of some nonreactive measures are sufficiently strong enough that standardization is not necessary, and measures can be individually tailored to specific clients. Although it is easier to apply a single uniform instrument to all clients in an experiment, establishing individual criteria for behavior change for each particular case has both practical and methodological advantages (Goldstein et al., 1966). The use of standardized instruments to indicate therapeutic changes may "wash out" individual

gains by nature of the instrument. To take an example, elevation on an MMPI scale may indicate improvement for one client and a relapse for another; an elevation on scale 8 (schizophrenia) may suggest a welcome increase of aesthetic interests in a normal client, but for the schizophrenic, this change would more likely reflect a higher incidence of pathological thoughts (e.g., Dahlstrom, Welsh, & Dahlstrom, 1960). Similarly, changes in Scale O (Social Introversion) have to be considered within the context of whether it is desirable for clients at that particular time in their lives to move in a direction of social introversion or extroversion. Since MMPI interpretation most commonly involves the configuration of several scales (Dahlstrom et al., 1960), the advantages of specifying improvement on an individual basis becomes even more apparent.

An example of how idiosyncratic measures can provide an appealing, nonstandardized alternative can be seen in the use of smoking and drinking behavior. Potential nonreactive means of measuring these behaviors will be discussed later. For a therapy client with deteriorating lung functioning or on the verge of alcoholism, a decrease in smoking or drinking respectively would be a powerful indicator of positive outcome. Yet such measures could not be blindly applied to an entire subject population, for some clients neither drink nor smoke, and for those who do, changes in consumption may not necessarily be indicative of improvement. Provided that the correct experimental design is used, the creative researcher could fit a nonreactive measure to almost every client that would reflect the goals of treatment.

Goldstein et al., (1966) stress the methodological ease with which individualized measures can be implemented into psychotherapy outcome research. The experimental design using such measures would require only that pairs of patients matched for outcome criteria be assigned to treatment (or alternate treatment) groups, or that assignments are made at random from a large pool of clients. Aside from the actual construction of individually tailored measures, the experimental design problems would be minimal. Moreover, such measures would allow for the use of more heterogeneous groups of clients; clients undergoing treatment to decrease sexual promiscuity could be included in the same experimental groups as those attempting to increase sexual activity.

As will be seen in the next section, few, if any, nonreactive measures are standardized. The fact that individualized measures will often be desirable is important in this regard, particularly because most of the nonreactive measures found in the literature can be applied to individual cases. From this perspective, the present lack of standardization data available on nonreactive measures is less discouraging.

POTENTIAL NONREACTIVE MEASURES

Since there are many nonreactive measures that could easily be added to a psychotherapy outcome battery that would strengthen a study methodologically, it is puzzling that they are seldomly found in psychotherapy outcome research. One possible reason for this is that researchers are not aware of their availability. This section will suggest a variety of different nonreactive and unobtrusive measures that can be used to assess psychotherapy outcome. For some of the suggested measures, there exists empirical evidence indicating that they are related to constructs that would reflect therapeutic improvement or deterioration. Investigators may feel comfortable directly incorporating such measures into outcome research.

Other measures will be suggested that make intuitive sense but lack evidence indicating that they reflect change in psychotherapy. There is a dual purpose behind including the latter measures. First of all, their mention may stimulate investigators to perform some of the interim research necessary to indicate that the measures are related to the variables in questions. Second, they may serve to stimulate the researcher's creativity. The list of possible nonreactive measures seems endless, and the range presented by Webb et al. (1981), though wide, is far from exhaustive. Our hope is that researchers will start to implement innovative nonreactive approaches to assessing behavior change as their experimental designs and modes of treatment dictate.

Some of the possibilities offered will, understandably, strike the psychotherapist as trivial in meaning. When considered independently, such measures *will* be of limited importance. However, if a number of these measures are combined with other nonreactive measures, as well as with a standard battery, they may help in important ways to crystallize a meaningful pattern (Goldstein et al., 1966). Such aggregates of nonreactive measures can, at times, be formed with minimal effort.

In some instances the suggested measures may appear ethically questionable. The authors implore researchers to use the suggested measures only after careful consideration of ethical issues. It is difficult to make absolute ethical pronouncements. What may be ethical under one circumstance may not be under another. There is no substitute for a mature ethical sensitivity. In some cases the nonreactive nature of some of the measures will be lost in the process of obtaining necessary participant consent. It will be possible to determine the feasibility of some of the suggested measures only by testing them out in actual studies.

The suggested measures will often be applicable only to particular experimental designs. Using characteristics of the client's clothing as a measure, for example, may lend itself well to a time series comparison: patterns of dress can be compared for individual clients before and after therapy. Application of such a measure in an experimental-control group comparison may be less satisfactory. On the other hand, telephone calls to assess assertiveness cannot be repeated often enough to be useful in a time series but may be an excellent measure for comparing control and experimental groups. Other proposed measures may be particular to such factors as the client's severity of illness, setting of treatment, age, and other factors.

Finally, in presenting this final section, it is necessary to acknowledge the heavy debt Webb et al.'s (1981) recent work on nonreactive and unobtrusive measures. Webb et al.'s highly inclusive work certainly should be consulted if the reader wants to pursue wants to pursue the area of nonreactive measurement further.

Physical Traces

The traces that clients leave behind them in the therapy room, waiting room, or psychiatric ward may provide valuable information. A series of such observations may serve as an indicator of overall improvement or deterioration. Palmer and McGuire (1973) estimated the extent to which psychiatric inpatients had contact with the community by the number of pieces of incoming mail they received. Holahan (1972) used cigarette butts left in the dayroom and coffee consumption to obtain an estimate of social interaction among psychiatric patients. His estimates through these methods agreed with those of an actual observer. That the number

of cigarette butts left during a therapy hour may serve as an indicator of client anxiety during that session seems a hypothesis worth testing.

The position of the furniture left in the therapy room is another trace measure worth considering. Feshbach and Feshbach (1963) noted that the diameter of a circle of children decreased as fear induced by ghost storytelling increased. Although the data are admittedly anecdotal, one of the present authors had the opportunity to observe the position of furniture left by psychotherapy clients who were responsible for setting up chairs each group session. Towards the end of treatment the chairs were placed closer together, which may have reflected the greater feelings of closeness reported by the clients. (Note, however, the construct validity problem so obvious in interpreting one phenomenon of circle closing in terms of anxiety and another in terms of social closeness.) Some years back, Johnson and Sechrest (1968) found that subjects treated for test anxiety turned in neater test papers than untreated subjects.

Archives

Archival data provide a promising source of nonreactive data because of their abundance and usually low cost. The psychotherapy researcher in particular may find this a prolific source of information because of the customarily thorough records kept by psychiatric hospitals and outpatient clinics. Moreover, such data are unlikely to be biased because their recorder is not aware of being studied by the outcome researcher. Nevertheless, Webb et al (1981) warn that two types of biases can occur: selective deposit and selective survival. The former involves a bias in the information that is entered into the records. Therapists, for example, may tend to report disproportionately the successful parts of their sessions in progress notes. Selective survival, on the other hand, results from records that are lost, destroyed, or purposely omitted.

There are seemingly endless indicators of behavior change that should be found in hospital records. Mental health researchers have already taken advantage of several of these measures, including hospitalizations, days out of the hospital, and median days of hospitalization (Karon & Vandenbos, 1974; Palmer & McGuire, 1973; Paul & Lentz, 1977; Wellish & Ro-Trock, 1980). Additional possible measures include number of appointments kept, ward ratings, nursing notes—including simply their volume, attendance at optional social affairs (Goldstein et al., 1966), number of trips to the seclusion room, the number of incidents of aggressive acting-out, and requests for medication. In one study which took advantage of this information (Fixsen, Phillips, Phillips & Wolf, 1976), treating a small group of delinquents in a family type arrangement was compared to institutional care and probation. The archival measures used, including recidivism, police and court contacts, grades, school attendance, and records of school dropouts all indicated the advantages of treating delinquents in a homestyle program. Julian and Kilman (1979) have also reported use of recidivism rate as an archival measure of outcome in the treatment of delinquents. The records at outpatient facilities may also contain potentially valuable information: for example, the number of therapy appointments kept, the number of times late for sessions, telephone calls, and medication compliance. Needless to say, it is necessary that the types of data used be consistent with the nature and severity of the client's disorder.

A particularly important source of archival data for psychotherapy outcome is

the frequency of sick calls. The notion that psychological factors play a large role in the etiology of medical conditions is widely accepted. In one study, the frequency of college student sick call visits was found to be related to the subject's amount of stress or his or her tendency to play the sick role (Mechanic & Volkart, 1961). Follette and Cummings (1967) have argued that psychotherapy can reduce the frequency of use of medical services. The possibility that sick call visits or physician's office visits can be useful for between-group comparisons in psychotherapy research seems well worth considering. Use of medical services may not be sufficiently sensitive as a measure and may be too susceptible to regression artifacts to be of value in a time series.

The assessment of outcome for adolescent psychotherapy clients is amenable to a number of archival indicators. With this population, relevant records yield unusually concrete data that can enable improvement to be operationalized.

As is true with adolescents, it may be necessary with adults to go outside of the therapy setting to obtain the necessary archival data. Fiske (1975) noted that marital records may be used as an outcome measure, and Cookerly (1980) used divorce rate to determine the efficacy of marital therapy. The client's employer may also be a useful source of data concerning productivity at work and attendance records. Bland and Orn (1980) used economic productivity as a outcome indicator with schizophrenics.

A promising use of archival records for the evaluation of therapeutic outcome has been suggested by Simons, Wade, Morton, and McSharry (1978). The authors used Goal Attainment Scaling (GAS) to determine improvement. This technique has been extensively used in various mental health, drug, criminal justice, rehabilitation, and education settings (cf. Garwick & Brintmall, 1974). With regard to unobtrusive measurement, the outstanding feature of Simons et al.'s (1978) use of GAS was that goals were established and attainment was measured retrospectively by independent raters who reviewed case records. Thus, there was no chance for reactivity to influence measurement since even the goal setting was not done by the therapist or client. The authors reported that their independently derived GAS scores were highly similar to therapist ratings of global improvement and client reports of GAS scores taken at followup. Moreover, they reported that the record-based GAS scores were sensitive to change, reliably determined, and cost effective. One problem with this technique was that in many of terminated case records sampled, inadequate intake, progress, or termination notes prevented the use of this rating techniques. The effects of this drawback could seemingly be reduced with minimal influence on reactivity, simply by encouraging more thorough record keeping.

Outcome may be reflected in records of handwriting and speech. Dollard and Mowrer (1947) developed the Discomfort-Relief Quotient as a way to compare the amount of tension or discomfort indicated in samples of writing or speech. Mowrer (1953) indicated that the DRQ changed systematically over the course of therapy. This measure can be obtained without the knowledge of the therapist or client, thus it is minimally reactive.

This brief presentation far from exhausts the possible uses of archival data. In most cases the types of data will be useful in making experimental and control group comparisons. The Discomfort-Relief Quotient is an exception to this.

Simple Observation

Simple observations are "simple" in that the investigator does not intrude in the production of material. Sources of this type of data may be the therapist, observers in the waiting room or behind two-way mirrors, secretaries, hidden cameras, and so on. This class of observations can be divided into physical appearance and nonverbal behavior.

The client's physical appearance can be assessed through simple observation and is a potential source of outcome measurement. Kane (1958, 1959, 1962) has related various patterns of dress to client characteristics that could be used as outcome indicators. Among the variables in dress he has considered are texture, pattern, amount of clothing, and color. These have been related to the patient's mood, traits, and personality changes. Dark colors, for instance, have been associated with depression. Preferences for certain aspects of textile design (although not for actual dresses, as yet) has also been related to need for achievement (Green & Knapp, 1959) and to extraversion and introversion (Sales, 1971). Psychiatrists (e.g., Kolb, 1977) have long advocated using observations of grooming and hygiene as data with which to make clinical inferences. In general, if the researcher is going to use physical appearance as an outcome variable, the therapist should be instructed not to comment on the client's dress or grooming.

Nonverbal behaviors have been the focus of a great deal of attention for their potential as sources of information about people. Harper and Wiens (1979) have reviewed the use of nonverbal behaviors as unobtrusive measures. Their discussion is particularly helpful because it considers the function that nonverbal behaviors play in the communication process as well as the information that these behaviors make available. The authors categorize those behaviors that can serve as unobtrusive measures as follows: noncontent aspects of speech, facial expression, body movements or kinetics, visual behavior and human use of space. Virtually every dimension of noncontent speech behavior has been touched upon in one study or another. Horowitz and his colleagues (Horowitz, Weckler, Saxon, Livaudais, & Boutacoff, 1977) have given thorough attention to the relationship between speech disruptions and a speaker's discomfort. The investigators found that speech disruptions varied systematically with the sensitivity of subject matter that subjects were requested to discuss. To measure speech disruptions, the content of subject's speech was typed verbatim and a frequency count was made of false starts, stammar, repeated words and nonsubstantive phrases such as "you know." The work of Horowitz and his colleagues is particularly relevant to the present discussion when one considers the possibilities of measuring speech disruptions in response to the discussion of certain psychotherapeutically relevant content areas. Progress could be measured by examining the speech disruptions on a pretest-posttest basis as clients talk about the conflict areas that brought them to treatment.

As was discussed earlier, one of the potential advantages of many nonreactive measures is that they may be individually tailored to a client's presenting problem. The use of voice tone in studying cardiac patients who exhibit "Type A" behavior provides an interesting example. Voice volume and speech rate characteristics have been described as some of the main identifiers of Type A patients (Scherwitz, Berton, & Leventhal, 1977; Schucker & Jacobs, Jr., 1977). Friedman, Brown, and Rosenman (1969) described a method by which voice intonations were converted

into an electrical analogue. On the resulting "voice analysis test," roughly 84% of Type A cardiac patients responded to this test in a manner that allowed their differentiation from persons exhibiting a Type B behavior pattern. Like the use of speech disruptions, the possibility of using voice tone analysis on a pre-post test analysis (without the subject's awareness, of course) is intriguing.

Whether the behaviors assessed as psychotherapy outcomes are dispositional or situational is an issue of interest (Mischel, 1968). For this reason, Murray's (1971) discussion of speech behavior, as it is related to situational and dispositional anxiety, is pertinent. After reviewing the relevant literature, Murray concluded that there is "a definite tendency for dispositional anxiety to be related to verbal quantity and negatively to silence" (p. 255). He also stated that the most frequently used measures of verbal quantity are time talking, number of clause units, number of interactions, and duration (time talking divided by number of interactions). For more specific examples of these measures, the reader may consult the work of Matarazzo and his colleagues (Hare, Waxler, Saslow, & Matarazzo, 1960; Matarazzo, Holman, & Wiens, 1967).

Verbal quantity may be the most useful noncontent speech behavior available to study depression as well as anxiety. Frequency of measures of speech were used by Rice, Abrams, and Saxman (1969) to characterize "flat" affect in depressed patients. Starkweather (1967) reported that when depressed patients improve, a difference can be observed in measures of their speech productivity, as well as in timing, energy, and pitch.

Even though facial expressions are regarded by several communication experts as the most important nonverbal channel (Dittman, 1972; Harper & Wiens, 1979; Knapp, 1972), this behavior has not been extensively employed in studies of psychiatric disorders or psychotherapy outcome. Trujillo and Warthin (1968) found that when compared with other patients, those with chronic duodenal ulcers have more vertical furrows on their forehead. They attributed this to the ulcer patient's increased frowning and noted that frowning is associated with rage, pain, discomfort, and hunger (e.g., Cannon, 1953; Darwin, 1965). Most of the work relating facial expression to psychiatric conditions has been done in the area of depression. Schwartz and his colleagues (Schwartz, Fair, Greenberg, Mandel, & Klerman, 1975) and Waxer (1976) have suggested that facial indices of happiness and sadness may serve as a measure of depression. Along the same lines, Rubenstein (1969) noted that depressive patients smiled more often after ECT.

The use of facial expression as an indicator of underlying subjective states requires vigorous definition since expressions of pure emotion are probably rare (Harper & Wiens, 1979). More often than not affect will be a blend of two or more common forms of emotion. For this reason, the work of Ekman and his colleagues (Boucher & Ekman, 1975; Ekman, 1972; Ekman & Friesen, 1978; Ekman, Friesen & Tomkins, 1971) is valuable because it identifies different areas of facial expression as they correspond with seven primary affect states (happiness, surprise, interest, fear, sadness, anger and disgust-contempt). This work has led to the development of the Facial Affect Coding System (Ekman et al., 1971), by means of which the affective experiences of individuals can be interpreted accurately from facial expressions.

Kinetics, or body movements, is a third class of nonverbal behavior that has stimulated research and attempts at codification. Ekman and Friesen (1969) have reported a method for classifying such movements; however, few researchers have

followed up this work with anything that could be considered applicable to psychotherapy outcome research. An analysis of kinetic behavior was made of hand movements that accompanied speech of schizophrenic patients (Grand, Freedman, Steingart, & Buchwald 1975). The patients who showed a greater number of body focused hand movements were more prone to hospitalization than those patients who used their hands expressively with speech. "Leg jiggling," a kinetic behavior regarded by Sechrest and Flores (1971) as a indicator of nervousness, was found to occur more frequently in coffee shops than cocktail lounges and more during class examinations than in ordinary classes. Jurich and Jurich (1974) used posture relaxation as one of several nonverbal expressions of anxiety. Thus, kinetics might have some value in studies of psychotherapy outcome.

Face gaze is an additional nonverbal behavior reported by Harper and Wiens (1979) that can be reliably coded. Strongman (cited in Webb et al., 1981) has indicated that eye contact is related to several interpersonal variables, among them dominance, aggression, humility and amorous intent. Others have found gaze to be positively related to extraversion (Argyle & Ingham, 1972; Kendon & Cook, 1969). There is also evidence that eye contact is related to psychiatric syndromes. Hinchliffe, Lancashire, and Roberts (1970, 1971) found that hospitalized depressives spent a greater proportion of time in eye-contact than surgery patient controls and "recovered" depressives spent less time in this behavior than their depressed equivalents. Similarly, Cranach (1971) reported that an increase in visual orienting responses in depressives seemed to precede improved verbal interaction. Using videotaped standard interviews, Rutter (1973) found that both schizophrenics and depressives spent less time looking at the interviewer than did nonpsychiatric controls.

A final nonverbal behavior that can be considered as a potential nonreactive measure is human use of space. Sommer (1959, 1965, 1967a, 1967b) has done extensive work in social distance and positioning and has described methods for operationalizing these variables in terms of linear distance. Social distance operationalized in these terms has been related to such potential outcome characteristics as competition and cooperation (Sommer, 1965), introversion-extraversion (Leipold, 1963), and vulnerability and inadequacy (Dykman & Reis, 1979).

Harper and Wiens (1979) caution that in using nonverbal behaviors for assessment it is difficult to derive stable and reliable measurements. They do, however, suggest some means by which this difficulty can be overcome:

Many of the above problems can be resolved through the use of recording devices. Videotaping provides a record of the audible and visible behavior that can be copied, played, and replayed. This capability allows the application of the most modern technological measurement devices. Audio-recordings in turn can be subjected to a randomized splicing technique (Scherer, 1971) which permits exact measurement of pitch characteristics of speech that is unrelated to its sequence. Slow motion analysis of videotapes can reveal micromomentary facial expressions and body movements that convey considerable information (Haggard and Isaacs, 1966). This technique can also be used to isolate the beginning and end of different movements for precise duration measures that would be impossible or difficult to accomplish through "real time" observation. With respect to proxemic phenomena, Scherer (1974) has described an ingenious means of geometrically interpreting photographs of individuals interacting to obtain precise interpersonal distance measurements. Utilization of multiple channel interaction recorders to obtain frequency and duration measurements (Morris and others, 1968) can ensure rapid data reduction. Ekman and Friesen's development of a Facial

Affect Coding System (Ekman and Friesen, 1978) further provides a precise means for evaluating facial affect expressions ranging from pure expressions to affect blends. Duncan and Fiske (1977) have developed a computer program that will automatically research and identify complex interaction sequences (p. 68).

Contrived Observations

Obtaining data by simple observation may not always be efficient or even possible, and the investigator must consider "perturbing" the client's environment. Perturbing stimuli can often be introduced without alerting persons to the fact that they are being studied (Sechrest, 1968). Contrived observations require particular attention to ethical issues, since such measurement often involves manipulation and/or deception of the subject.

The work of McFall and his colleagues (McFall & Lillesand, 1971; McFall & Marston, 1970) in assessing the outcome of assertiveness training has already been discussed. The reader will recall that these investigators had confederates posing as salespeople call subjects and put pressure on them to either acquiesce or decline an offer. Several additional clever ideas for assessing outcome have been reported since McFall's work.

Kazdin (1974) measured the effectiveness of an assertiveness training program by calling subjects and evaluating their responses to unreasonable requests to do volunteer hospital work. However, he later (1979) questioned the face validity of this measure when he considered that acquiescing to such a response may have reflected altruism instead of nonassertiveness. Hersen, Eisler, and Miller (1974) used a contrived task to measure the effectiveness of assertiveness training with psychiatric patients by "shortchanging" them at the hospital canteen and observing their responses. The temporary short-changing of the patients involved crediting them with less money in their canteen booklets than they were originally promised. Their interaction with the experimental accomplice at the canteen was videotaped, and the patient's response was assessed with regard to overall assertiveness and latency of response.

In another study involving a contrived interaction on a psychiatric ward, (Fredriksen, Jenkins, Foy, & Eisler, 1976) the outcome of a training program to modify abusive verbal outbursts in two adult patients was evaluated. The patients were confronted with situations on the ward that previous to the training would have elicited verbally abusive outbursts. The patients' responses were videotaped, and five target behaviors (looking time, irrelevant comments, hostile comments, inappropriate requests and appropriate requests) were assessed. Neither patient was able to identify the on-ward situations used in assessment, which suggested that the measures were successfully unobtrusive.

Unobtrusive observation of contrived interaction has also been used to assess the effectiveness of social skills training. Gutride, Goldstein, and Hunter (1973) arranged for psychiatric patients to interact with a second "patient" (an experimental accomplice) in a hospital waiting room. The patients, who had undergone a program to increase social interaction behavior, were observed through a one-way mirror as they reacted to a prearranged sequence of behaviors from the experimental accomplice. The observer rated the presence of several interactional patient behaviors at 30-second intervals: eye contact, forward leaning, physical contact, smiling, conversation initiation, conversation responsivity, and length of

talking. The inter-rater agreement for the measures was found to exceed 90 percent, and the results they yielded parallel those from the other outcome measures used in the study. Bellack, Hersen, and Lamparski (1979) used a similar method to evaluate heterosexual social interaction. The interaction of college students and opposite-sexed confederates was videotaped after the students were told to wait due to a "slight scheduling mix-up". The measure involved ratings of a series of target behaviors similar to those used by Gutride et al. (1973).

Hollandsworth, Glazeski, and Dressel (1978) used contrived observation to help evaluate a treatment program in which a 30-year-old college graduate with a long history of anxiety induced failures in job interviews, received social skills training. The subject's social-communicative skills were rated by two judges, who, unknown to the subject, assumed the role of shoppers and observed him interacting with customers in his natural work environment before and after training. Both verbal behaviors (fluency of speech, appropriateness of contact) and nonverbal behaviors (eye contact, composure, personal appearance) were evaluated. The authors reported that although they made sure that this measure was administered ethically by having the subject sign a statement consenting to the assessment procedure, debriefing indicated that he was unaware of being observed, and therefore the unobtrusiveness of the measures was maintained.

Kazdin (1979) has pointed out some potential problems that are especially relevant to the use of contrived observation. He notes that the use of observers to assess behavior change may result in "observer drift" or "the tendency of observers to change the manner in which they apply the definitions of behavior problems over time" (p. 27). The use of observers necessitates the assessment of interobserver agreements by external observation. Consequently, another potential problem with the use of observers is that the process of determining interobserver reliability may itself be reactive. According to Kazdin, both these and other problems that arise from the use of observers can be alleviated by training them well and continuously monitoring their behavior directly or by videotape. An additional problem with contrived observations discussed by Kazdin (1979) is that subjects may see through the manipulation designed to elicit the target behavior. This is particularly the case when subjects are sensitized to this possibility through the process of obtaining their permission to use such assessment procedures. Kazdin recommends gathering information to evaluate whether or not the measure was unobtrusive. In the studies of both Frederiksen et al., (1976) and Hollandsworth et al. (1978), debriefing revealed that subjects were apparently unaware of the contrived situations in which their behavior was measured. By reporting such data, the reader's confidence that the measures were unobtrusive will be increased.

Measures with Dependably Low Levels of Reactivity

Most measures require that the subject be aware of being studied, but this knowledge does not invariably have a strong influence on behavior. In fact, if it did, we might use the measure as a change agent. There are some measures that, although being obvious indicators of behavior change to the subject, are nevertheless likely to be minimally influenced by such knowledge. Tests of ability, in which most persons will be motivated to give a good performance, such as intelligence or achievement tests, fall into this category. One manifestation of impairment, in fact,

may be disinterested in giving a good performance. Holtzman (1975) noted an IQ increase of 11 points in one of his patients after psychoanalysis.

The problem of reactivity is most often encountered in a context that opens up the possibility of "faking good" (i.e., of subjects producing better responses than are typical of their performances). Hence, situations that tend to produce optimal performance may be minimally reactive. Nonetheless, reactivity may occur in the direction of a poor performance, and the investigator should be vigilant of those instances when it may be in the best interests of a client to simulate a poor performance (i.e., malingering).

Several investigators have reported the use of perceptual and motor tasks to evaluate changes in patients, usually in populations that are more disturbed. Hybl and Stagner (1952) reported that clients after therapy demonstrated some improvement in "frustration tolerance" as measured by performance on simple motor tasks. Peters and Jones (1951) used performance on two simple motor tasks, Porteus Mazes and Mirror Tracing, to study changes in schizophrenics. They found improvement on these tasks in treatment groups but not controls. Indicators of personality change may be obtained from perceptual-motor tasks. Performance on the Witkin Rod and Frame Test, for example, has been related to self-esteem, active involvement with one's environment, and analytical thinking (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). Coopersmith (1964) found that subjects who highly estimated their personal worth were better able to obtain sensory constancy on a perceptual-motor task. Goldstein and Chotlos (1966) found that alcoholics showed significantly greater field independent performance on the Witkin Rod and Frame test following a 3 month treatment.

Although some physiological measures may be subject to the client's voluntary control, there are others that we can be confident are, when used for research purposes, influenced by the client's awareness of being measured. The Galvanic skin response is one such measure and was used by Hollandsworth et al. (1978) to measure anxiety during pre- and post-training in *in vivo* job interviews after social skills training. Although their methods are likely too expensive and time consuming to follow in most outcome research, Greenberg and Pearlman (1976) used time in REM sleep to predict good response to psychotherapy, elopement, and need for medication in psychiatric patients. Another minimally reactive physiological measure is surface electromyographic recordings from selected facial muscles (Schwartz et al., 1975). Schwartz et al. (1975) argued that EMG measures on facial muscles can reflect the content of thoughts, independently of expression, and reported that responses on this measure paralleled improvement on objective tests.

Trexler and Karst (1972) used a Finger Sweat Print to measure anxiety in their study, which compared the outcome of rational emotive therapy, placebo, and no treatment effects on public speaking anxiety. The Finger Sweat Print measure is described by Droppleman and McNair (1971). It involves a tape band procedure used on the subject's index fingers. Droppleman and McNair reported a high interjudge reliability of this measure ($r = .94$), but Trexler and Karst described the FSP as somewhat unstable based on their test-retest reliability figures. A reduction in speech anxiety based on the FSP was found on some of Trexler and Karst's (1972) analyses, which was consistent with the results of their more reactive outcome indices.

Surveillance of alcohol, cigarette, and drug use is another example of an assessment situation in which the person's knowledge of being studied may influence

the target behavior, at least over the short run. It may be difficult, however, to control such addictive behaviors over any substantial period of time merely to meet the demands of continued assessment. Both alcohol and drug use can be monitored by means of urine samples, which may be obtainable on random or other schedule tending to reduce reactivity Goldstein & Brown, 1970; (Goldstein et al., 1966). Similarly, cigarette use can be brought under surveillance by measuring serum thiocyanate in samples or, even better, by measuring serum thiocyanate in samples of saliva or blood. Drug and alcohol use have been used as partial measures of psychotherapy outcome (Spitzer, Endicott, Fleiss, & Cohen, 1970), and in some cases even cigarette usage might be relevant to psychotherapy outcome. It should not be supposed that surveillance can always be produced without reactivity. We know, for example, of weight-watching groups in which some members use diuretics before weekly weigh-ins to produce an apparent, but spurious, weight loss.

SUMMARY

Almost without exception, the measures used in psychotherapy outcome research are highly reactive. This shortcoming is not to be taken lightly, since Smith, Glass, and Miller (1980) have indicated that reactivity in measurement may be distorting the experimental effect size in these studies. One must be sympathetic to the researcher in that it is next to impossible to design a full-scale psychotherapy outcome study that tests for the effects of reactivity. Further, the measures currently used appear to be the most psychometrically sound means of assessing outcome that are available. The objection voiced here is that the reactive measures are being used alone. Even when researchers combine several measures they are usually all reactive. To remedy this common methodological shortcoming, researchers need to begin to include nonreactive measures routinely in outcome batteries. Such measures can in valuable ways supplement, but not replace, more traditional outcome instruments. Moreover, some measures available are low-cost, require relatively little extra effort, would not require a change in usual experimental designs, and can be individually tailored to clients. Nonreactive measures must, however, be applied with their psychometric and ethical restrictions kept in mind. Potential sources of unobtrusive measures include physical traces, archives, simple observations, and contrived observations. Although this paper has offered several suggestions for nonreactive measures that could be implemented in psychotherapy outcome research, it has not been our intent to catalog the available possibilities. Rather, these examples were presented in the spirit of stimulating more thought and research in this seriously lacking area.

REFERENCES

- American Psychological Association. *Code of ethics of the American Psychological Association*. Washington, D.C.: Author, 1973.
- Argyle, M., & Ingham, R. Gaze, mutual gaze, and proximity. *Semiotica*, 1972, **6**, 32–49.
- Bellack, A. S., Hersen, M., & 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.
- Bland, R. C., & Orn, H. Schizophrenia: Schneider's first rank symptoms and outcome. *British Journal of Psychiatry*, 1980, **137**, 63–68.

- Boucher, J. D., & Ekman, P. Facial areas of emotional information. *Journal of Communication*, 1975, **25**, 21–29.
- Burdock, E. I., & Hardesty, A. S. An outcome index for mental health hospital patients. *Journal of Abnormal and Social Psychology*, 1961, **63**, 666–670.
- Campbell, D. T. The mutual methodological relevance of anthropology and psychology. In F. I. K. Hsu (Ed.), *Psychological anthropology approaches to culture and personality*. Homewood, Ill.: Dorsey Press, 1961, 333–352.
- Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.
- Cannon, W. B. *Bodily changes in pain, hunger, fear and rage*. Boston: Charles T. Branford Co., 1953.
- Cantril, H. *Gauging public opinion*. Princeton: Princeton University Press, 1944.
- Cialdini, R. B., & Baumann, D. J. Littering: A new unobtrusive measure of attitude. *Social Psychological Quarterly*, 1981, **44**, 254–259.
- Cook, T. D., & Campbell, D. T. *Quasi-experimentation; Design and analysis issues for field settings*. Chicago: Rand McNally, 1979.
- Cookerly, J. R. Does marital therapy do any lasting good? *Journal of Marital and Family Therapy*, 1980, **6**, 393–397.
- Coopersmith, S. Relationship between self-esteem and sensory (perceptual) constancy. *Journal of Abnormal and Social Psychology*, 1964, **68**, 217–221.
- Cranach, M. The role of orienting behavior in human interaction. In A. H. Esser (Ed.), *Behavior and environment: The use of space by animals and men*, New York: Plenum Press, 1971.
- Cronbach, L. J. Response sets and test validity. *Educational and Psychological Measurement*, 1946, **6**, 475–494.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. *An MMPI Handbook* (Vol. 1). Minneapolis: University of Minnesota Press, 1960.
- Darwin, C. R. *The expression of the emotions in man and animals*. Chicago: The University of Chicago Press, 1965.
- Dittman, A. T. *Interpersonal messages of emotion*. New York: Springer, 1972.
- Dollard, J., & Mowrer, O. H. A method for measuring tension in written documents. *Journal of Abnormal and Social Psychology*, 1947, **42**, 3–32.
- Droppleman, L. F., & McNair, D. M. An experimental analogue of public speaking. *Journal of Consulting and Clinical Psychology*, 1971, **36**, 91–96.
- Duncan, S. D., & Fiske, D. W. *Face-to-face interaction: Research, Methods, and Theory*. Hillsdale, N.J.: Erlbaum, 1977.
- Dykman, B. M., & Reis, H. T. Personality correlates of classroom seating position. *Journal of Educational Psychology*, 1979, **71**, 346–354.
- Ekman, P. Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation. Vol. 19*. Lincoln: University of Nebraska Press, 1972.
- Ekman, P., & Friesen, W. V. The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1969, **1**, 49–93.
- Ekman, P., & Friesen, W. V. *Facial affect coding system*. Palo Alto, Ca.: Consulting Psychologists Press, 1978.
- Ekman, P., Friesen, W. V. & Tomkins, S. S. Facial affect scoring technique: A first validity study. *Semiotica*, 1971, **3**, 37–58.
- Eysenck, H. J. An exercise in mega-silliness. *American Psychologist*, **33**, 1978, 517.
- Feshbach, S., & Feshbach, N. Influence of the stimulus object upon the complementary and supplementary projection of fear. *Journal of Abnormal and Social Psychology*, 1963, **66**, 498–502.
- Fisher, S., & Greenberg, R. P. *The scientific credibility of Freud's theories and therapy*. New York: Basic Books, 1977.
- Fiske, D. W. The use of significant others in assessing their outcome of psychotherapy. In I. E. Waskow & M. B. Parloff, (Eds.), *Psychotherapy change measures: Reports of the clinical branch outcome measures project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Fiske, D. W. Methodological issues in research on the psychotherapist. In A. S. Gurman & A. M. Ruzin (Eds.), *Effective psychotherapy: A handbook of research*. New York: Pergamon Press, 1977.
- Fiske, D. W., Hunt, H. F., Luborsky, L., Orne, M. T., Parloff, M. B., Reiser, M. F. & Tuma, A. M. The planning of research on effectiveness of psychotherapy. *Archives of General Psychiatry*, 1970, **22**, 22–32.
- Fixsen, D. L., Phillips, E. L., Phillips, E. A., & Wolf, M. M. The teaching-family model of home group

- home treatment. In W. E. Craighead, A. E. Kazdin, & M. J. Mahoney (Eds.), *Behavior modification: Principles, issues and applications*. Boston, Mass.: Houghton Mifflin, 1976.
- Follette, W. T., & Cummings, N. A. Psychiatric services and medical utilization in a prepaid health plan setting. *Medical Care*, 1967, **5**, 25–35.
- Frederiksen, L. W., Jenkins, J. O., Foy, D. W., & Eisler, R. M. Social-skills training to modify abusive verbal outbursts in adults. *Journal of Applied Behavior Analysis*, 1976, **9**, 117–125.
- Friedman, M., Brown, A. E., & Rosenman, R. H. Voice analyses test for detection of Type A behavior pattern. *Journal of the American Medical Association*, 1969, **208**, 828–836.
- Garwick, G., & Brintmall, J. (Eds.). *Proceedings of the second annual Goal Attainment Scaling Conference*. Minneapolis, Minn.: Program Evaluation Resource Center, 1974.
- Goldstein, A., & Brown, B. W. Urine testing schedules in metadone maintenance treatment of heroin addiction. *Journal of the American Medical Association*, 1970, **214**, 311–315.
- Goldstein, G., & Chotlos, J. W. Stability of field independence in chronic alcoholic patients. *Journal of Abnormal Psychology*, 1966, **71**, 420.
- Goldstein, A. P., Heller, K., & Sechrest, L. *Psychotherapy and the psychology of behavior change*. New York: Wiley, 1966.
- Gottschalk, L. A., & Gleser, G. C. *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley: University of California Press, 1969.
- Gottschalk, L. A., Winget, C. N., & Gleser, G. C. *Manual of instructions for using the Gottschalk-Gleser Content Analysis Scales: Anxiety, Hostility and Social Alienation-Personality Disorganization*. Berkeley: University of California Press, 1969.
- Grand, S., Freedman, N., Steingart, I., & Buchwald, C. Communicative behavior in schizophrenia: The relation of adaptive styles to kinetic and linguistic aspects of interview behavior. *Journal of Nervous and Mental Disease*, 1975, **161**, 293–306.
- Green, H. B., & Knapp, R. H. Time judgment, aesthetic preference, and need for achievement. *Journal of Abnormal and Social Psychology*, 1959, **58**, 140–142.
- Greenberg, R., & Pearlman, C. A., Jr., REM sleep and adaption of psychiatric patients: An application of sleep studies. *American Journal of Psychiatry*, 1976, **133**, 1147–1150.
- Gutride, M. E., Goldstein, A. P., & 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.
- Haggard, E. A., & Isaacs, K. S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In L. A. Gottschalk & A. H. Auerbach, (Eds.), *Methods of Research in Psychotherapy*. New York: Appleton, Century, and Crofts, 1966.
- Hare, A. P., Waxler, N., Saslow, G., & Matarazzo, J. D. Simultaneous recording of Bales and Chapple interaction measures during initial psychiatric interviews. *Journal of Consulting Psychology*, 1960, **24**, 193.
- Harper, R. G., Wiens, A. N., & Matarazzo, J. D. *Nonverbal communication: State of the art*. New York: Wiley, 1978.
- Harper, R. G., & Wiens, A. N. Nonverbal behaviors as unobtrusive measures. In L. Sechrest (Ed.), *Unobtrusive measurement today*. San Francisco: Jossey Bass, 1979.
- Hersen, M., Eisler, R. M., & Miller, P. M. An experimental analysis of generalization in assertive training. *Behaviour Research and Therapy*, 1974, **12**, 295–310.
- Hinchliffe, M., Lancashire, M., & Roberts, F. J. Eye-contact and depression: A preliminary report. *British Journal of Psychiatry*, 1970, **117**, 571–572.
- Hinchliffe, M., Lancashire, M., & Roberts, F. J. A study of eye-contact changes in depressed and recovered psychiatric patients. *British Journal of Psychiatry*, 1971, **119**, 213–215.
- Hogarty, G. E. Informant ratings of community adjustment. In I. E. Waskow & M. B. Parloff (Eds.), *Psychotherapy change measures: Reports of the Clinical Branch Outcome Measures Project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Holahan, C. Seating patterns and patient behavior in an experimental dayroom. *Journal of Abnormal Psychology*, 1972, **80**, 115–124.
- Hollandsworth, J. G., Glazeski, R. C., & Dressel, M. E. Use of social-skills training in the treatment of extreme anxiety an deficient verbal skills in the job-interview setting. *Journal of Applied Behavior Analysis*, 1978, **11**, 259–269.
- Holmes, L. D. Tafu: Stability and change in a Samoan village. Reprint No. 7. Wellington, N.Z.: Polynesian Society, 1958.
- Holtzman, P. S. Cognitive and perceptual tests for evaluating psychotherapy outcome. In I. E. Waskow

- & M. B. Parloff (Eds.), *Psychotherapy change measures: Reports of the Clinical Branch Outcome Measures Project* Rockville, Md.: National Institute of Mental Health, 1975.
- Horowitz, L. M., Weckler, D., Sazon, A., Livaudais, J. D., & Boutacoff, L. I. Discomforting talk and speech disruptions. *Journal of Consulting and Clinical Psychology*, 1977, **45**, 1036–1042.
- Hybl, A. R., & Stagner, R. Frustration tolerance in relation to diagnosis and therapy. *Journal of Consulting Psychology*, 1952, **16**, 163–170.
- Johnson, S. M., & Sechrest, L. B. Systematic desensitization in the treatment of test anxiety. *Journal of Consulting and Clinical Psychology*, **32**, 280–286.
- Jourard, S. M. *The transparent self*. New York: Van nostrand Reinhold, 1971.
- Julian, A., & Kilman, D. R. Group treatment of juvenile delinquents: A review of the outcome literature. *International Journal of Group Psychotherapy*, 1979, **29**, 3–37.
- Jurich, A. P., & Jurich, J. A. Correlations among nonverbal expressions of anxiety. *Psychological Reports*, 1974, **34**, 199–204.
- Kane, F. Clothing worn by out-patients to interviews. *Psychiatric Communications*, 1958, **1**, (2).
- Kane, F. Clothing worn by an out-patient: A case study. *Psychiatric Communications*, 1959, **2**, (2).
- Kane, F. The meaning of the form of clothing. *Psychiatric Communications*, 1962, **2**, (2).
- Kanfer, F. H. Report on outcome measures in behavior therapy. In I. E. Waskow & M. B. Parloff (Eds.), *Psychotherapy change measures: Reports of the clinical branch outcome measures project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Karon, B. P., & Vandebos, G. R. Thought disorder in schizophrenia, length of hospitalization, and client status ratings: Validity for the Feldman-Drasgow Visual-Verbal Test. *Journal of Clinical Psychology*, 1974, **30**, 264–266.
- Katz, D. Do interviewers bias poll results? *Public Opinion Quarterly*, 1942, **6**, 248–268.
- Kazdin, A. E. Effects of covert modeling and model reinforcement on assertive behavior. *Journal of Abnormal Psychology*, 1974, **83**, 240–252.
- Kendon, A., & Cook, M. The consistency of gaze patterns in social interaction. *British Journal of Psychology*, 1969, **60**, 481–494.
- Kiesler, D. J. Patient measures based on interviews and other speech samples. In I. E. Waskow and M. B. Parloff, *Psychotherapy change measures: Report of the clinical branch outcome measures project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Knapp, M. L. *Nonverbal communication in human interaction*. New York: Holt, Rinehart, & Winston, 1972.
- Kolb, L. C. *Modern clinical psychiatry*. Philadelphia: W. B. Saunders, 1977.
- Lang, P. J. Psychophysiological analysis in psychotherapy research. In I. E. Waskow & M. B. Parloff, *Psychotherapy change measures: Report of the Clinical Research Branch Outcome Measures Project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Leipold, W. D. Psychological distance in a dyadic interview as a function of introversion-extraversion, anxiety, social desirability and stress. Unpublished doctoral dissertation, University of North Dakota, 1963.
- Lewinsohn, P. Relationship between height of figure drawings and depression in psychiatric patients. *Journal of Consulting Psychology*, 1964, **28**, 380–381.
- Lewis, B. I. Mechanism and management of hyperventilation syndromes. *Biochemical Clinics*, 1964, **4**, 89–96.
- Machover, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C. Thomas, 1949.
- Mahl, G. Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology*, 1956, **53**, 1–15.
- Matarazzo, J. P., Holman, D. C., & Wiens, A. N. A simple measure of interviewer and interviewee speech durations. *Journal of Psychology*, 1967, **66**, 7–14.
- Matarazzo, J. D. & Wiens, A. N. *The interview: Research on its anatomy and structure*. Chicago: Aldine-Atherton, 1972.
- McFall, R. M., & Lillesand, D. B. Behavior rehearsal with modeling and coaching in assertion training. *Journal of Abnormal Psychology*, 1971, **77**, 313–323.
- McFall, R. M., & Marston, A. R. An experimental investigation of behavior rehearsal in assertive training. *Journal of Abnormal Psychology*, 1970, **76**, 295–303.
- Mechanic, D., & Volkart, E. H. Stress, illness behavior and the sick role. *American Sociological Review*, 1961, **26**, 51–58.
- Mischel, W. *Personality and assessment*. New York: John Wiley & Sons, 1968.
- Morris, R. L., Johnson, G. I., Bailey, D. D., & Wiens, A. N. A twenty-four-channel temporal-event digital

- recording system. *Medical Research Engineering*, 1968, 7, 32–37.
- Mowrer, O. H. *Psychotherapy: Theory and research*. O. Hobart Mowrer and 21 contributors. New York: Ronald Press Co., 1953.
- Murray, D. C. Talk, silence and anxiety. *Psychological Bulletin*, 1971, 75, 244–260.
- Palmer, J., & McGuire, F. L. The use of unobtrusive measures in mental health research. *Journal of Consulting and Clinical Psychology*, 1973, 40, 431–436.
- Paul, G. L., & Lentz, R. J. *Psychosocial treatment of chronic mental patients: Milieu versus social-learning programs*. Cambridge, Mass.: Harvard University Press, 1977.
- Pepitone, A., & Gray, D. M. Effect of self-esteem on drawings of the human figure. *Journal of Consulting Psychology*, 1964, 28, 452–455.
- Peters, H. N., & Jones, F. D. Evaluation of group psychotherapy by means of performance tests. *Journal of Consulting Psychology*, 1951, 15, 363–367.
- Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*. Chicago, Ill.: Year book Publication, 1946.
- Rathje, W. L. Trace measures. In L. Sechrest (Ed.), *Unobtrusive measurement today*. San Francisco: Jossey-Bass, 1979.
- Rathje, W. L., & Hughes, W. W. The garbage project as a nonreactive approach: Garbage in . . . Garbage out? In H. W. Sinaiko & L. A. Broedling (Eds.), *Perspectives on attitude assessment: Surveys and their alternatives*. Washington: Smithsonian Institution, 1975.
- Rice, D. G., Abrams, G. M., & Saxman, J. H. Speech and physiological correlates of 'flat' affect. *Archives of General Psychiatry*, 1969, 20, 566–572.
- Riesman, D. Orbits of tolerance interviewers and elites. *Public Opinion Quarterly*, 1956, 20, 49–73.
- Riesman, D., & Ehrlich, J. Age and authority in the interview. *Public Opinion Quarterly*, 1961, 25, 39–56.
- Roback, H. B. & Webersinn, A. L. Size of figure drawings of depressed psychiatric patients. *Journal of Abnormal Psychology*, 1966, 71, 416.
- Robinson, D., & Rohde, S. Two experiments with an anti-semitism poll. *Journal of Abnormal and Social Psychology*, 1946, 41, 136–144.
- Rosenthal, R. *Experimental effects in behavioral research* (Enlarged edition). Halsted/Wiley, 1976.
- Rubenstein, L. Facial expressions: An objective method in the quantitative evaluation of emotional change. *Behavior Research Methods and Instruments*, 1969, 1, 305–306.
- Rutter, D. R. Visual interaction in psychiatric patients: A review. *British Journal of Psychiatry*, 1973, 123, 193–202.
- Sales, S. M. Need for stimulation as a factor in social behavior. *Journal of Personality and Social Psychology*, 1971, 19, 124–134.
- Salzman, L. F., & Harway, N. I. Size of figure drawings of psychotically depressed patients. *Journal of Abnormal Psychology*, 1967, 72, 205–207.
- Scherer, K. R. Randomized splicing: A note on a simple technique for masking speech content. *Journal of Experimental Research in Personality*, 1971, 5, 155–159.
- Scherer, S. E. Proxemic behavior of primary school children as a function of their socioeconomic class and subculture. *Journal of Personality and Social Psychology*, 1974, 29, 800–805.
- Scherwitz, L., Berton, K., & Leventhal, H. Type A assessment and interaction in the behavior pattern interview. *Psychosomatic Medicine*, 1977, 39, 229–240.
- Schucker, B., & Jacobs, Jr., D. R. Assessment of behavioral risk for coronary disease by voice characteristics. *Psychosomatic Medicine*, 1977, 39, 219–228.
- Schwartz, G. E., Fair, P. L. Greenberg, P. S., Mandel, M. R., & Klerman, G. L. Facial expression and depression II: An electromyographic study. *Psychosomatic Medicine*, 1975, 37, 81–82.
- Scott, W. H. Measures of perceptual and cognitive tendencies in psychotherapy research. In I. E. Wasikow & M. B. Parloff, *Psychotherapy change measures: Report of the Clinical Research Branch Outcome Measure Project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Sechrest, L. Testing, measuring, and assessing people. In E. Borgatta & W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally, 1968.
- Sechrest, L. Another look at unobtrusive measures: An alternative to what? In W. Sinaiko & L. Broedling (Eds.), *Perspectives on attitude assessment: Surveys and their alternatives*. Washington, D.C.: Smithsonian Institution, 1975.
- Sechrest, L., & Flores, L. The occurrence of a nervous mannerism in two cultures. *Asian Studies*, 1971, 9, 55–63.
- Sechrest, L., & Phillips, M. Unobtrusive measures: An overview. In L. Sechrest (Ed.), *Unobtrusive measures*

- today. *New directions in methodology for behavioral science* (No. 1). San Francisco: Jossey-Boss, 1979.
- Sechrest, L., & Phillips, M. Unobtrusive measures: An overview. In L. Sechrest (Ed.), *Unobtrusive Measurement Today*, San Francisco: Jossey-Bass, 1979.
- Simons, L. S., Wade, T. C., Morton, T. L., & McSharry, D. M. Treatment outcome and follow-up evaluation based on client case records in a mental health center. *Journal of Consulting and Clinical Psychology*, 1978, **46**, 246–251.
- Sletto, R. F. *A construction of personality scales by the criterion of internal consistency*. Hanover, N.H.: Sociological Press, 1937.
- Smith, M. L., Glass, G. V., & Miller, T. I. *The benefits of psychotherapy*. Baltimore: The Johns Hopkins University Press, 1980.
- Sommer, R. Studies in personal space. *Sociometry*, 1959, **22**, 247–260.
- Sommer, R. Further studies of small group ecology. *Sociometry*, 1965, **28**, 337–348.
- Sommer, R. Classroom ecology. *Journal of Applied and Behavioral Science*, 1967, **3**, 489–503. (a)
- Sommer, R. Small group ecology. *Psychological Bulletin*. 1967, **67**, 145–152. (b)
- Spitzer, R., Endicott, J., Fleiss, J., & Cohen, J. The psychiatric status schedule. *Archives of General Psychiatry*, **23**, 1970, 41–55.
- Starkweather, J. A. Vocal behavior as an information channel of speaker status. In K. Salzinger & S. Salzinger (Eds.), *Research in verbal behavior and some neurophysiological implications*. New York: Academic Press, 1967.
- Strupp, H. H. A reformulation of the dynamics of a therapist's contribution. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research*. New York: Pergamon Press, 1977.
- Trexler, L. D., & Karst, T. O. Rational-emotive therapy, placebo, and nontreatment effects on public speaking anxiety. *Journal of Abnormal Psychology*, 1972, **79**, 60–67.
- Trujillo, N. P., & Warthin, T. A. The frowning sign multiple forehead furrows in peptic ulcer. *The Journal of the American Medical Association*, 1968, **205**, 470.
- Waskow, I. E., & Parloff, M. B. *Psychotherapy changes measures: Report of the Clinical Research Branch Outcome Measures Project*. Rockville, Md.: National Institute of Mental Health, 1975.
- Waxer, P. Nonverbal cues for depth of depression: Set versus no set. *Journal of Clinical and Consulting Psychology*, 1976, **44**, 493.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B., *Nonreactive Measures in the Social Sciences*. Boston: Houghton Mifflin Company, 1981.
- Wellish, D. K., & Ro-Trock, G. K. A three-year follow-up of family therapy. *International Journal of Family Therapy*, 1980, **2**, 169–175.
- Wiggins, J. S. *Personality and prediction: Principles of personality assessment*. Reading, MA.: Addison-Wesley, 1973.
- Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenogch, D. R., & Karp, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., & Wapner, S. *Personality through perception: An experimental and clinical study*. New York: Harper, 1954.
- Wolpe, J., & Lang, P. J. A fear survey schedule for use in behavior therapy. *Behavior Research and Therapy*, 1964, **2**, 27–30.