

successfully satisfying the needs and challenges faced by behavioral therapists. Openness to diversity of the kind illustrated by this Special Series can only serve the behavior therapy field and those who practice within it. However, implementation of many of the suggestions presented here will take the joint efforts of researchers and practitioners to empirically establish the extent to which concepts advanced in these papers can be empirically validated and successfully implemented in a behavioral practice.

## References

- Achenbach, T. M. (1993). Implications of multiaxial empirically based assessment for behavior therapy with children. *Behavior Therapy, 24*, 91-116.
- DuPaul, G. J., & Barkley, R. A. (1993). Behavioral contributions to pharmacotherapy: The utility of behavioral methodology in medication treatment of children with Attention Deficit Hyperactivity Disorder. *Behavior Therapy, 24*, 47-65.
- Forehand, R., & Wiersma, M. (1993). The role of developmental factors in planning behavioral interventions for children: Disruptive behavior as an example. *Behavior Therapy, 24*, 117-141.
- Kazdin A. E. (1993). Evaluation in clinical practice: Clinically sensitive and systematic methods of treatment delivery. *Behavior Therapy, 24*, 11-45.
- Mash, E. J., & Hunsley, J. (1993). Behavior therapy and Managed Mental Health Care: Integrating effectiveness and economics in mental health practice. *Behavior Therapy, 24*, 67-90.

RECEIVED: September 2, 1992

FINAL ACCEPTANCE: September 30, 1992

## Evaluation in Clinical Practice: Clinically Sensitive and Systematic Methods of Treatment Delivery

ALAN E. KAZDIN

Yale University

The present paper makes the case for systematic assessment and evaluation in clinical practice. The purpose of systematic evaluation is to enhance client care and to improve the basis for drawing inferences about treatment and therapeutic change. Requirements of a methodology for clinical practice are outlined that begin with the priority of clinical care and patient improvement. Assessment, design, and data evaluation strategies as well as the processes they reflect in the context of clinical work are outlined. Ongoing assessment of client progress over the course of treatment is emphasized as the central feature of systematic evaluation. Implications of assessment for setting goals, specifying means, and making decisions to augment patient care are also presented and illustrated. Further work required to develop a methodology of clinical practice, and common objections to systematic evaluation in clinical work, and professional issues that affect adoption of systematic evaluation are also presented.

There is a widely recognized hiatus between clinical research and clinical practice (Barlow, Hayes, & Nelson, 1984; Stricker & Keisner, 1985). The hiatus is captured by many discussions of scientists and practitioners and the effort to combine the best of both worlds in clinical training, research, and practice (e.g., Stricker, 1992). Although the hiatus often is lamented, it is also viewed as understandable given the different priorities, contexts, and demands of clinical practice and clinical research. In clinical practice, the goals are to manage and to treat the individual case. Psychotherapy is individually tailored to the characteristics of the patient or client based on his or her special circumstances, presenting complaints, and personal and interpersonal resources.<sup>1</sup> Therapy

Completion of this paper was facilitated greatly by the support of a Research Scientist Award (MH00353) and a grant (MH35408) from the National Institute of Mental Health. The support of these projects is gratefully acknowledged. Correspondence concerning this paper should be directed to: Alan E. Kazdin, Department of Psychology, Yale University, P.O. Box 11A Yale Station, New Haven, CT 06520-7447.

<sup>1</sup> For present purposes, psychotherapy refers broadly to psychosocial interventions designed to influence affect, cognitions, and behavior. The full range of approaches (e.g., insight-oriented, cognitive-behavioral, family) are encompassed in the present use of the term psychotherapy. Related, the terms client and patient are used to refer to persons (e.g., children, adolescents, adults) seen in therapy. The terms are used interchangeably without implied endorsement of a particular model of patienthood. For purposes of presentation, the comments focus on individual psychotherapy. However, the thrust of the points applies to other foci (e.g., patients seen as couples and in families).

is provided in a flexible way; the focus, techniques, and processes may change as needed. In contrast, clinical research has as its primary goal the isolation of variables to evaluate theory or to test specific hypotheses. Several practices related to the experimental design (e.g., random assignment, no-treatment control groups) and treatment delivery (e.g., number, content, and foci of the treatment sessions) are prescribed in a fixed (controlled) fashion.

A central concern has been that clinical practice may not be informed by psychological research. In fact, clinical practice has drawn on basic research quite frequently. Extrapolations of alternative models of learning to psychotherapy are prominent examples of efforts to translate and to apply psychological theory and laboratory paradigms to therapy. Indeed, many instances can be identified historically in which the investigators who conducted basic research (e.g., Bechterev, Pavlov) were directly involved in extrapolations to clinical work (see Kazdin, 1978). Conversely, and of course familiar to readers of the present journal, professionals involved in practice frequently have drawn upon laboratory research paradigms to develop specific treatment techniques (e.g., Cautela, 1967; Salter, 1949; Wolpe, 1958).

Although clinical practice has drawn on psychological theory and laboratory paradigms, there are specific domains that have not penetrated clinical practice as much as one would expect (Garfield, 1980). As the major case in point, research on psychotherapy is often considered to have had little impact on clinical practice. Indeed, occasionally eminent researcher-practitioners have stated that their own research has not greatly affected their practice (see Bergin & Strupp, 1972). Surveys of practitioners have helped to identify self-reported reasons why diverse facets of research seem alien to clinical practice (e.g., Cohen, Sargent, & Sechrest, 1986; Haynes, Lemsky, & Sexton-Radek, 1987). The ways in which research is conducted, the focus on groups (means), the isolation of treatment techniques, and the use of nonreferred cases are a few of the many reasons why research is not viewed as particularly relevant to clinical practice (see Barlow et al., 1984; Bergin & Strupp, 1972). For these reasons, the limited impact of psychotherapy research on practice is understandable.

The connections between substantive findings of psychology and clinical psychology and clinical practice are critically important in their own right. A concern of arguably greater significance is the methodological tenets and rules of knowledge that separate clinical practice and clinical research. In clinical practice the professional is called on to assess and to evaluate the nature and scope of client dysfunction, the resources within the client and his or her life relevant to treatment, the likelihood that various treatments will be beneficial, and the impact of treatment as therapy unfolds, among other things. Impressions, intuition, and seasoned experience are major tools relied on for the assessment before and during treatment. Clinical judgment decides what the client "needs," how treatment is progressing, whether therapeutic change has been achieved, and so on. In clinical practice, the methods of assessment and evaluation emphasize therapist judgment.

In research, the bases for assessment and evaluation are generally operational and replicable. Key features of research are specification of variables, operationalization of constructs through systematic assessment, and formal

(specifiable) methods of data evaluation. Certainly one should not view scientific tenets and practices with unconditional positive regard. Epistemological tenets of scientific research and many specific practices derived from them (e.g., hypothesis testing, statistical tests) themselves are laced with subjectivity and can be easily challenged (see Kazdin, 1992a, 1992b). At the same time, the weaknesses of scientific methods are not an argument for abandoning systematic evaluation in clinical work.

Clinical work and the methods of procuring "data" have their own problems. In fact, the accumulating verdict regarding judgmental biases in decision making in general and, more specifically, within the context of clinical work has not been kind. Research on causal attributions and expectations of clinicians, illusory correlation, clinical prediction, overconfidence in inferences, and biases in information processing have emphasized the limits of perceptions (see Kanfer & Schefft, 1988; Newman, 1983; Nezu & Nezu, 1989). If there is any body of research to guide clinical practice, it might be the corpus related to judgmental and perceptual biases. The collection of systematic information in the context of treatment does not erase bias. Yet, systematic evaluation can overcome some of the limitations of judgment. Sources of bias can be placed in an empirical arena, subjected to research, and incorporated into improved methods. Also, systematic evaluation provides the opportunity for the codification of accumulated experience to better inform treatment decisions.

The problem of judgmental biases is a strong basis on which to argue for more systematic assessment and evaluation in the context of therapy. However, there is a related, more persuasive, and clinically relevant reason to make the case. Specifically, the priority of clinical work is to improve the functioning of a particular client and usually to do so here and now. The questions for therapy practice are: How can we help this patient; with what strategies and procedures; how much progress is being made; in what ways has this client changed; and what further changes, if any, are needed or desirable? Evaluation in clinical work is designed to assess functioning of the client; to decide what interventions are suitable; to apply the interventions; to examine their impact on the goals of treatment; and to make decisions and changes as needed regarding the continuation, discontinuation, and alteration of treatment. The priority of the individual client is not an argument against systematic assessment and evaluation. Indeed, given the importance of ensuring benefits to the client, systematic assessment and evaluation *a fortiori* are very important. It is precisely because the individual (or client such as group or family) is precious in clinical work that we wish to maximize the quality of the information that we use to draw inferences and to make decisions regarding the therapeutic course.

The purpose of the present paper is to make the case for systematic assessment and evaluation in clinical practice. A methodology of clinical practice is outlined that is designed to serve as the basis for drawing clinically important and useful inferences. Components of the methodology, underlying thought processes, and specific practices encompassed by systematic evaluation in clinical practice are discussed and illustrated. Emphasis is accorded the special role of systematic assessment of ongoing treatment progress and

the implications of such assessment for identifying treatment goals and means. Further work required to develop assessment strategies for clinical work, design options, and data evaluation methods are discussed. Finally, common objections to systematic evaluation in clinical practice and professional issues that affect adoption of systematic evaluation are also presented.

### Toward a Methodology for Clinical Practice *Critical Components for Clinical Methods*

Interest in methods to study the individual are by no means new. Historical tracings of this interest often emphasize the work of Allport (1937), who drew the distinction between nomothetic and idiographic approaches. The *nomothetic* approach refers to the study of groups and an effort to identify laws that characterize people in general. The *idiographic* approach focused on the study of individuals, particularly their unique characteristics. Allport emphasized that these approaches generate quite different types of research and differing findings. The systematic study of the individual has a history extending to several areas of psychology as traced (see Barlow & Hersen, 1984; Kazdin, 1982) and illustrated elsewhere (Davidson & Costello, 1969; Kazdin & Tuma, 1982; Yin, 1984).

In the context of clinical psychology, evaluation of individual cases has a special role. The very definition of clinical psychology frequently includes as the appropriate domain the study of the individual (e.g., Korchin, 1976; Watson, 1951). Also, the case study has played a pivotal role in generating hypotheses, supporting specific theoretical perspectives, and providing opportunities for studying rare phenomena in clinical work. Evaluation of the individual case in clinical work has relied almost exclusively on anecdotal methods. There have been efforts to devise and to develop more rigorous methods of evaluation for clinical work. As a notable example, Shapiro (1957, 1964) advocated and illustrated the importance of systematic and individualized assessment and application of research methods to augment clinical care of individual patients. Among professionals involved in clinical work, there exists a keen interest in integrating research methods and clinical work (Haynes et al., 1987). Even so, there is no articulated, formalized, and disseminable methodology for evaluating the single case in day-to-day clinical practice.

An important point of departure for developing a methodology for clinical practice is to identify the central requirements. Clearly, the first requirement of a methodology for clinical practice is recognition of the priority of clinical care. Unequivocally, methodological practices are to be used toward the benefit of the individual patient. Second, assessment and evaluation practices and other methodological requirements must be in harmony with exigencies of the clinical situation. From the standpoint of research, characteristics of clinical situations in which treatment is delivered invariably have been conceived as "constraints." This notion has unfortunate implications, suggesting that evaluation cannot be achieved or will be sufficiently flawed as to not be very worthwhile. Yet, we must begin with clinical situations and develop systematic assessment and evaluation with these characteristics in mind.

In addition to clinical priorities and exigencies, the methodology itself has special features. *Methodology* refers generally to the principles, practices, and procedures that permit one to draw valid inferences. In clinical work, we are interested in conclusions about patient change, procedures to effect these changes, processes leading to or thwarting change, and so on. Methodology can be conceived as encompassing three general components: assessment, design, and data evaluation. *Assessment* refers to the systematic collection of data that permits one to describe client progress. *Design* refers to ways of arranging the presentation or delivery of the intervention to facilitate inferences. *Evaluation* refers to the ways in which the assessment data are examined to draw inferences. The components work together to draw inferences and to rule out various factors that can compete with drawing conclusions.

Although one can emphasize practices of assessment, design, and evaluation, methodology is also a way of thinking about phenomena and systematizing the information we obtain to draw inferences. The thought processes reflect concerns about ways of operationalizing critical constructs, posing hypotheses about interventions and processes leading to change, and testing assumptions about interventions and their impact. The thought processes also entail bearing in mind the various threats to drawing inferences (i.e., those influences that may interfere with having the desired impact or with drawing conclusions about the agent of change). A methodology for clinical practice requires that these thought processes are retrained but in harmony with clinical work.

Assessment, design, and evaluation are not alien to clinical practice. Invariably, practitioners are always drawing inferences, actively or passively making decisions on what they perceive, and so on. A methodology for clinical practice does not introduce assessment, design, and evaluation. Rather, it brings these practices into harmony with tenets of science (e.g., testing hypotheses, operationalizing critical concepts, fostering replication). The special feature is to utilize evaluation concepts and practices to advance the therapeutic progress of individual patients. To that end, the primary goal of a methodology for clinical practice is to assess, evaluate, and demonstrate change. Assessment and evaluation are key features. In research, primary attention is accorded demonstration of a causal relation between variables and isolating the impact of individual variables. In clinical work, the focus on identifying causal agents is likely to be less central.

### *Leads from Single-case Research Methods*

In contemporary research, there has been a continued interest in developing methods of evaluating the individual case. The most systematic and visible effort along these lines has emerged from behavioral research. The development of behavioral research, particularly behavior analysis, accounts for the proliferation and expansion of single-case research designs. As is well known, the work of Skinner is responsible for the development of behavior analysis, a position that at once encompasses a substantive view about behavior and its development (so-called radical behaviorism) and a methodological approach (single-case research) (see Kazdin, 1978). In Skinner's work (1938, 1957), operant

conditioning and single-case research methods were intertwined. The methodological approach included the study of the individual subject (organism), continuous assessment of ongoing behavior, measurement of overt behavior, and evaluation of different experimental manipulations within the same subject(s) over time (Sidman, 1960).

Single-case experimental designs have been advanced as one way to introduce systematic assessment and evaluation in clinical practice. The unique feature of these designs is the capacity to conduct experimental investigations with the individual case (e.g., Barlow & Hersen, 1984; Kazdin, 1982). The designs permit assessment of therapeutic change and "causal" inferences about the basis of the change. Single-case research methodology includes essential components mentioned earlier, namely, assessment, design, and evaluation strategies, all of which can be mobilized to examine the individual case.

*Assessment.* Of all features of single-case methods for use in clinical practice, the ongoing assessment of the individual (or group) is most fundamental. Ongoing or continuous assessment refers to reliance on *repeated observations of performance over time*. The client's performance is measured on several occasions, usually before the intervention is applied and continuously over the period while the intervention is in effect. Observations obtained before treatment (referred to as *baseline*) are designed to *describe* the existing level of performance and to *predict* the likely level of performance for the immediate future if an intervention is not provided. The prediction is achieved by *projecting or extrapolating* into the future a continuation of baseline performance. Evaluation of changes over the course of different phases (baseline, intervention) involves comparing the data patterns within and across phases.

Apart from continuous or ongoing assessment, measures are usually developed to individualize the focus for the person seen in treatment. To convey this feature more sharply, the distinction has been made between nomothetic and idiographic behavioral assessment strategies (Cone, 1988). Nomothetic assessment consists of more familiar measures developed for the study of groups. Test development (e.g., Wechsler Intelligence Scales, Minnesota Multiphasic Personality Inventory) focuses on a variety of types of validity, for example, to predict or correlate group performance with nontest criteria. Similarly, test standardization focuses on normative data as a basis on which to evaluate performance. In contrast, idiographic assessment is individualized, based on the particular clinical dysfunction, context, and circumstances of the individual patient. Traditional psychometric properties and methods of scale development may not apply because of the need to individualize the measure and to assess performance continuously over time.

For a methodology of clinical practice, systematic and ongoing assessment is the most pivotal feature. It is in the assessment of change that judgmental biases and subjectivity are likely to loom large. The initial interest of treatment is whether there is a reliable change. There is no substitute for systematic assessment. Hence, assessment in some ongoing way will be pivotal to the methodology of clinical practice. Also, emergent assessment strategies with the single case point to the need to develop measures specifically suited to that purpose, a topic to which we shall return.

*Experimental arrangements.* The manner in which the intervention is presented to the client can be arranged in many different ways. The arrangement of conditions, referred to as research design, is intended to permit valid inferences about the intervention or manipulation of interest. In single-case research, inferences are usually made about the effects of the intervention by comparing different conditions presented to the same subject over time. Repeated assessment provides the data to permit comparisons over time for the same subject. There are a number of experimental designs and design variations, each of which evaluates the extent to which the intervention, rather than extraneous influences, is responsible for change (see Barlow & Hersen, 1984; Kazdin, 1982).

In developing a methodology for clinical practice, we wish to be particularly careful about selecting designs. It is the arrangement of conditions (e.g., withholding, delaying, or presenting treatment for a predetermined period) that has been singularly the most incompatible feature with clinical work. We would like to know about what causes change and arrangements of intervention conditions can be important. A methodology for practice must ensure that these features are deployed only if they can benefit the individual client. We want either to include design options in a methodology of clinical practice or to achieve to the extent possible the inferential security that such designs provide in some other way.

*Data evaluation.* In group research, the differences between groups are evaluated statistically. Statistical evaluation reflects an effort designed to separate "chance" from veridical effects and to determine whether the differences between conditions (groups) are reliable. Single-case data can be evaluated statistically and a number of techniques are available (e.g., Gottman, 1981; Kazdin, 1982). As a matter of philosophy underlying single-case designs and their approach toward intervention, the evaluation usually is nonstatistical (Baer, 1977; Parsonson & Baer, 1978). Evaluation of data nonstatistically has the same goal as statistical analysis, namely, to identify if the effects are consistent, reliable, and unlikely to have resulted from chance fluctuations. The method is sometimes referred to as *visual inspection*, which, unfortunately, suggests that the method is particularly subjective and unusually vulnerable to biases that statistics were designed to overcome.<sup>2</sup> Actually, nonstatistical data evaluation depends on multiple characteristics of the pattern in the data to determine the reliability of change (see Kazdin, 1992a). These data characteristics include changes in mean level of performance across phases, shifts in level of performance when treatment is initially introduced or altered, changes in slope or trend lines across different phases in the single-case design, and the latency of change once a shift in treatment is made. Such characteristics yield many different data patterns; when the characteristics act in concert, rather clear inferences can be drawn.

In a methodology of clinical practice, we wish to evaluate the reliability

<sup>2</sup> There is acknowledged subjectivity in both statistical and nonstatistical methods of data evaluation (see Berger & Berry, 1988; Kazdin, 1992a).

of the change as well, i.e., whether the change is beyond expected fluctuations in performance, chance, or some "artifact." At the same time, clinical work has a special interest in clinically important changes that are not necessarily addressed by statistical and nonstatistical approaches central to group and single-case designs. An advantage of single-case designs has been attention to descriptive features of the data and how these can be used. In a methodology for clinical practice, guidelines and evaluation will be required to assess progress and to make decisions about treatment.

*Limits of the designs for clinical work.* Single-case designs have been proposed for use in clinical work because of the focus on individuals and because some of the designs seem compatible with implementing interventions and evaluating treatment progress (see Alter & Evens, 1990; Barlow et al., 1984; Bloom & Fischer, 1982; Jayaratne & Levy, 1979; Kratochwill & Piersel, 1983). The designs have been used in several clinical demonstrations of the effectiveness of alternative treatments. Clearly, when the designs are feasible and compatible with treatment, their use is to be encouraged because of the strong bases they provide for inferences about treatment.

Widespread use of the designs, at least as they are currently proposed, may not be feasible for clinical practice. First, the designs make very special demands on the administration of treatment. Typically, baseline (no treatment) and treatment phases are alternated or treatment is introduced to different behaviors or domains of functioning in a way that withholds or delays the intervention. The demands of the design and of clinical goals with the individual patient can easily come into conflict. Efforts have been made to identify selected elements that might be used in various forms (e.g., Hayes, 1981) and quasi-experimental versions of the designs (Kazdin, 1981) to make adoption of the designs more compatible with clinical work. Even so, experimental arrangements are not a high priority for clinical work, because isolating intervention effects and demonstrating causal relations are not seen as advancing care of the individual client.

Second and related, single-case research has focused on behavioral assessment, a complex topic in its own right (see Bellack & Hersen, 1988; Nelson & Hayes, 1986). The usual focus of behavioral assessment is on overt behaviors that serve as the targets for treatment and/or as criteria on which to evaluate progress. For a methodology of clinical practice, we would want to accommodate a broad range of assessment domains and methods, including but extending well beyond overt behavior, given the diverse conceptual views and goals of treatment.

Third, single-case designs have been used most often to evaluate behavioral interventions. There is no necessary conceptual link between the single-case methodology and behavioral interventions. Indeed, child psychoanalysis has been carefully evaluated in a single-case (e.g., multiple baseline) design to test critical concepts underlying the target focus as well as the effectiveness of treatment (see Fonagy & Moran, 1990). For a methodology of clinical practice, it is important to underscore that the design features need not be restricted to a particular conceptual model of dysfunction, treatment approach, or assessment method. The task is to identify critical elements of single-case de-

signs that can be used to promote client care and empirical evaluation of treatment within clinical practice.

### General Comments

Single-case design methodology has critical components that provide underpinnings for a methodology of clinical practice. This is obvious when compared to the more traditional group research methods. For example, commonly used design practices of group research (e.g., no-treatment and waiting-list control groups, random assignment to conditions) do not apply to clinical evaluation with the individual case seen in clinical practice. Although group research methods do not seem relevant to the methodology of studying the individual, they ought not to be discounted altogether. A method for clinical practice can draw from many practices that have flourished in traditional group designs. Examples include the development of treatment manuals to codify intervention procedures, evaluation of clinical significance of therapeutic change, and development of a broad set of assessment devices on which to draw for clinical evaluation (see Kazdin, 1992a). In developing a methodology for clinical practice, advances from group and single-case approaches can provide useful leads.

### Systematic Assessment and Evaluation in Clinical Practice

As mentioned previously, the questions for clinical work are: How can we help this patient; with what strategies and procedures; how much progress is being made; in what ways has this client changed; and what further changes, if any, are needed or desirable? To address these questions, we are interested in measuring change, making decisions (e.g., to alter, continue, terminate) about treatment, drawing connections between treatment (e.g., experiences of the client, procedures) and therapeutic change, and examining the influences of nontreatment factors (e.g., changes in support system, living conditions, employment) on client functioning.

A fully developed methodology for clinical practice would involve elaboration of assessment, design, and data evaluation strategies. Each is a weighty topic in its own right to address assumptions, concrete practices, and interpretative issues. For present purposes, emphasis here is on assessment, because this is conceptually and logically prior to design and data evaluation methods. Systematic ongoing assessment can improve our understanding of relations between treatment and change and provide information for the immediate benefit of the client. Introducing systematic assessment to clinical practice is not merely the addition of a few measures to supplement clinical judgment. The very notion of systematic assessment requires critical thoughts about the therapy and the patient at multiple levels.

### Specifying and Assessing Treatment Goals

Critical to assessment and evaluation is the explicit identification of treatment goals. The goals refer initially to the objectives for the individual client



(e.g., stress reduction, increase in social activities, greater reciprocal communication among family members) that treatment is designed to address. Therapy can have many different goals and, indeed, in clinical work these are tailored to individual clients. For example, a goal may be to influence how the client feels (affect), thinks (cognition), and acts (behavior). Each broad category consists of many different foci and assessment methods. Affect might pertain to how one feels about others, oneself, or another aspect of one's life. The goals may extend beyond the individual to include relationships with others (e.g., spouse, child, peers) or functioning of systems (e.g., family, group). There may be multiple goals as a function of the problems or therapeutic foci as provided by the presenting problem. Identifying the goals is a prerequisite to the selection of measures.

Prioritizing the goals is important as well. There may well be several interrelated areas to which the treatment is directed. They ought to be prioritized in a way that permits initial assessment as well as treatment decisions. It is unlikely that treatment will affect all areas equally well or even in the same way or direction (e.g., improvement). Assessment of treatment outcome often shows that different measures can yield different conclusions (Lambert, 1983; Kazdin, 1992a). Hence, it is useful to identify the initial salient goals and to select measures to assess progress toward them. The selection of one or a few initial goals does not mean that this is the only focus of treatment. Rather, it means that assessment will sample critically relevant facets of client functioning on which to examine progress and treatment impact. The goals may vary over time, based on priorities identified with the patient and progress in treatment. For example, subjective distress may serve as the initial impetus. The focus may shift toward less immediate but no less important domains such as symptoms, coping styles, and ways of relating to other individuals. The critical point is that specifying goals does not mean that these goals become a destination toward which the course of treatment is irrevocably plotted. The goals of therapy are often multiple, and making them explicit is an initial as well as an ongoing step.

Identifying the goals refers to the domains or foci and areas of assessment. One might wish to specify the magnitude of change in relation to the goals, that is, the amount of change (degree of improvement) or absolute level of functioning. Goals and amount of change may be feasible or of interest in a given problem area (e.g., weight control). However, for many foci of treatment (e.g., coping skills, stress, marital satisfaction), the amount of improvement that is suitable or important may be individualized based on special circumstances (e.g., presence of a chronic stressor). For present purposes, specification and operationalization of treatment goals are central.

As one example of how goals can be specified and assessed, consider Goal Attainment Scaling (e.g., Kiresuk & Garwick, 1979). This method of assessment consists of a flexible way to identify and to assess individually tailored goals. Goals are identified explicitly for treatment, usually in conjunction with the client. For each treatment goal, a scale is selected with a graded series of outcomes. The outcomes range from least to most favorable. The number of goals (domains), the specific anchor points for each one, and the number points (e.g., 1-5 point scales) for each goal can vary with the individual. The central

feature illustrated by Goal Attainment Scaling is the use of individualized assessment. Other methods can accomplish similar goals, as illustrated later.

### *Specifying and Assessing Procedures, Processes, or Means*

In addition to making the goals explicit, it is also important to specify the means of achieving the goals. The means refer to the strategies, methods, techniques, and other planned interventions that the therapist uses. The variations may well be broad, as suggested by the very large number of therapy techniques that are recognized in child, adolescent, and adult psychotherapy and the multiple combinations often used in clinical practice. Typically, one of two general types of means is relevant to various therapies. First, the means of achieving therapeutic change may refer to the procedures used in treatment; that is, to what the therapist does, what the client is asked to do, and so on. Second, the means of achieving change may refer to emergent processes or relationship issues (e.g., experiencing emotions, developing an alliance). Specifying procedures or processes is not an end in itself. The primary goal is to utilize the information to benefit the client, based on how well treatment was implemented and the ends that were achieved. The notion of treatment integrity, used frequently in therapy research, is particularly important in this regard (see Kazdin, 1992a). Treatment integrity refers to the extent to which the treatment was carried out as intended. In psychotherapy outcome studies, interpretation of the results has been greatly aided by assessment of treatment integrity. In some cases, for example, few or no differences between alternative treatments could be explained, because treatments were not implemented as intended or because persons received treatment to which they were not assigned (e.g., Feldman, Caplinger, & Wodarski, 1983). It is possible that the treatments would have differed in outcome if they had been implemented as intended.

In the context of clinical work, ongoing assessment of client progress (change on the measures related to the goals) may reveal that there is no therapeutic change. Assessment of procedures or processes that the therapist believes are important may provide useful information regarding how to proceed. The information may reveal that treatment procedures (e.g., addressing certain topics, experiencing special types of insight) were not implemented very well or processes within sessions (e.g., developing an alliance, dealing with a particular conflict) were not achieved. Hence, it is reasonable to try different strategies to alter these processes. Alternatively, the assessment of treatment processes or means may reveal that the processes have been evoked fairly well but no therapeutic changes are evident. Here a different approach to the treatment might be reasonable to try. In general, assurances are needed that the procedures were tried or that the processes identified by the therapist were achieved. Making the procedures or means explicit in relation to the goals is an initial step toward this end. As with the goals of treatment, specification of the means to obtain them is an initial step toward developing assessment procedures.

### *Selecting Measures*

Specification of the goals of treatment identifies the constructs of interest to the therapist and client. The next step is to operationalize the constructs

by noting the specific measure(s) that will be used. Selecting the measure requires decisions about the source of information (e.g., self, spouse, parent, clinician report) and assessment method (e.g., "objective" measures of personality or psychopathology, client diaries, card sorts, projective methods, interviews, direct observation, and others). In principle, the actual instruments or measures that can be used include the full array of psychological instruments that have been developed.

It is possible that standardized instruments could be used in unmodified form. For example, unmodified subscales from the MMPI-2 (Butcher, 1990) might serve as a viable candidate for clinical use. In other cases, standardized instruments may need to be modified in terms of wording and format. For example, salient items from a large symptom list, from a standardized instrument, or from an open-ended clinical interview might be identified and agreed on as reflecting the issues to be the focus of treatment or the outcome to be used to evaluate the focus. Several items to represent individual domains can be selected; each item within a domain can be rated for severity of dysfunction to provide an individualized measure (e.g., see Shapiro, 1964). Finally, it is likely that many measures will need to be developed that are individually tailored to the client and the treatment. In such cases, familiarity with scale development and available instruments from which to draw will be useful. A large number of measures useful in clinical work, based on a variety of different assessment methods and sources of information, have been identified elsewhere (see Alter & Evens, 1990; Bloom & Fischer, 1982; Nelson, 1981).

The clinical focus may well dictate both the construct and the appropriate modality of assessment. For example, feelings of hopelessness might well be operationalized in a variety of different ways. Patient perceptions on a measure that assesses hopelessness (i.e., Beck & Steer, 1988) would be quite reasonable and perhaps particularly close to the treatment focus. Of course, we have learned from research that there are dangers in single operationism, i.e., in defining a construct by one measure. Clinical work might utilize more than one measure. For example, the client might complete a measure of hopelessness at the end of each therapy session. Perhaps every 2 weeks a spouse or significant other could be phoned to answer questions about the frequency with which the client engaged in pleasant activities or to respond to an adjective checklist to characterize the client's affect. Yet even if systematic evaluation were restricted to a single measurement operation, this would improve upon the information base usually available in clinical practice.

Apart from measurement of client functioning, measurement of the treatment means or processes is important as well. The specific type of treatment and putative processes or features leading to change dictate the assessment focus. The therapist proposes (hypothesizes) that specific means are central to therapeutic change. If these can vary with treatment administration, assessment is likely to be important. The reason is that, as part of decision making, we wish to know whether the components or processes of treatment were reasonably delivered and tested. As with the client assessment, the task is to identify critical constructs and then to operationalize these in a way that permits assessment.

The use of measures for ongoing continuous assessment in clinical work is a departure from the usual use of measures and scale development (Cone, 1988; Shapiro, 1957). There are formidable issues and problems here because existing measures cannot simply be used in very different ways and be assumed to assess the constructs of interest, a topic to which we shall return. At this point, it is important to note that the research base will be required to develop and to evaluate assessment strategies for clinical use. Systematic evaluation need not be placed on hold until such research is completed, because there are assessment strategies already available for repeated assessment, for individualized use, and so on.

#### *Assessment Occasions*

Assessment in therapy can be completed at different points in time, each with different purposes. *Discrete (one-shot) assessment before treatment* (pretreatment assessment) is the most common point of assessment. Clinics routinely provide assessment batteries to obtain basic subject and demographic information, to evaluate broad domains of functioning, to attain information for diagnostic purposes, and to rule out conditions that might require other interventions or attention (e.g., medical diagnosis). In clinical work pretreatment assessment is designed to assess emotional and behavioral problems, personality, and functioning in everyday life. Ideally, a significant portion of the information obtained at this stage determines the goals and foci of treatment.

*Discrete assessment after treatment is completed* (posttreatment assessment) is familiar as well. In clinical settings, impressions and global evaluations may be placed in the form of narrative summaries (e.g., discharge summaries) rather than formal posttreatment assessment on psychological measures. In treatment research, posttreatment assessment batteries invariably are used to evaluate change and to compare alternative treatment or control conditions. In clinical work, assessment after treatment usually is not of optimal benefit to the individual client, unless the information is used to continue in treatment, to make recommendations for referral, or to confirm that goals have been reached or significantly approximated.

*Ongoing, continuous assessment before and during the course of treatment* is central to evaluation in clinical practice. Ongoing assessment can be used to chart where the client is at the beginning of treatment and to see if changes are made over time. Several data points are needed, not merely to assess the mean level of functioning over time, but also to have an idea regarding variability and trends on the measure. Ongoing assessment begins prior to treatment, after the initial goals and treatment foci are identified. The purpose is to have some initial idea of the level of dysfunction in relation to the domains that serve as the major focus. Pretreatment assessment for as little as one or two assessment occasions (although more occasions generally are better) provides a baseline to help evaluate subsequent progress. Also, it is quite possible that the level of performance on the measure is at an extreme due to stress or crisis, and that marked changes from the first to second assessment occasion would be expected due to statistical regression, passing of the crisis,

and repeated testing on the instrument. If improvement is large during pretreatment assessment (baseline), this may have implications for redefining the goals of treatment, the means to obtain them, or the selection of measures to evaluate treatment.

As highlighted in the context of single-case designs, the initial assessment provides descriptive information (baseline) about the level of performance and its variation. It is possible that only one assessment occasion will be feasible or, indeed, that no assessment is feasible, due to the urgent nature of the treatment. In most psychotherapy cases, it is not clear that intervention is absolutely essential at the first contact. Usually, assessment can begin while there is some effort to manage the situation. In cases where treatment begins immediately, there may also be the prospect of retrospective baseline, where an estimate is provided by the client and others in his or her life. Apart from baseline assessment, evaluation during the treatment phase is pivotal to evaluate change during this period.

Frequency of assessment is dictated by numerous factors, including the assessment method, the specific instrument, and ease of assessment. If assessment is restricted to the weekly therapy session, it would be useful to obtain some information each session. To that end, assessment can be incorporated into treatment. For example, the first or final 10 to 15 minutes of treatment can be devoted to completion of the assessment. It may be that the information is of interest for use in the session, and hence assessment at the beginning of the hour is important. Alternatively, reflection on what transpired in treatment may be of special interest. For some clinical problems, assessment is carried out at home (e.g., logs in which to record relationship issues, migraine headache frequency, urges, pleasant activities), and hence need not utilize much of the therapy hour. The central requirement is to conduct assessment on several assessment occasions over the course of treatment.

#### *Implementation of Assessments*

Typically, therapy does not involve systematic assessment in an ongoing way. Consequently, an initial step is to integrate assessment and treatment as part of the service designed to improve the patient. There are several practical decisions and options for implementing the assessments that can serve the goals of empirical evaluation. Many of the options may be dictated by individual characteristics, problems, and circumstances of the patient; by the nature and type of treatment service; and by characteristics and preferences of the therapist. For example, among assessment options, therapists might conduct a brief assessment at the beginning or end of each treatment session. Alternatively, patients can complete measures and bring them to the sessions or can call and leave recorded messages on a clinic answering machine if a log is kept or to be reported. The therapist may rely on others to obtain as well as to summarize the information. Secretarial staff or assistants can conduct a brief telephone interview covering critical items or can direct the client to a room immediately before a session so the client can complete the measure(s). Similarly, a staff member or volunteer can sum scores or enter the information

or data on an office computer-software program to summarize or to graph the information. A number of options are possible for administering or assisting with assessment or summarizing the data. Ethical protections for office information apply to all such information in the usual way. Options should not be constrained by presumed assumptions that assessment procedures or persons involved in them might invariably tarnish the treatment process.

Whatever the assessment method, implementation requires discussion with the patient. Use of the measures during clinical treatment is introduced in a way that conveys the interconnectedness of treatment and evaluation, so that the client understands that assessment is an important part of the care and service that is provided. The rationale for assessment can be provided at the outset, when treatment is explained. Salient points include the importance of an initial evaluation and goal selection. Goals, and hence measures, may change during treatment. Thus, the focus on initial presenting complaints and their assessment would not preclude a subsequent shift.

Even if goals do not change, evaluation of the assessment procedure is important. Before or early in treatment it would be important to assess whether the level of client functioning is dysfunctional. It is possible that the measure would show that the client is functioning well on the measure. For example, if depression were assessed with an adult client and low Beck Depression Inventory scores (e.g., 0-3; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) were evident in pretreatment, reconsideration of the construct or measure would be important. If further discussion and interview reveal that depressive symptoms are manifest, it may be necessary to develop another measure, including an individualized measure.

#### *Design*

As mentioned previously, methodology for clinical work will also require consideration of ways of arranging treatment to draw inferences, i.e., research designs. The purpose of research designs is to help rule out or make implausible alternative explanations that might be advanced to account for changes seemingly associated with the intervention. Although we routinely and often mechanically engage in practices to accomplish these goals in research (e.g., random assignment of subjects, no treatment control groups), it is not the practices per se that are central. Indeed, it is often the case that exalted methodological practices (e.g., random assignment of subjects) are not likely to achieve their goals (e.g., distribution of various "nuisance" variables equally across groups or conditions) (see Hsu, 1989; Kazdin, 1992a). This does not cast aspersions on the entire research enterprise; rather, it merely underscores the importance of understanding the rationale of various methodological practices. Once the rationales are understood, novel practices can be implemented when more routine practices are not viable options or are not likely to be effective.

The concerns (threats to validity) toward which clinical research is directed are extensive (see Kazdin, 1992a). In clinical practice, the central design issue is whether change is likely to be due to any of the events that transpired in treatment. Conclusions about the impact of treatment can be augmented by



obtaining pretreatment (baseline) information, by conducting ongoing assessment over the course of treatment, and by charting changes in treatment as these normally occur in clinical practice. As noted previously, the pattern of data over the course of therapy can help in deciding whether change has occurred and the likely basis for change. The assessment of therapeutic procedures or processes can provide an evaluation of the likelihood that influences intended to achieve change contributed to outcome. The overall pattern of data in relation to understanding the goals and clinical dysfunction also contribute to the plausibility of alternative explanations of the results. In the clinical situation much information can be brought to bear to rule out threats to validity (see Kazdin, 1981). Even so, ruling out explanations is likely to be of secondary interest in a methodology of clinical practice.

#### Data Evaluation

Evaluation refers to identifying whether change has occurred, is reliable, and departs from the fluctuations one would expect without the intervention. Evaluation relies on the assessment data, which are examined systematically. Ongoing assessment provides data before and during the course of treatment. Several methods are available to evaluate single-case information. Of all methods, graphical display (e.g., a simple line graph) is particularly useful for seeing the pattern in the data obtained over time (see Kazdin, 1982; Parsonson & Baer, 1978).

Nonstatistical data evaluation methods, drawn from single-case experimental research, can be used for clinical evaluation. As mentioned previously, several characteristics of the data (changes in means, levels, and trends; latency of change across phases) can be useful in deciding the reliability of the changes. In single-case experimental research, these criteria are examined across several different experimental design phases beyond pretreatment and during treatment phases. In research, experimental arrangements (designs) and data evaluation act in concert to strengthen the inferences that can be drawn. In clinical work, characteristics of the data still can be invoked to evaluate change and to make decisions about treatment.

Graphical displays and nonstatistical data evaluation are examples and by no means exhaust the methods of evaluating change. A variety of other evaluative aids are available, varying in degrees of sophistication and complexity. For example, several methods of graphing data point for multiple subjects from group research (e.g., stem-and-leaf plots, box plots) might also be considered to plot multiple data points for individual subjects as a means to describe and to evaluate progress (see Rosenthal & Rosnow, 1991). Also, descriptive and inferential statistical techniques are available for continuous data obtained from the single case, although assumptions of analyses and requisite information (e.g., large number of data points) can limit the options (see Gottman, 1981; Kazdin, 1982). Efforts to extend statistical evaluation to individual subjects over repeated occasions are not new (e.g., Payne & Jones, 1957). Such work can be systematized to identify tools for evaluation and to elaborate and investigate their properties in order to maximize their interpretability.

The above methods suggest options for evaluating the changes in patient functioning to identify whether improvements have been achieved. The methods, whether nonstatistical or statistical, focus on the reliability of the change. The importance of the change in terms of clinical work is obviously critical. *Clinical significance* refers to the practical value or importance of the effect of an intervention, that is, to whether it makes any "real" difference to the clients or to others. Therapists and their clients invariably are concerned with an important, noticeable change that affects everyday functioning. Attaining clinically significant change is obviously important. Although many methods have been identified to reflect clinical significance, there is no consensus on any particular measure. Also, many issues for research and clinical work remain to be addressed regarding their use and interpretation (see Jacobson, 1988; Kazdin, 1992a). For example, it is quite possible that the client reports marked change, is satisfied with treatment gains, and wishes to move on to another issue or to terminate therapy. Even so, the magnitude of change on a given measure may not meet some criterion for clinical significance. In short, clinical significance is an important consideration. The initial concern in introducing assessment into treatment is with the reliability of the change. That is, the priority is to establish that there has been a change.

#### Case Illustration

A methodology for clinical practice begins with development of assessment for the individual case. Selection and identification of initial goals and ongoing assessment of progress are central. A case description highlights these initial steps to show use in practice by focusing on assessment, leaving aside other features, including design and data evaluation methods.

*Brief background.* Georgiana is a white, 39-year-old woman who referred herself for outpatient treatment. She is married, with two children (ages 16 and 17). She and her husband are both college educated; based on education, income, and husband's occupation they are from a middle socioeconomic class. Georgiana is not employed outside of the home; her husband is a manager of a computer software firm. She and her husband have been married for 18 years.

Georgiana scheduled an appointment because she said she was depressed and needed to talk to someone. During the initial interview at the clinic, Georgiana saw a male therapist, who asked, with open-ended questioning, about the reasons she sought treatment, sources of satisfaction and dissatisfaction in her life, relationships with significant others, symptoms, and related matters. Toward the end of this discussion, the therapists queried Georgiana about what she expected and wished to obtain from treatment.

At the interview, Georgiana indicated that she had been treated for depression on two separate occasions in her life, once during college and once after the birth of her first child. On each occasion she was placed on medication. She reported some relief but also complained about side effects and did not feel really helped, overall, with her problems. Currently, she said she thought she was depressed again. She reported feeling "empty and lost" about her life

and marriage. She said that she lacked meaning and direction in her life. She felt her life was "wasted" because she was not engaged in anything productive. She felt alienated from her husband and her children. In the case of her husband, she felt great emotional distance due to years of reduced intimacy, joint activity, and time together. Her children were very important to her, but she felt she was not needed by them very much now that they were teenagers. Georgiana identified as her own goals for treatment simply feeling better about her life and not being depressed. She wished to feel some direction and to improve relations with her spouse and others. The therapist explained that he felt that a cognitive-behavioral treatment would be appropriate for these goals and explained how the treatment worked.

**Assessment.** In the initial interview, the therapist introduced systematic assessment after the open-ended discussion provided an initial formulation of the focus. The assessment procedures were used to help make goals and directions of treatment more explicit and to quantify the domains that were to be the foci. Three measures were used by the therapist. The first measure was designed to reflect the domains Georgiana identified as important. At the end of about one hour of the interview, the therapist identified 4 major concerns, themes, or areas as a beginning for them to work on: (1) depressive thoughts and feelings, (2) little involvement in meaningful and fulfilling activities for herself, (3) disengagement from her family, and (4) few supportive contacts outside of the home. These 4 areas were identified by the therapist, based on the interview, and were discussed to see if they captured Georgiana's experience, since they did not follow exactly from Georgiana's formulation of the problems. For each theme Georgiana was asked to help construct statements that were graded to indicate different levels of functioning. The goal was to comprise a 4-point scale for each theme in which the 1 = worsening of the problem, 2 = current functioning/feelings, 3 = some improvement, and 4 = her goal of how she wanted to be/feel on this domain. The therapist referred to this as the *Georgiana Scale (G Scale)* to help guide them during treatment. The therapist conveyed the concept of the 1-4 point scale, but Georgiana provided the content of each of the statements. He asked her to describe a way to characterize her current functioning or where she was now, what it would be like if she became worse, what some improvement might look like, and how things would be if she really made the kind of change she wanted. Table 1 presents the 4 themes and the graded statements Georgiana and the therapist constructed. For the assessment on the G Scale, she was instructed to select the statement under each theme area that characterized how she felt during the previous week. After the scale was developed in the initial session, Georgiana was asked specifically if the second statement of each theme area really captured her current feelings; she reported it did.

After this scale was developed, two other measures were described and given to her to complete. She also completed the *Beck Depression Inventory (BDI)*; Beck et al., 1961) to reflect severity of depressive symptoms. The measure includes 21 items; for each item, 1 of 5 (but scored 1-3) statements is selected. The statements differ in severity of the depressive symptoms. Georgiana also completed the *Quality of Life Inventory (QLI)*; Frisch, Cornell, Villanueva,

TABLE 1  
FOUR THEMES AND ITEMS FOR GEORGIANA

### I. Depressive Thoughts and Feelings

1. I feel more depressed and dejected than I did before I started treatment.
2. I feel about the same level of depression and dejection as I did when I started.
3. I feel a little better about my mood, and things are not as bad as before.
4. I feel a lot better; I do not think about my feelings as negative, and I have more energy to get out and do things.

### II. Involvement in Meaningful and Fulfilling Activities

1. I really feel paralyzed about doing anything anymore.
2. I am not doing anything differently now or anything special I like compared to when I started treatment.
3. I feel better that I have some direction and focus in what to do.
4. I am totally involved in some things as a career that gives me good feelings about life.

### III. Disengagement from Her Family

1. I do less with my husband and children than before and don't seem to care about doing things with them.
2. Things really have not changed about my feelings. Everyone at home does his or her own thing, and my husband and I mostly just eat meals together.
3. My husband and I are a little better. We go out once in a while and are a couple again.
4. My husband and I are really "together." We are intimate in many ways, and I can feel that he cares for me.

### IV. Supportive Contacts Outside the Home

1. I am isolated from people in general, including my relatives who live in town.
2. Once in a while I see someone when I shop or at a school event with my children. We chat a bit but nothing beyond the superficial.
3. I meet someone to have coffee with or to go to an event or shopping with during the day.
4. I meet a few people by myself or some couples that my husband and I can get together with, and we do this on a regular basis.

**Note.** The theme areas were derived from an open-ended interview with Georgiana. The specific statements were generated by her to reflect what it would be like to become worse, to remain the same, to improve a little, and to achieve her goal for that theme. These alternative outcomes reflect numbers 1 through 4 respectively under each theme. Each time this particular measure was completed, Georgiana selected the alternative statement under each theme that was closest to how she felt during the previous week.

& Retzlaff, 1992). The measure is a self-report scale that assesses overall quality of and satisfaction with life. Seventeen domains of life (e.g., love relationship, home, learning, recreation, friendships, philosophy of life, work, health, neighborhood) are covered. The weighted score is used for each domain, based on the client's rating of the importance of that domain in his or her life theme (0 = not at all important, 2 = very important) multiplied by the satisfaction derived for that domain (-3 = very dissatisfied to +3 = very satisfied). The psychometric properties of the scale (e.g., measures of internal consistency, test-retest reliability, criterion, and concurrent validity) have been favorable

(see Frisch, 1992). More pertinent for the present use, the scale has been used in clinical work with individual patients. A completion time of approximately 10 minutes further facilitates its use.

The initial contact with Georgiana lasted approximately 2 hours. Approximately 1.5 hours was spent with the therapist for the interview and development of the G Scale. After this, Georgiana completed the BDI and QOLI and was scheduled to return the following week. She was asked to come 20 minutes before the session. When she returned for the second session, she completed the BDI and QOLI before the session began. After these were completed, she brought them to the therapist.

The therapist began the session by asking her to select 1 statement from each of the 4 theme areas that they had discussed. The material had now been typed (similar to Table 1). They briefly discussed whether the areas were still important to her and whether she felt that their last interview missed critical material. The therapist conveyed that the initial goals were a place to begin and that the information within the sessions and from the assessments would be important to make any midcourse corrections as needed.

At this point, the therapist described and began treatment. The treatment employed by the therapist consisted of cognitively based therapy combined with features of interpersonal psychotherapy. Cognitive therapy focused on maladaptive cognitions about herself related to her depressive affect, poor self-esteem, feelings that life was not worthwhile, and internal attributions regarding her views of herself (see Beck, Rush, Shaw, & Emery, 1979). Interpersonal psychotherapy focused on her interpersonal relations, roles and the sources of satisfaction and emotion associated with each, and many of her feelings about herself as a spouse and parent (see Klerman, Weissman, Rounsaville, & Chevron, 1984). Therapy was an integration of these treatments (e.g., cognitions about her relations, emotions associated with them, her own attributions and those of others). She was also given specific tasks at the end of each session that were designed to increase mutually shared activities with her husband and to expand her activities and roles outside of the home. These activities were part of "assignments" developed with Georgiana at the end of each session based on work within the session, any prior progress on other activities, and so on. Treatment was conducted on a weekly basis. Highlighted here are the assessment information and its use rather than the details of the treatment.

Each week, Georgiana was asked to come about 20 minutes before the session and to complete the G Scale, BDI, and QOLI. At the first and second sessions, the full scales were administered. However, there were symptoms and domains within the two standardized scales that were not problematic or relevant to Georgiana. Abbreviated versions of the BDI (15 items) and the QOLI (13 domains) were constructed and used throughout treatment, based on elimination of selected items.

The assessment information included scores from the G Scale, BDI, and QOLI. From each scale, performance was quantified with a summary total score for that measure to examine whether any systematic pattern or change might be evident as treatment progressed. The assessment information is

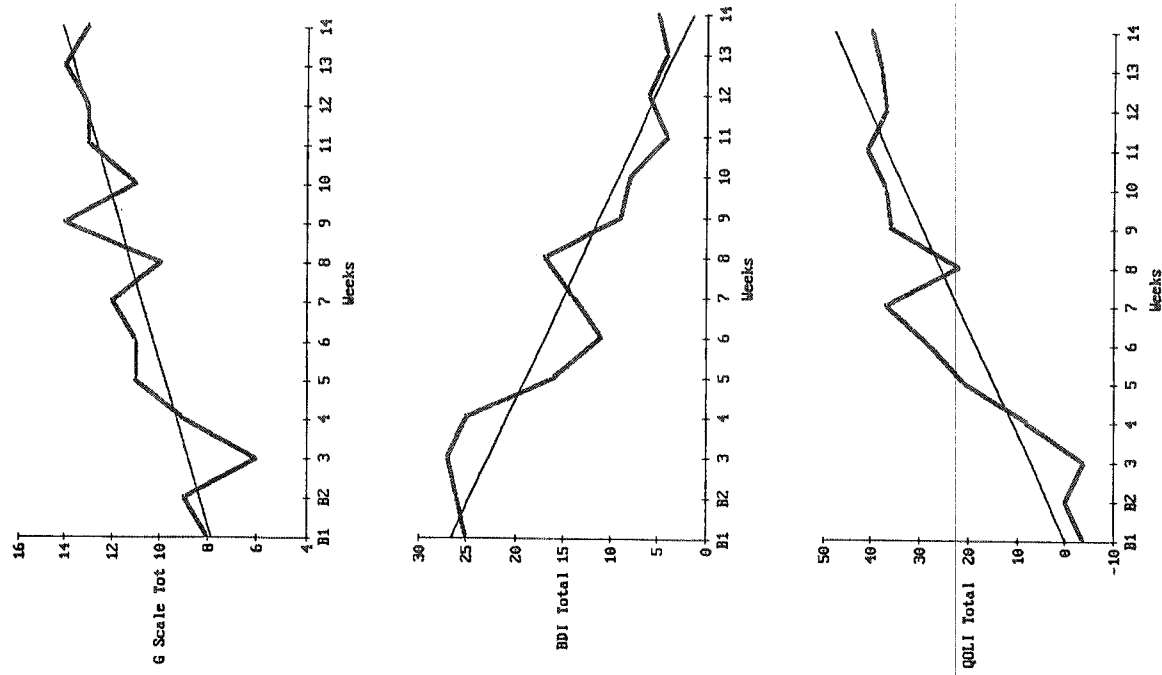


Fig. 1. Session-by-session scores for Georgiana on three measures including the "Georgiana Scale" (G Scale, upper panel), the modified Beck Depression Inventory (BDI, middle panel), and the modified Quality of Life Inventory (QOLI, lower panel). The scores for each measure represent those items (BDI, QOLI) that were selected as relevant to the client and do not reflect the complete scales. Data are represented for 14 sessions (weeks). The first week was devoted entirely to interview and assessment. The second week began with completion of assessments followed by the initiation of treatment. The first two weeks (B1, B2) refer to baseline or pretreatment as assessment. Given the direction of scoring of the measures, improvement would be reflected in increases for the G Scale and QOLI and a decrease in the BDI. Fitted to each graph is a linear regression line over the course of all data points.

graphed in Figure 1. The two assessment occasions prior to treatment were delineated as baseline (weeks B1, B2 in the figure). Added to each graph is a linear regression line to characterize the slope that best fits the data. Overall, the individual data points and regression line suggest a trend toward improvement over time.

Although the overall scores are useful in summary form, the mean for all of the items of a given measure (e.g., BDI) with an individual case can suffer the same liability of using means in group research; to wit, they can obscure information that would be important to identify. In Georgiana's case, the G Scale and the QOLI indicated that she had made little progress in her relationship with her husband. The relationship issues emerged more fully in the treatment session of week 14. At the beginning of the session, the therapist indicated that he thought this would be a good time to discuss at length the original goals of treatment and how she had been doing, based on the assessment information and Georgiana's appraisal of treatment.

Georgiana indicated that she had felt much better about herself and her life. Her thoughts about her life, what is important, and her direction were much better. Over the course of treatment, she had initiated a number of activities. She began a class at a local university and now planned to seek a degree in nursing. She had developed more interaction with a neighbor, whom she met almost daily to engage in routine activities. Also, through her class she met a few people whom she enjoyed. Time with her children was more enjoyable. In general, she felt much better about her overall well-being. At the same time, she felt her relationship with her husband was not helped at all by treatment. Although she and her husband went out on a couple of "dates," she felt this turned out merely to be time together with no connection or closeness. She said she loved her husband and could not imagine being without him but that there seemed to be no closeness or contact like there used to be. At this point, the therapist suggested that they focus more on this for a few sessions but that the immediate goal would be for her and her husband to consider joint steps toward improving their marriage. The method highlighted previously for developing the G Scale was used again to identify theme areas within her marriage that were significant and to identify the anchor points. The measure was added at this time to the assessment.

Treatment was continued for 5 more weeks. Georgiana no longer completed the BDI and QOLI weekly. These were now completed every other week. Weekly assessment continued for the original G Scale and for added thematic areas related specifically to her marital relation. After 5 weeks, Georgiana and her husband agreed to begin marital counseling and Georgiana was no longer seen individually.

*Commentary.* There are useful features of this case that illustrate an approximation toward a methodology of clinical practice. First, efforts were made to make initial goals of treatment explicit and to quantify these goals. The assessment procedures included a highly individualized scale (the Georgiana Scale) using a method of scale development bearing resemblance to others used clinically (e.g., Shapiro, 1964). The two standardized measures (BDI, QOLI) were altered to address the specific domains that seemed relevant to

the client. The information was charted to evaluate progress, and this information was used both to alter treatment and to further assessment.

There are many limitations one can identify as well, although some may be unfair because all of the details of the case cannot be presented. First, the three measures are reasonable in part because of the domains they assess and the clinical feasibility of their administration and completion. Yet, each is a self-report scale. The argument could be made that self-report is pivotal, given self-referral as the basis for initiating treatment. However, the common method variance (all self-report) and likely overlap of foci (e.g., G Scale depression theme and BDI) might have made these measures highly correlated. Assessment with one other method would have augmented the evaluation of progress. Examples might have been engaging her husband (with Georgiana's active consent, of course) in periodic assessment (e.g., telephone call to rate an adjective checklist or Georgiana's depression or activities, or his own feelings about their relationship). Related, all of the assessments were completed in the treatment setting, even though they were designed to reflect feelings during the prior week. If all self-report measures were to be used, it may have been useful to diversify the formats and observation setting. For example, Georgiana could have kept a daily log of cognitions or pleasant activities engaged in between the sessions.

Second, the issue of marital dissatisfaction might have been identified earlier and taken a more central role early in treatment. From the brevity of the description, we do not have a sense of the range of clinical problems and their priorities. Consequently, we cannot tell if marital functioning was central early in treatment or appropriately deferred until other issues were addressed. The therapeutic strategies and assessment procedures might have been different. Alternatively, the marital issues may have become more salient as treatment progressed precisely because the client felt progress in other domains. The lack of progress on her relationship with her husband may have been clarified by assessment. The information suggests that the assessment and focus of treatment warranted change. The case reveals a more general problem; namely, how decisions are reached in treatment. The case shows the collection of systematic assessment information. How this information is to be used and the criteria for shifting treatment foci are not automatic. These matters are not raised by collection of systematic data but rather brought to light more clearly because of assessment.

Third, the data presented to us to evaluate ongoing treatment are helpful, to be sure. At the same time, the means from each measure include multiple items. For the therapy process, it would be important for the therapist to consider areas (items) within a measure that may not have changed, even though psychometrically we are aware of the limits of individual items. Some evidence for use of individual items comes from evaluation of the G Scale items that focus on Georgiana's relationship with her husband. The lack of progress here was used as a basis for further discussion of this theme, a reasonable use of the information.

Fourth, further comment on the data clinically would be very useful. For example, at week 8, the individual graphs in Figure 1 suggest that there was

a slight increase in symptoms and dysfunction. This is a case in which qualitative and quantitative information are so important in combination. Although the change in the figure might be chance fluctuation, we would like to know if something in Georgiana's life during the previous week or something within the treatment is related to the change. The information is crucial, in part, because if something has exacerbated Georgiana's symptoms, it would be useful to identify the influence and to consider addressing the matter as part of treatment.

Fifth, the use of regression lines to describe the data is helpful in charting overall trends. Also, computer programs that permit data entry can calculate these automatically, so the computational tasks are not onerous. The interpretation of the regression lines in the present case must be made cautiously. The lines suggest an overall improvement and hence are very useful in conveying a pattern. Presumably, the absence of improvement in a given area would have led to differences in treatment. At the same time, the changes over time cannot be interpreted as being the result of treatment. Denoting change is of course different from understanding the basis of the change. In this case, clinically, we might be satisfied in having systematically charted progress.

Finally, we do not have an idea of how treatment was executed, i.e., how Georgiana adhered to the tasks in the session, as well as to the homework assignments. The improvement suggests she profited from treatment in the domains that were assessed. Yet, we would like to have an idea of whether she adhered to assignments and whether the therapist felt that cognitive issues and interpersonal functioning were suitably addressed within each session. Such information over the course of treatment would help a bit to identify relations between treatment and outcome.

Overall, the case highlights the use of systematic assessment. Evident from even the brief description is that more could be made explicit about operationalizing treatment procedures and goals and about the decision-making process regarding the course that the therapist charts. Why the therapist chose cognitive therapy and interpersonal therapy combined, as well as the decision rules in clinical work, warrant explication. More relevant to the point, however, is the clear effort to individualize assessment and to make explicit the selection of treatment goals, and the evaluation of progress toward those goals in an ongoing way.

### *Development of the Methodology*

#### Issues and Limitations

Developing a methodology for clinical practice, as suggested by the case description, raises scores of issues with different implications for the field. To begin, there is, of course, the methodology of clinical practice itself. The prior discussion has tried to outline in broad terms how methodological practices and the underlying thought processes that methodology fosters can augment clinical work. Yet the methodology is not yet developed and the emergent issues are not yet identified.

Among immediate concerns are the methodological tools available for assessment. Several assessment conditions for clinical work depart from those that serve as the basis for the bulk of test validation. To begin, standardized assessment is designed to provide a constant set of items to all individuals. This can be useful for selecting a subset of items to suit individuals who are seen in treatment. However, more individualized assessment may be needed. Assessment methods for clinical work require elaboration of strategies to obtain information, as well as identification of specific measurement instruments. Existing measures might be helpful in content and format, but the assessment needs and methods of their validation may be fundamentally different for clinical work from those used in current nomothetic scale development and validation.

Relatedly, the repeated administration of a measure over time constitutes a major departure from the usual test administration procedures. Validity and interpretation of the measure might well be affected by repeated assessment. Research in clinical psychology needs to develop and to validate assessment strategies for ongoing assessment in clinical work. At the same time, many measures (e.g., "objective" personality tests, symptom checklists) with established construct validity might be adapted to clinical work. Nevertheless, it is true that measures need to be developed for assessment in the clinical situation, for repeated administration, and for assessment conditions that are different from those in which most measures were standardized. One of the central purposes of clinical research is to provide the underpinnings of clinical practice. To that end, assessment practices in clinical work can be evaluated and further developed in research.

Similar issues can be raised in relation to design and data evaluation strategies. The use of elements of single-case designs warrant further elaboration outside the context of behavioral assessment and interventions. The most frequent conditions in clinical practice are likely to be when assessment is conducted before treatment and during treatment; when assessment is conducted during treatment only, without pretreatment assessment; and when assessment is conducted and goals and measures change over the course of the treatment. These situations present methodological challenges but not insurmountable obstacles. In the move toward drawing valid inferences, the most crucial step is that of obtaining systematic assessment information on the intervention procedures or processes as well as on client progress. Assessment aids the inferential processes by permitting the examination of change. Also, the pattern of continuous assessment data can assist in suggesting whether a given change in treatment might have contributed to change.

Similarly, data evaluation methods are available but not well explicated for clinical use. Nonstatistical data evaluation in treatment research is generally unfamiliar beyond the cadre of researchers involved in behavior analysis. Descriptive statistics of the sort mentioned previously (e.g., changes in means, slope) can be used for inferential purposes. However, the methods need to be developed for general use. Elaboration of evaluation methods for clinical use is required to provide tools and to clarify their use and limitations.



### Objections and Myths About Evaluation

The approach advocated here is to propose that systematic assessment and evaluation be integrated into clinical work. Stated in the general form, it might be difficult to envision opponents. Yet, clinical practice already can be burdensome and demanding. It is likely that systematic assessment and evaluation would be seen as an additional burden with no clear benefit. Several objections are commonly voiced as reasons to eschew systematic assessment and evaluation. The objections are critically important, because they help to elaborate the special contribution of evaluation to clinical care.

*No need for evaluation—I know change when I see it.* In clinical work, we are often confident that we can readily detect changes in the client. No formal assessment is necessary if the clinician can reliably identify the changes. It certainly is true that systematic evaluation is not needed for some of the changes that could take place. Very dramatic changes (e.g., elimination of all obsessions, compulsions, signs of sexual dysfunction, or symptoms of Tourette's syndrome), so-called "slam-bang" effects (Gilbert, Light, & Mosteller, 1975), may make formal evaluation unnecessary. Yet, in most cases, treatment effects are not likely to be so dramatic or unidimensional that they can be unequivocally determined without formal evaluation. Patients can profit directly from systematic evaluation to identify change and the amount, scope, and stability of that change.

Change in one domain or modality and intermittent or brief improvements can be seen, recalled, and viewed as sufficient signs of progress or improvement. Impressionistic information and selective recall maximize the impact of judgmental biases (Nezu & Nezu, 1989). It is also quite possible that improvements would be overlooked if a domain that changed is not one systematically recalled by the therapist. In addition, exacerbation of symptoms early in treatment or the absence of change over time might be particularly critical to identify and likely to be missed. The use of a more formal assessment procedure and methods of data evaluation are designed to help redress the recognized limits of perception, whether therapist perception or client perception. This does not mean that psychological measures are flawless. Measures permit a consistent and standard method of assessment. In the case of therapy, ongoing assessment provides a relatively constant barometer of patient functioning in domains of interest and is likely to be more reliable than impressionistic observations alone.

*Evaluation interferes with treatment.* A concern is that assessment may interfere with the therapeutic relationship. The therapist is responsible for treatment; adding to that the role of the assessor or evaluator may, on conceptual grounds, mix roles and be viewed as antitherapeutic. Introducing evaluation, it is sometimes argued, competes with the very relationship that the therapist wishes to foster. There is no doubt that particular models or theoretical conceptualizations of treatment might argue against having the therapist in the role of the person who provides treatment and who evaluates progress. Yet, it is unfortunate to presume without supporting evidence that evaluation harms when alternative assumptions (e.g., not evaluating the patient can permit harm to occur, evaluation may have no impact, evaluation may help) are plausible

as well. It is probably the case that the way clients perceive systematic evaluation has to do with the therapists' views of evaluation and with the way the evaluation objectives and methods are presented to the clients. If evaluation is presented as a matter of course, as central to treatment, and as purposeful, then the clients' views are likely to be positive. For better or worse, patients are quite accepting of the rationales and treatments that are provided for their care. Ongoing assessment can be conceived as central to treatment and presented to clients in that fashion.

*Everyone is an individual and assessment obscures individuality.* The shield that is most often raised against systematic evaluation is the view that every patient who comes to treatment is an individual. Administration of measures, particularly some standardized set of measures, may ignore the individuality of the patient. Yet systematic assessment can be quite individualized (see Shapiro, 1957, 1964). The clinician can decide with the patient those domains of functioning that are most relevant and build the assessment devices to reflect that. As priorities change in treatment, the evaluation procedures (measures) can be altered as well. Clinical practice, unlike the usual research context, permits individualization of both assessment and treatment. The methods can be tailored to the patient's priorities, resources, and characteristics. What is needed for clinical practice is a set of strategies for clinicians, so that multiple assessment and evaluation options are provided to address individual circumstances of treatment (Cone, 1988).

A point worth noting in passing is that standardized assessment has a critically important role in therapy in addition to individualized assessment. The profile of a patient across standardized measures and the standing of a patient relative to a normative group of peers of the same age, sex, and ethnicity, for example, can provide meaningful data that may also guide treatment. For example, the vast data on MMPI-2, a standardized measure, might be informative as a profile for domains of patient personality and psychopathology. The standardized nature is not a threat to patient individuality, but rather an opportunity to examine that individuality against a broader backdrop. From the standpoint of clinical evaluation of ongoing treatment, the more salient point is worth underscoring, namely, that systematic, quantitative assessment and evaluation can be readily individualized.

*Assessment trivializes patient problems.* Assessment requires that the therapist operationalize the therapeutic focus. A reasonable concern is that, in so doing, the patient's problems are reduced in a way that does not represent their experience. Many issues raised in treatment are difficulties involving relationships, feelings of ambivalence, guilt, angst, loss of meaning, and crises of identity. Presented with one or more of these, the therapist is asked to identify assessments to examine the impact of treatment. The constructs and foci may seem to present obstacles for such a demand. For example, if the therapeutic focus is "capacity to love," what measure is available? If a measure were developed, would this not trivialize the patient's problem and experience?

Operational definitions are designed to reflect the construct or domain of interest; they are not designed to encompass all of it. If we assess capacity for love or capacity for relationships by a 10-item inventory, this does not mean

that we are only interested in changes on these 10 items. We use the measure as a guide to and a sample of the broader area. We devise the items individually because they reflect (are highly correlated with) critical foci for this patient, but we are interested in the nontest referents (constructs, latent variables) these are designed to reflect. A measure is a guide for the therapist; it is designed to provide a check on a relevant and, hopefully, central feature of the domain.

A concern is that complex human experiences by definition cannot really be measured and that operational definitions are not even distant approximations. Yet, the onus of clinical psychology is to develop measures of the constructs that may be central to clinical work. Therapists do not have the luxury of awaiting such data, and, indeed, may be in a better position to obtain and generate the needed information. If the domain cannot be operationalized, we have another concern, namely how will the therapist know that improvement is evident? Complex constructs of interest can be studied and validated in the context of treatment. For example, in psychological and medical treatments, we are often interested in improving the "quality of life," as briefly illustrated in the case example noted earlier. There are few constructs as broad, global, and, perhaps, as clinically important. Quality of life is not illusive from the standpoint of assessment and assessment validation (Frisch, 1992; Keith & Schalock, 1992).

*Clinical problems change, so regimented assessment cannot be followed.* An objection against evaluation in clinical work is based on the dynamic (not psychodynamic) nature of treatment. Specifically, in most psychotherapy there is not a single, simple patient problem that remains constant. Problems change and foci are redefined over the course of treatment. Indeed, evidence suggests that over half of patients seen in therapy add new target complaints over the course of treatment (Sorenson, Gorsuch, & Mintz, 1985). The changing focus of treatment and the multiple nature of the foci are not arguments against assessment. Rather, they make assessment all the more important. It is critical to identify changes in problem domains and priorities of the foci from the standpoint of the patient and therapist.

Systematic assessment is particularly important to evaluate change and patterns of change. If the information indicates that the initial pressing issue is no longer important or appreciably less important relative to some other domain, the assessment and evaluation can be altered. An excellent feature of clinical work and clinical evaluation, unlike most clinical research trials, is that the assessment and treatment are not fixed. Therapists and clients can set new goals *in medias res* and present or withdraw assessments to reflect these changes. The changes in treatment and treatment foci are not in conflict with assessment. In fact, the changes are likely to be very important. We would want better than impressionistic information to chart these changes and to make informed decisions about treatment.

### Professional Issues

Introducing systematic assessment and evaluation routinely in clinical practice has dramatic implications regarding professional training, a point that

can only be acknowledged here. Central among the issues is the model training in clinical psychology. The model advocated here might be referred to simply as a *psychological model*. Psychologists have by training, con experience, and orientation a concern for the methodological tenets of social research. The commitment first is to assessment, evaluation, testing of hypotheses, empiricism, and so on. In clinical work, these need to be adjusted to ensure that they are mobilized to help directly this patient here and there. The model required for clinical psychology is not necessarily special, although the training in therapy and the assessment and evaluation techniques themselves might well be quite unique. Yet, the commitment to systematic assessment and evaluation, one would expect, comes from entry into the field rather than from theoretical orientation, specialization area, and so on.

In our clinical work as therapists, it is fine to use that very special vocabulary that one is wont to hear; to wit, we make decisions based on our own of what "the patient needs," our "gut-level reaction," how we "resonate to experience," and whether we or our patients are "comfortable" with that. Therapy requires its own language, as do other speciality areas within the field. Yet a given is that these vocabularies will be systematically assessed and evaluated, placed in an empirical and replicable arena, and, with art to develop knowledge, in the case of clinical work, used to help individual patients.

Other professional issues that can be raised in the present context are a training curriculum for clinical students to teach the specifics of methodology of clinical practice, as well as clinical therapy experiences in which these specifics are integrated and supervised are not minor issues. The issue of developing the methodology is probably not as difficult as the task of dissemination. There remain ethical and accountability issues as well. Professional treatment procedures and outcomes in an empirical arena is by no means trivial. How the information is used and the threat to therapists and to patients are not minor.

With merely a hint at some of the methodological and professional complexities, it is not clear why one would want to proceed with systematic evaluation. The central argument might be made by considering what the core features of psychology are as a discipline and the unique perspective training implied by the profession itself. However, the argument is weak because methodology of clinical practice has implications to diverse fields beyond psychology where therapy and related procedures are applied. If the goal is to help patients and to address the concerns of this individual here and there then the case for systematic evaluation is, one would think, more persuasive. In clinical work, where the individual is so important and where direct behavior is a goal, *unsystematic* evaluation is hard to advocate.

### Conclusion

The central theses of the present paper are that systematic evaluation is to be used in clinical practice, that systematic information improves professionalistic evaluation of treatment progress, and that client care is imp

by such evaluation. More generally, a methodology of clinical practice is needed that encompasses assessment, design, and data evaluation strategies to enhance clinical care. Central to this methodology is operationalization and assessment of treatment goals and means, as well as ongoing progress of the patient over the course of treatment. Assessment can aid therapists in reaching decisions about how to proceed, what more might be needed, and when to shift strategies.

Assessment and data evaluation are important in overcoming inherent limitations and biases of human judgment. The use of psychological measures and alternative methods of data evaluation does not eliminate biases, but rather enters these biases into an empirical arena in which they can be better studied. The clinician need not slavishly adhere to one or two measures and the data they show. Qualitative and impressionistic information, now the main and usually sole source of information over the course of treatment, can still be drawn upon to evaluate treatment progress. However, attention to more systematic information gathering may provide a stronger basis for decision making to help the individual patient seen in treatment.

Throughout the discussion, a methodology of clinical practice was presented in relation to clinical priorities, i.e., helping the individual in treatment. Systematic evaluation of individual clients can contribute to the knowledge base more broadly. A methodology of the individual case can help to establish lawful relations at the level of individuals. Moreover, accumulation of cases from clinical work can be used to identify patterns such as variations of treatment effects for different types of patients and clinical problems (e.g., Maletsky, 1980, 1991; Marquis, 1991). Evaluation in clinical work can test models of treatment studied in research and the generality of research results to different conditions (types of persons, therapists, settings) of clinical practice. Also, patterns of responding that emerge from accumulated cases in clinical work, such as the mixed or varied effects of treatment with a given type of client, can serve to generate hypotheses for therapy research. In short, there is an important opportunity to be gained from systematic evaluation in clinical practice beyond the benefits to the individual patient. Again, the priority of systematic evaluation has to be the individual case. The accumulation of information and the use of that information (e.g., secondary data analyses) to draw more general conclusions can be extremely important as well.

Introducing systematic evaluation into clinical practice can inadvertently fuel concerns that are not at issue here. The goal of developing a methodology for clinical practice is not intended to turn clinicians into researchers, to increase the median publication rates of clinical psychologists, nor to bring researchers more into contact with actual clinical practice. Such concerns are often viewed as part of the "problem" of the hiatus of clinical research and practice. At the same time, the concerns preserve the hiatus by providing a frame of reference that recognizes the legitimate differences between the clinician and the researcher. The approach advocated here begins from a slightly different perspective; namely, by defining psychology as a profession committed to scientific approaches and tenets toward the subject matter and by placing knowledge, its acquisition, and application on empirical footing. Methods of

research join us, as psychologists, as ways of thinking about and approaching situations. Clinical practice has very special goals, procedures, issues, and priorities. The tenets of science are not erased or abandoned by exigencies of clinical work. Rather, they are adapted to meet and to serve the goals of clinical work. The tenets of science do not require the use of true experimental if the phenomenon of interest places limits on "control" and "experimental manipulation" of conditions. Systematic assessment of process and outcome, evaluation of correlations, developing and testing hypotheses, and creative use of data evaluation strategies can be applied to help guide treatment and to augment understanding.

Of course, it is one matter to advocate systematic evaluation in the abstract, but quite another to suggest a viable methodology for use in daily work. Elements for devising such a methodology are available, elements that begin to chart assessment, design, and data evaluation options to aid clinical work. The contribution of single-case research was highlighted as the central source of such a methodology, because the single-case designs and overall approach have been used to evaluate individual cases and to examine the bases for therapeutic change (Barlow & Hersen, 1984; Kazdin, 1982). Single-case designs include assessment strategies, special experimental arrangements, and methods of data evaluation that can provide underpinnings for a methodology of clinical practice. There are many specific practices that have come to be associated with the use of single-case designs in behavior analysis (e.g., direct observations of overt behavior, use of behavioral interventions). These specific practices can be evaluated on their own merit. However, they are not central to the model proposed for evaluation in clinical practice. The nature of the intervention and the constructs and measures relied on are not dictated by systematic assessment and evaluation. Indeed, it is not the conceptual heritage that clinical psychologists, by training, are intended to share, but rather their methodological approach toward human functioning.

Evaluation in clinical practice raises both concrete and broad issues that were not fully addressed here. As noted earlier, measurement strategies and individual assessment instruments need to be developed further. Among the salient requirements, the strategies and measures must at once be feasible, valid, and meaningfully interpretable over multiple assessment occasions, and must be able to reflect change. Individualization of goals and measures for clients seen in treatment (e.g., Problem-Oriented Record, Goal Attainment Scaling) are not new. The remarkable assessment and evaluation expertise in psychology that has served the development of measurement theory and instrument development and validation can be deployed to develop individualized assessments for clinical use. Formalization of the area of idiographic assessment would help to codify existing tools available for clinical use and help to clarify the research agenda in order to ensure that the appropriate assessment strategies and instruments are available for clinical use across a range of treatment foci, therapeutic approaches, and clients.

A more general issue, not addressed by evaluation in practice, pertains to accountability. In clinical practice for a given case, the practitioner cannot be held responsible for the outcome, progress, improvement or lack thereof

evident in the client. Clinical dysfunction and problems clients bring to treatment are often multiply determined. Insufficient knowledge about the etiology, course, and risk factors for a given problem and about how to controvert such influences present genuine limitations and indeterminacies of treatment. Also, the knowledge base is derived from the study of groups. Findings, even if well established, may not mean that a particular pattern of influences or set of determinants apply to a particular patient. For these reasons, we cannot be held responsible for achieving therapeutic changes that all parties involved would like. We can be held accountable for careful evaluation and for informed decision making. This begins with responsible evaluation of the problem and application of reasonable interventions. In addition, quantitative and ongoing assessment and evaluation of treatment are reasonable clinical responsibilities. Quantitative assessment and evaluation might be advanced as central to all mental health professions, because the goals, accountability, and concerns for individual clients are shared in clinical work. At the same time, among the disciplines, psychological training is unique because of what the field and advanced (doctorate) degree represent. Hypothesis testing, assessment, and evaluation are relevant to both research and practice.

In principle, advancing systematic evaluation in clinical practice is not new. The requirements and examples are not new either (e.g., Davidson & Costello, 1969). There are, however, new opportunities. Research methods and practices have advanced. Development of measures of process and outcome, specification of treatments in improved form (e.g., manuals, guidelines), and concern with and assessment of treatment integrity, for example, can be used to benefit applications of treatment with individual patients. The continued concern of the hiatus of clinical research and practice and model of clinical training may provide impetus for advancing clinical work. Perhaps there is no need to train hyphenated psychologists (clinician-researcher, scientist-practitioner). It may be clinically naive to consider that such a small hyphen can hold together such opposing priorities, concerns, and orientations. Training *psychologists* might serve clinical goals very well; to wit, in the use of systematic assessment and evaluation; operationalization of critical constructs; reliance on, respect for, but not uncritical worship or acceptance of, data; and so on.

There are many obstacles to the integration of systematic assessment and evaluation into clinical practice. The central thesis was not designed to deny what *are* the characteristics of clinical practice, but rather to reconsider what they *ought* to be and then to move toward reconciling these. In clinical practice, reconciling what ought to be with what can be often is a central goal of treatment. Perhaps it is not out of place to consider the task here in relation to the enterprise of therapy. A high priority for the field is the development of a suitable methodology for clinical work that begins with the demands of clinical practice and deploys critical thought processes and methods to advance patient care.

## References

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Alter, C., & Evens, W. (1990). *Evaluating your practice: A guide to self-assessment*. New York: Springer.
- Baer, D. M. (1977). "Perhaps it would be better not to know everything." *Journal of Applied Behavior Analysis*, 10, 167-172.
- Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist practitioner: Research and accountability in clinical and educational settings*. New York: Pergamon.
- Barlow, D. H., & Hersen, M. (1984). *Single-case experimental designs: Strategies for studying behavior change* (2nd ed.). Elmsford, NY: Pergamon.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.
- Beck, A. T., & Steer, R. (1988). *Beck Hopelessness Scale: Manual*. San Antonio, TX: The Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 53-63.
- Bellack, A. S., & Hersen, M. (Eds.) (1988). *Behavioral assessment: A practical handbook*. New York: Pergamon.
- Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, 76, 159-165.
- Bergin, A. E., & Strupp, H. H. (1972). *Changing frontiers in the science of psychotherapy*. Chicago: Aldine.
- Bloom, M., & Fischer, J. (1982). *Evaluating practice: Guidelines for the accountable professional*. Englewood Cliffs, NJ: Prentice Hall.
- Butcher, J. N. (1990). *The MMPI-2 in psychological treatment*. New York: Oxford University Press.
- Cautela, J. R. (1967). Covert sensitization. *Psychological Reports*, 20, 459-468.
- Cohen, L. H., Sargent, M. M., & Sechrest, L. B. (1986). Use of psychotherapy research by practicing psychotherapists. *American Psychologist*, 41, 198-206.
- Cone, J. (1988). Psychometric considerations and the multiple models of behavioral assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 42-66). New York: Pergamon.
- Davidson, P. O., & Costello, C. G. (Eds.) (1969). *N = 1 experimental studies of single cases*. New York: Van Nostrand Reinhold.
- Feldman, R. A., Caplinger, T. E., & Wodarski, J. S. (1983). *The St. Louis conundrum: The effective treatment of antisocial youths*. Englewood Cliffs, NJ: Prentice-Hall.
- Fonagy, P., & Moran, G. S. (1990). Studies on the efficacy of child psychoanalysis. *Journal of Consulting and Clinical Psychology*, 58, 684-695.
- Frisch, M. B. (1992). Use of the Quality of Life Inventory in problem assessment and treatment planning for cognitive therapy of depression. In A. Freeman & F. Dattilio (Eds.), *Comprehensive casebook of cognitive therapy* (pp. 27-52). New York: Plenum.
- Frisch, M. B., Cornell, J., Villanueva, M., & Reizlaff, P. J. (1992). Clinical validation of the Quality of Life Inventory: A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychological Assessment*, 4, 92-101.
- Garfield, S. L. (1980). *Psychotherapy: An eclectic approach*. New York: Wiley.
- Gilbert, J. P., Light, R. J., & Mosteller, F. (1975). Assessing social interventions: An empirical base for policy. In C. A. Bennett & A. A. Lumsdaine (Eds.), *Evaluation and experiment: Some critical issues in assessing social programs* (pp. 39-193). New York: Academic Press.
- Gottman, J. (1981). *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge: Cambridge University Press.

- Hayes, S. C. (1981). Single-case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology, 49*, 193-211.
- Haynes, S. N., Lemsky, C., & Sexton-Radek, K. (1987). Why clinicians infrequently do research. *Professional Psychology: Research and Practice, 18*, 515-519.
- Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 131-137.
- Jacobson, N. S. (Ed.) (1988). Special issue: Defining clinically significant change. *Behavioral Assessment, 10*, Whole Issue No. 2.
- Jayaratne, S., & Levy, R. L. (1979). *Empirical clinical practice*. New York: Columbia University Press.
- Kanfer, F. H., & Scheff, B. K. (1988). *Guiding the process of therapeutic change*. Champaign, IL: Research Press.
- Kazdin, A. E. (1978). *History of behavior modification*. Baltimore, MD: University Park Press.
- Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49*, 183-192.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kazdin, A. E. (1992a). *Research design in clinical psychology* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Kazdin, A. E. (Ed.) (1992b). *Methodological issues and strategies in clinical research*. Washington, DC: American Psychological Association.
- Kazdin, A. E., & Tuma, A. H. (Eds.). (1982). *New directions for methodology of social and behavioral sciences: Single-case research designs*. San Francisco: Jossey-Bass.
- Keith, K. D., & Schalock, R. L. (1992). The Quality of Life Questionnaire. *The Behavior Therapist, 15*, 106-107.
- Kiresuk, T. J., & Garwick, G. (1979). Basic Goal Attainment Scaling procedures. In B. R. Compton & B. Gallaway (Eds.), *Social work processes* (rev. ed., pp. 412-420). Homewood, IL: Dorsey.
- Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1984). *Interpersonal psychotherapy of depression*. New York: Basic Books.
- Korchin, S. J. (1976). *Modern clinical psychology*. New York: Basic Books.
- Kratochwill, T. R., & Piersel, W. C. (1983). Time-series research: Contributions to empirical clinical practice. *Behavioral Assessment, 5*, 165-176.
- Lambert, M. J. (1983). Introduction to assessment of psychotherapy outcome: Historical perspective and current issues. In M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 3-32). New York: Wiley.
- Maletzky, B. M. (1980). Self-referred versus court-referred sexually deviant patients. *Behavior Therapy, 11*, 306-314.
- Maletzky, B. M. (1991). *Treating the sexual offender*. Newbury Park, CA: Sage.
- Marquis, J. N. (1991). A report on seventy-eight cases treated by eye movement desensitization. *Journal of Behavior Therapy and Experimental Psychiatry, 22*, 187-192.
- Nelson, R. O. (1981). Realistic dependent measures for clinical use. *Journal of Consulting and Clinical Psychology, 49*, 168-182.
- Nelson, R. O., & Hayes, S. C. (Eds.) (1986). *Conceptual foundations of behavioral assessment*. New York: Guilford.
- Newman, F. L. (1983). Therapist's evaluation of psychotherapy. In M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 498-536). New York: Wiley.
- Nezu, A. M., & Nezu, C. M. (Eds.) (1989). *Clinical decision making in behavior therapy: A problem-solving perspective*. Champaign, IL: Research Press.
- Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change* (pp. 101-165). New York: Academic Press.

- Payne, R. W., & Jones, H. G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13*, 115-121.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
- Salter, A. (1949) *Conditioned reflex therapy: The direct approach to the reconstruction of personality*. New York: Creative Age Press.
- Shapiro, M. B. (1957). Experimental method in the psychological description of the individual psychiatric patient. *International Journal of Social Psychiatry, 3*, 89-102.
- Shapiro, M. B. (1964). The measurement of clinically relevant variables. *Journal of Psychosomatic Research, 8*, 245-254.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century.
- Skinner, B. F. (1957). The experimental analysis of behavior. *American Scientist, 45*, 343-371.
- Sorenson, R. L., Gorsuch, R. L., & Mintz, J. (1985). Moving targets: Patients' changing complaints during psychotherapy. *Journal of Consulting and Clinical Psychology, 53*, 49-54.
- Stricker, G. (1992). The relationship of research to clinical practice. *American Psychologist, 47*, 543-549.
- Stricker, G., & Keisner, R. H. (Eds.) (1985). *From research to clinical practice: The implications of social and developmental research for psychotherapy*. New York: Plenum.
- Watson, R. I. (1951). *The clinical method in psychology*. New York: Harper & Row.
- Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford, CA: Stanford University Press.
- Yin, R. K. (1984). *Case study research: Design and methods*. Beverly Hills, CA: Sage.

RECEIVED: May 12, 1992

FINAL ACCEPTANCE: August 31, 1992