Assessing Social Validity in Clinical Treatment Research
Issues and Procedures

Sharon L. Foster
California School of Professional Psychology–San Diego
Eric J. Mash
Department of Psychology University of Calgary

ABSTRACT

Social validity is a term coined by behavior analysts to refer to the social importance and acceptability of treatment goals, procedures, and outcomes. This article discusses dimensions of social validity, methods used to evaluate various aspects of social validity, and the applicability of these concepts and methods in clinical treatment research.

Correspondence may be addressed to Sharon L. Foster, California School of Professional Psychology–San Diego, 6160 Cornerstone Court East, San Diego, California, 92121. Electronic mail may be sent to sfoster@mail.cspp.edu
Received: August 24, 1998
Revised: December 15, 1998
Accepted: December 17, 1998

The concept of "social validity" has its roots in applied behavior analysis, a field that uses Skinnerian principles to study methods for producing changes in observable behavior (Baer, Wolf, & Risley, 1968). Since its inception, applied behavior analysis eschewed "traditional" measures and methods of assessing client change by focusing on single individuals, repeatedly and objectively measuring observable behavior, and using controlled single-organism designs to assess treatment outcomes. By the late 1970s, behavioral journals were filled with articles showing the effects of various procedures on a variety of well-defined and carefully observed behaviors.

At the same time, anecdotal reports of concerns about the meaningfulness of the targets of intervention and the difficulty of implementing treatment procedures began to emerge (Wolf, 1978). Two seminal articles, the first by Wolf (1976, 1978) and the second by Kazdin (1977), responded to these concerns and defined the concepts now considered under the rubric of social validity. These articles outlined two major issues for behavior analysts. The first involved the need to show that an intervention will be accepted and viable if implemented in a community setting (Schwartz & Baer, 1991). The second charged clinical scientists with documenting the social importance of treatment goals and outcomes (i.e., showing that treatment addressed one or more meaningful or important problems in the client's life and that it produced clinically important changes in those problems). A third important article, by Van Houten (1979), underscored and extended these points. Most current approaches (e.g., Gresham & Lopez, 1996) follow Wolf's, Kazdin's, and Van Houten's leads and define three distinct but related elements of intervention that can be assessed for their social validity: (a) the goals of treatment, (b) the treatment procedures, and (c) the
outcomes produced by treatment procedures. Goals can be assessed for both their importance (i.e., what justifies working toward the particular treatment goals?) and their acceptability (i.e., does "society" find the treatment goal to be worthwhile or desirable?). Treatment procedures are usually assessed solely for acceptability, whereas outcomes are assessed for their social importance (i.e., does the degree of client change represent an important improvement for the client?).

Kazdin (1977) and Wolf (1978) also proposed two methods for evaluating social validity. Wolf (1978) called specifically for the use of subjective evaluation to gauge the social acceptability and applied importance of goals, procedures, and outcomes, arguing that subjective evaluations complement more objective measurement and that "social importance' [is] a subjective value judgment that only society [is] qualified to make" (pp. 206—207). Kazdin focused specifically on treatment outcomes, echoing Wolf's view that subjective evaluations could provide quantitative data that reflect qualitative judgments about client behavior and capture global judgments of the client's performance or impact on others. Kazdin also mentioned normative comparisons (see also Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999) as an alternative social validation approach for examining the importance of outcomes. Most research assessing social validity in behavior analytic research has conformed to these methodological dictates (see Armstrong, Ehrhardt, Cook, & Poling, 1997; Kennedy, 1992; and Storey & Horner, 1991, for reviews), despite occasional calls for extensions of social validation methodologies (e.g., Hawkins, 1991; Winett, Moore, & Anderson, 1991).

Although the term social validity is most commonly used by behavior analysts, the concepts underlying it are not unique to behavioral approaches. Discussions in other literatures use terms like clinical importance, applied importance, clinical significance, qualitative change, educational relevance, ecological validity, and cultural validity to explore issues related to the viability of treatment procedures and the importance of client change. Deliberations about the relevance of traditional modes of therapy for ethnically diverse clients raise important social validity issues. Changes in service delivery models related to managed health care have also played a role in bringing social validity into focus in an environment in which social "managers" often determine whether an intervention is acceptable and reimbursable. Recent emphases on the importance of distinguishing between quantitative measures of change in protected environments (such as university research settings) versus the social significance or importance of change in a real-world context (e.g., Weisz, Donenberg, Han, & Weiss, 1995) also complement points raised by Wolf and Kazdin 2 decades ago.

The purpose of this article is to describe briefly the rationale for each of the three components of social validity noted above (i.e., treatment goals, procedures, and outcomes) and to examine methods used to evaluate each, emphasizing the two traditional methods used to assess social validity (normative comparisons and subjective evaluations). We focus primarily on the role of social validity assessment in treatment—outcome studies, although many issues apply to clinical practice and prevention settings as well. Within each area, we address two different but related goals of social validation: (a) demonstrating that goals or outcomes are meaningful and important in the client's life and (b) addressing the social acceptability of the goals, procedures, or outcomes to relevant "consumers" of the intervention.

Before turning to our discussion, a few general points are in order. First, examining the social validity of a particular treatment is not a substitute for examining treatment efficacy in other ways (Schwartz & Baer, 1991). Rather, social validity assessment should supplement outcome measures that assess changes in client functioning in areas directly targeted by treatment. Second, although social validity has sometimes been conceptualized as an outcome, others (Finney, 1991; Schwartz & Baer, 1991; Winett et al., 1991) have argued that it is best thought of as a process. That is, examining whether a goal or an outcome is important,
or a treatment is acceptable, can occur at various points in time—as a program is being developed, during early pilot phases, and after wide-scale dissemination. Evaluations can also occur during different points in the treatment process—at the beginning, while treatment is ongoing, immediately after termination, and at follow-up (Fawcett, 1991). When and how to assess social validity depends on whether the purpose of the assessment is to develop, improve, or appraise the intervention; whether the aim of the assessment is to examine goals, procedures, or outcomes of treatment; and whether the investigator wishes to evaluate the acceptability of the treatment procedures and goals or the importance of treatment goals and outcomes.

Finally, social validity has typically been discussed as though it were a single construct that could be scored dichotomously—an intervention either "has" or "lacks" social validity. This tendency oversimplifies the construct of social validity in at least two ways. First, social validity is a multidimensional construct that can be broken into at least two other general constructs (acceptability and importance), which themselves are multidimensional in nature. Assessments of acceptability and importance may or may not converge: A trivial treatment goal, for example, may be acceptable to clients but not important for improving their psychological well-being. In addition, different components of treatment can differ from each other in their acceptability or importance. For example, ineffective treatment procedures may be acceptable to clients, but the poor outcomes they produce may not. A second oversimplification is the assumption that different dimensions of social validity can be scored dichotomously, as either present or absent. At this point, no well-established criteria exist for determining particular cutoff points for what constitutes "enough" social validity. Therefore, social validity should be conceptualized as a series of continua that reflect its different dimensions rather than as a unidimensional dichotomy (Wolery & Gast, 1990).

The Importance and Acceptability of Treatment Goals

Goals of treatment are crucial because they define the nature of the core difficulties treatment is designed to address, dictate the content of measures for assessing outcome, and provide links to literatures that may provide the conceptual and empirical foundations of the intervention. Social validation approaches can examine two aspects of treatment goals: their social acceptability to various consumers of the intervention and their importance.

Rosen and Proctor (1981) differentiated types of treatment outcomes. Their distinctions are relevant to treatment goals and have implications for the type of social validity assessments that are appropriate. Ultimate goals are the problems that lead a client to seek treatment: Their accomplishment, at least in theory, is necessary and sufficient for termination and its appraisal as successful. Ultimate goals are most often determined by or in collaboration with the client, often comprise the client's presenting problems, and by definition are important to client. Instrumental goals, in contrast, can be defined as desired outcomes that lead to ultimate outcomes without further intervention. These are determined by the therapist, who hypothesizes that accomplishing these goals is necessary to accomplish the ultimate goals. Instrumental goals are often theory- or data-driven, whereas ultimate goals should not vary greatly or depend on the theoretical orientation of the therapist, and are generally determined by clients or by surrogates who act on behalf of clients (e.g., staff who care for individuals with intellectual disabilities, parents). For example, social skills training approaches assume that teaching a child prosocial skills (instrumental outcomes) will lead a child to make and keep friends (ultimate outcomes). Many couples therapists assume that teaching a couple better communication and problem-solving skills (instrumental outcomes) will result in greater subjective satisfaction with the relationship (ultimate outcome).

One key distinction between ultimate and instrumental goals lies in how they are established and assessed.
Specifically, ultimate goals should be defined and assessed by the client or a significant individual in the client's life. Instrumental goals, in contrast, should be defined in terms of their causal or predictive relationship to the ultimate goals. This distinction has important implications for social validity questions about how to establish the importance of treatment goals—a point elaborated below as we examine the use of normative comparisons and subjective evaluations of goals as ways of examining the importance and acceptability of treatment goals.

It is also important to recognize that goals that are expressed in terms of a client's problems often have counterparts—adaptive goals to increase positive aspects of a client's functioning or environment (Evans, 1993). Put another way, every undesirable behavior or state of affairs that the client wishes to decrease in treatment has a counterpart behavior or set of circumstances that should be increased to take the place of the problem. Examining the acceptability and importance of both sets of goals is important, as these cannot be assumed simply to function in inverse ways. Consider, for example, contemporary debates about adolescent sexual behavior. Although we suspect that most of the general public would view decreasing risky adolescent sexual behavior as a desirable goal, specific groups disagree on which behaviors are desirable and should be promoted: increased condom use or increased abstinence.

**Normative Comparisons**

With the normative comparison approach, levels of a target problem or deficit are compared with normative data (Kendall & Grove, 1988; Kendall & Norton-Ford, 1982). Levels that fall outside the "normal" range are used to justify the importance of an ultimate or instrumental goal. Although a true normative approach would compare an individual's performance with that of a randomly selected community sample, more often investigators use "local micronorms" for these comparisons (Gresham & Lopez, 1996). Local norms are established by collecting data on individuals in the client's immediate environment. The investigator then either compares the client's scores with these data or uses these data to select participants for the intervention. For example, Bierman and Furman (1984) observed children's conversational skills and collected data on children's acceptance by peers. They then selected the children with the poorest peer acceptance and worst skills (compared with others in their classrooms) to participate in a social skills training program.

Normative comparisons are most useful when a problem occurs along a continuum and deviation from normative levels has been shown (through empirical means) to be problematic (see Kendall et al., 1999, for an extended discussion of normative comparisons). Under these conditions, the normative comparison approach can provide some justification that the client's difficulties warrant treatment. Nonetheless, the normative comparison approach does not directly address (a) the social acceptability of treatment goals (which is better addressed through evaluations by some segment of the population), (b) the distress caused by target problems to the client or to others in the client's environment, or (c) whether the nonnormative levels of performance impair the individual's day-to-day functioning.

A number of practical and conceptual issues also limit the use of normative data for establishing the importance of ultimate goals. First, appropriate normative data are not always available (Kazdin, 1977). Second, setting a cutoff point for what constitutes the "normal range" or boundaries of acceptable performance may be difficult (Storey & Horner, 1991). In addition, the normative comparison group should be selected with care and should be representative of members of the environment in which the client is asked to function (Storey & Horner, 1991). For example, comparing a developmentally delayed youngster's social performance with that of normally developing same-sex children might be preferable to comparison with other delayed children if the child was placed in a mainstreamed classroom.
Collecting local norm data involves additional issues. This procedure requires informed consent by those who provide data for the normative database, and the effort involved can be prohibitive when the investigation involves a large number of clients and time-consuming assessment procedures. The local norms approach also requires that investigators establish criteria for selecting individuals who will contribute the local norms, determine the number of individuals necessary to provide an adequate sample, and decide whether these individuals will be "average" or "exemplary" performers (Van Houten, 1979). Furthermore, measures used to select members of the normative group should themselves be reliable and valid (Storey & Horner, 1991).

Some have also criticized the use of norms on conceptual grounds. Fuqua and Schwade (1986) have pointed out that some treatment goals can be justified on ethical or legal grounds, regardless of their normative status. These might include decreasing sexual harassment or physical abuse or improving independent living skills of individuals with developmental handicaps in the interest of placing them in less restrictive environments. In addition, because a behavior is normative does not guarantee that it is adaptive (e.g., some unsafe sexual practices may be normative among sexually active adolescents but nevertheless maladaptive). Similarly, many nonnormative behaviors may not be particularly maladaptive. For example, Asher, Markell, and Hymel (1981) criticized interventions that attempted to increase rates of interaction in children who interacted relatively infrequently with their peers. Asher et al. argued that low rates of interaction per se were empirically unrelated to other indicators of peer-related difficulties and that improvements in rates of interactions did not lead to improvements in other domains related to peer acceptance. Finally, normative performance may not provide an indication of "ideal" or "competent" levels of performance (Van Houten, 1979).

Normative levels are also insufficient for establishing the importance of instrumental goals. Unlike ultimate goals, which are defined as important because they cause the client (or someone in the client's environment) sufficient distress to seek treatment, instrumental goals lead to accomplishment of the ultimate goals of treatment. Thus, the importance of these goals is best determined empirically by linking them longitudinally and experimentally with ultimate goals (see Winett et al., 1991, for a discussion of this issue from an epidemiological perspective). When different instrumental goals can produce changes in the ultimate goals, these instrumental goals should be selected by considering the relative effectiveness, efficiency, and social acceptability of the alternatives for accomplishing the ultimate goals.

Subjective Evaluations

The second approach to examining the goals of treatment lies in having some consumer of treatment services evaluate one or more goals for either their importance or their acceptability. As an alternative, sometimes subjective evaluation is used in formulating an intervention and results in no formal data. In this case, the investigator uses input from clients or relevant community members to construct the goals of treatment rather than to document the importance or acceptability of already-established goals. The assumption behind this approach is that it builds social importance or acceptability into treatment goals by involving relevant members of the social community. Examples include (a) screening out clients who do not report significant personal distress about the problem being investigated or who do not display impairments in functioning (as is required by many Diagnostic and Statistical Manual of Mental Disorders diagnoses) and (b) interviewing members of an inpatient staff about disturbing client behaviors that should be addressed during treatment. Another example is Cone's template matching approach (Cone & Hoier, 1986). With template matching, the investigator collects data from potential consumers of the client's behavior. These data specifically assess the performance expectations the client will be asked to meet in the situations or
environments in which the client wishes to function. The investigator then uses this information to create a "template" of desirable and undesirable behaviors for particular environments or situations, which become the targets of assessment and treatment. For example, possible playmates of rejected children could be interviewed about the behaviors they like and dislike in play situations. The therapist could then assess the extent to which rejected children display these behaviors and develop a treatment plan for increasing deficiencies and reducing excesses (Cone & Hoier, 1986).

In other cases, subjective evaluation of goals involves formal data collection and can involve asking clients or others to evaluate potential treatment goals. This has been done most often with treatments involving children and individuals with disabilities, perhaps because these individuals frequently are not involved in selecting goals of treatment for themselves. For example, Isett, Roszkowski, Spreat, and Reiter (1983) asked staff members of group homes to rate how tolerable each of a series of maladaptive behaviors was when performed by individuals with mental retardation. Subjective evaluation can address either the acceptability or the importance of components of social validity, depending on the content of the assessment and the population of consumers who complete the evaluation. Questions that ask clients (or their significant others) about the personal importance, distress, or disruption caused by target problems can establish the importance of ultimate goals in seeking treatment. In contrast, questions that ask members of the community to rate the importance of treatment goals provide information about lay beliefs about goals and thus—despite their alleged focus on "importance"—may be most directly relevant to the acceptability of the goal to people in the client's social environment.

Subjective evaluations are generally not directly relevant to evaluating the importance of many instrumental goals, which—as just indicated—should be established through scientific investigation. Indeed, one could argue that asking clients or lay persons to evaluate the importance of instrumental goals is tantamount to assessing the extent to which their beliefs correspond with scientific knowledge. Procedures that ask experts in nonpsychological areas to identify and operationalize treatment goals, however, can be useful when scientific data are not available. For example, Plienis et al. (1987) asked teenagers to rate topics they liked to talk about, then built highly rated topics into their social skills training practice. Williams and Cuvo (1986) consulted written texts and observed individuals performing apartment upkeep skills in order to develop task analyses of skills required for home maintenance for individuals with disabilities. Subjective evaluations are, however, important in establishing the acceptability of instrumental goals to clients and others in the clients' environment: An ultimate goal may be acceptable, but instrumental goals may not be. In addition, subjective evaluations of instrumental goals may reveal circumstances in which clients (or others) do not believe that the instrumental goals will produce the ultimate goals and may highlight areas in which clients need educating about how the steps of therapy are likely to lead to resolution of the difficulties that brought the client to treatment.

One key issue in using subjective evaluations by consumers lies in determining the relevant consumer of the mental health services. Schwartz and Baer (1991) described four consumer groups, each with a stake in the outcome of treatment. The first group, direct consumers, directly receives the intervention and includes the client and members of the client's family or environment (e.g., parents, spouse, and teacher) who are directly involved in the treatment. Indirect consumers, a second group, may pay for the treatment or be affected in some important way by the treatment. These consumers might include family members, friends, third-party payers, or coworkers who do not participate in the treatment but who would be strongly affected by the treatment outcome. A third group, members of the immediate community, interacts with the direct consumers on a regular basis but is less involved than are indirect consumers (e.g., neighbors, colleagues, or classmates). The final group, members of the extended community, includes individuals not acquainted with the client but who live in the same community.
Some consumer groups are more relevant to particular social validity issues than to others. For example, evaluating the importance of the ultimate goals of treatment might be best assessed by identifying the set of consumers that is most affected by the client's difficulties (usually direct and indirect consumers) and asking them to appraise the importance of the goal for the client. On the other hand, if subjective evaluation is used to gauge the acceptability of treatment, the investigator should identify who in the community will facilitate or impede the implementation of the treatment or program in the natural environment (Schwartz & Baer, 1991). Evaluations by potential clients, by significant others, by professionals and staff who are likely to implement the treatment, and by vocal spokespersons in the community may all be important in whether a treatment is likely to be implemented or accepted.

A number of caveats have been voiced about using subjective evaluations to assess the importance and acceptability of treatment goals. First, subjective opinion may or may not correspond with actual behavior (Hawkins, 1991; Schwartz & Baer, 1991): People may not take action to promote, impede, or enroll in an intervention, even when they strongly like or dislike the goals of the program. Second, the subjective opinion of community members in general is ill suited to determining whether the goal of the intervention is significant for the client or the client's mental health (Gresham & Lopez, 1996; Hawkins, 1991; Kazdin, 1977; Lloyd & Heubusch, 1996) because the former requires a judgment by the client or someone else and the latter is a professional judgment. Members of the community may or may not be well-informed and may base judgments on a host of factors other than relevant information. This is particularly true when subjective evaluation is used to judge the importance of instrumental goals, which are better determined by expert professionals using their knowledge of the scientific bases of the client's difficulties and their professional expertise. Third, investigators should select the consumer group carefully so as to avoid "stacking the deck" in favor of positive evaluations of the intervention goals (Wolery & Gast, 1990) and should clearly describe their sampling procedures. Fourth, a strength of community judgments is the fact that they are contextually and historically embedded. However, as Schwartz and Baer (1991) pointed out, sometimes community opinions may conflict with the opinions of the professional community about social importance. They cited prejudicial views as examples. In these cases, some argue, social validity information may signal a need to alter community opinions and standards. Such information could, nonetheless, be quite important in anticipating resistance to the treatment procedures by at least some members of the community.

Finally, a major purpose of collecting goal acceptability ratings is to assess whether the goals of treatment fit community values. But defining the "community" is by no means easy; society is populated with identifiable and diverse subgroups, some of which have diametrically opposed viewpoints on the social importance and acceptability of treatment goals. One example lies in viewpoints about the acceptability of attempting to help individuals who are gay or lesbian to change their sexual orientation—a goal that might be strongly endorsed by some religious groups but equally strongly opposed by gay rights advocates and by many psychologists (e.g., Davison, 1976).

Assessments of goal acceptability should clearly identify the consumer group to be assessed and why the group is relevant to the assessment issue. Investigators should be particularly sensitive to these issues when treatments are used with new or more diverse populations than the groups in which an intervention was originally studied—a situation in which social validation is particularly important. When goals are evaluated by members of the community other than the client or a clearly identified important member of the client's environment (e.g., parent, teacher, or staff member), the investigator should also follow procedures designed to ensure representative sampling of that group and provide descriptive information about characteristics of the group. Finally, of course, instruments used to assess goal importance or acceptability should pass
psychometric muster (i.e., have solid evidence for their content validity as well as reliability and other forms of validity).

**Functional Impairment and Adaptation**

Because of the limitations of subjective evaluations, Hawkins (1991) has suggested that investigators adopt an alternative approach for examining the importance of treatment goals: to examine their contribution to functional performance or impairment in the client's life. That is, treatment goals should be selected because they maximize positive short- and long-term outcomes and minimize costs for the client and those in the client's environment, something Hawkins called *habilitative validity*. One major method for establishing habilitative validity involves demonstrating the linkages between treatment goals and adaptive or maladaptive outcomes. Although these notions are appealing, they are not without difficulties, including the need to operationalize and assess important short- and long-term outcomes associated with treatment goals and the need to reconcile circumstances in which a particular goal produces conflicting outcomes, some negative, some positive. An additional problem lies in the fact that one cannot necessarily assume a priori whether many outcomes are "adaptive" or not, and thus the outcomes themselves may need investigation for their social validity.

Kazdin (1997) and Kazdin and Kendall (1998) recently recommended that treatment—outcome studies with children devote greater attention to client functioning in daily life requirements and activities, echoing similar recommendations voiced by Hoagwood, Jensen, Petti, and Burns (1996). Indeed, many of these authors implicitly nominated a variety of areas as universally meaningful and important for clients, regardless of whether these constitute the ultimate or instrumental goals of treatment. These include performance in role demands in one's work or school environment, restrictiveness or access to various settings that permit autonomy and development, quality of family and peer relationships, degree to which the client's difficulties limit access to conditions that promote adaptation and growth (e.g., social support), burden of care for family members, service use related to the client's problem (e.g., inpatient treatment or need for special services), and monetary costs of the client's problems (e.g., disability payments or treatment and service costs).

Of course, no single study can assess all of these, nor is each equally relevant for all of the myriad types of client difficulties. In addition, the content validity of instruments used to assess important domains of functioning should be examined carefully for information to justify the operationalization of these intuitively appealing but often hard-to-define concepts. The importance of the functional impairment and adaptation focus, however, is that it goes beyond normative comparisons and subjective evaluation to underscore the importance of examining treatment goals that address important domains of functioning, or alternatively,of linking goals empirically to adaptive or maladaptive outcomes.

**Acceptability of Treatment Procedures**

Assessments of treatment acceptability are part of the broader issue of viability of treatment and are particularly important as researchers transport their interventions to community settings and extend treatment applications to diverse populations within and outside of the United States. In fact, recent American Psychological Association (APA) guidelines for developers of psychological interventions (Task Force on Psychological Intervention Guidelines, 1995) explicitly include issues relating to social validity in their second "clinical utility" axis. Specifically, the clinical utility axis relates to evaluations of "the ability (and willingness) of practitioners to use, and of patients to accept, the treatment in question, and to the range of applicability of
that treatment” (Task Force on Psychological Intervention Guidelines, 1995, p. 13). Investigators have assessed treatment acceptability in two main ways: by using experimental analogue studies and by using consumer satisfaction questionnaires completed by clients.

### Experimental Analogue Studies

One common way of assessing treatment acceptability has involved presenting students, potential clients, members of the public, or mental health workers with hypothetical vignettes depicting client problems and possible treatment procedures, then asking raters to evaluate the plans using a rating scale. Much of this research has focused on components of behavioral interventions, such as behavioral contracts, reinforcement procedures, and time-out, although a few investigations have included nonbehavioral interventions, such as medication and paradoxical intervention (see Calvert & Johnston, 1990; Elliott, 1988; Miltenberger, 1990; Reimers, Wacker, & Koeppel, 1987; and Storey & Horner, 1991, for reviews).

Investigators have used several scales to assess treatment acceptability in these analogue investigations. Among the most popular is Kazdin’s 15-item Treatment Evaluation Inventory (TEI; Kazdin, 1980); a shortened 9-item version is also available (Kelley, Heffer, Gresham, & Elliott, 1989). A second instrument, the Intervention Rating Profile (Witt & Martens, 1983), contains 20 items. Several variations of this instrument have also been used. These include a 15-item version (Martens, Witt, Elliott, & Darveaux, 1985), a version for children (Elliott, 1986), and a longer version with additional items that assess assumed effectiveness of the intervention (the Behavior Intervention Rating Scale) (Von Brock & Elliott, 1987). A number of studies have also used Semantic Differential scales to evaluate treatment acceptability, but these scales have less face validity than the TEI and other instruments, the content of which directly addresses questions such as whether the respondent finds the procedures acceptable, whether the procedures require too much effort, and whether the procedures are suitable for the client population.

Although treatment acceptability studies have shown some consistency in their findings (Elliott, 1988; Miltenberger, 1990; Reimers et al., 1987), these studies have several limitations in their external validity. First, many do not use client populations to evaluate treatment acceptability; thus, their findings may be more relevant to general community reactions to treatments than to the reactions of potential clients. Second, generalizability from analogue to real-world settings remains an empirical question with these studies (Calvert & Johnston, 1990). Although ratings are assumed to reflect a client’s willingness to enter a particular treatment for a particular problem, verbal reports of treatment acceptability do not guarantee that clients would actually seek or accept the treatment if they needed it or, conversely, that they would object to its being offered in their community. Third, studies have focused almost exclusively on behavioral procedures used with children and individuals with developmental disabilities, perhaps because these populations are not considered capable of true informed consent and thus must be treated with particular care because of the need to preserve these clients’ well-being. This has led to limited attention to treatment acceptability of interventions for adults and adolescents, however. Fourth, research has devoted very little attention to nonbehavioral interventions. Fifth, analogue assessments generally present information about treatment as it might be presented to the client at the outset of treatment. However, in many clinical settings the clinician proposes a treatment plan after assessing the client’s difficulties and provides a rationale that links the treatment to the assessment results. Indeed, analogue studies have shown that the type of rationale (Cross-Calvert & McMahon, 1987), the way the investigator indicates that the treatment will be implemented (Kazdin, 1980), and whether the treatment is determined collaboratively versus unilaterally (i.e., by either the client or the clinician; Kusick, Gutkin, & Witt, 1991) can influence acceptability ratings of treatments for children. Sixth, most interventions in analogue studies are examined in isolation, and findings about them may not generalize to treatments that use combinations of interventions.
Finally, treatment acceptability may vary over time, particularly as clients experience the nature and results of treatment procedures. Reimers, Wacker, Cooper, and De Raad (1992) examined this issue using a modified version of the TEI. Parents receiving behavior management services for their children in an outpatient clinic completed the inventory four times: immediately after hearing the therapist's recommendations of positive contingency systems for use with their child and 1, 3, and 6 months after their initial clinic visits. They also rated their child's behavior on a behavior problem questionnaire and, during the last three assessments, rated the degree to which they had complied with the therapist's recommendations. Interestingly, acceptability ratings were highly consistent over time ($r_s = .90$ between 1 month and 3 months, .89 between 3 months and 6 months). In addition, acceptability was strongly related to change in behavior problems, both concurrently and longitudinally, although the use of parent report to assess both acceptability and behavior problems introduced a method variance confound in these results. Finally, cross-lag panel correlations indicated that compliance at 1 month predicted acceptance at 3 months. In contrast, contrary to clinical supposition, acceptance did not predict compliance longitudinally. This finding, however, is seriously limited by the single-item unvalidated self-report measure of compliance. Nonetheless, this study underscores the importance of looking at acceptability over time in clinical populations and of relating client views to other important dimensions of therapy process and outcome.

As with any evaluation relevant to social acceptability, the investigator must establish the target population(s) in studies of treatment acceptability (Schwart & Baer, 1991) and should sample that population adequately. Many different people control whether a treatment can and will be implemented successfully in community settings. These include clients, significant others involved in the treatment, the individuals who will implement and supervise the treatment (e.g., teachers, inpatient staff, nurses, or clinicians), directors of services, and funding sources.

A second important issue involves determining what aspect of "acceptability" should be addressed. Although investigators have typically operationalized acceptability by asking respondents to make global judgments (such as how appropriate the procedure is, how disruptive the procedure might be to daily life, and whether they believe it will produce desirable effects), acceptability can be examined in other ways as well. Lennox and Miltenberger (1990), for example, suggested that evaluations of treatment acceptability assess the acceptability of side effects of treatment, potential for treatment strategies to be abused (e.g., aversive procedures or pharmacological agents), compatibility of the treatment with legal and regulatory mandates, staff cooperation, staff (or clinician) competence with the procedures, treatment efficiency, and cost-effectiveness evaluation. Although one could debate whether these variables directly assess acceptability per se, each is clearly relevant to the broader issue of treatment viability or feasibility and points to the multidimensional nature of the treatment acceptability construct.

**Consumer Satisfaction**

Another method of assessing treatment acceptability has relied on self-report questionnaires clients complete, ordinarily after treatment. Often called consumer satisfaction measures, these questionnaires generally ask clients to rate various aspects of treatment on some sort of global rating scale (see Lebow, 1982, and McMahon & Forehand, 1983, for reviews). Examples include the Client Satisfaction Questionnaire (Larsen, Attkisson, Hargreaves, & Nguyen, 1979) for adult treatment and the Therapy Attitude Inventory (Brestan, Jacobs, Rayfield, & Eyberg, 1999) for parent training. Reimers et al. (1992) also used a modified version of the TEI with clients in a clinic setting.

Several factors are important to consider in assessing client satisfaction with treatment procedures. First,
clients can be satisfied or unsatisfied with (a) the degree of change in ultimate or instrumental goals, (b) the
treatment procedures, and (c) the therapist. These are conceptually distinct elements of treatment
acceptability. Brestan et al. (1999) factor analyzed the Therapy Attitude Inventory completed by parents at
the end of parent training and found two factors: one that contained items assessing improvement in child
behavior, the other principally assessing ratings of the strategies used in treatment.

Second, satisfaction can be evaluated at various levels of analysis (Fawcett, 1991). At the molar level, for
example, clients can be asked about their satisfaction with treatment as a whole, whereas a more molecular
analysis might ask about client satisfaction with treatment elements, such as intake, assessment, and
homework. As clinicians and investigators place greater emphases on combined treatments (Kazdin,
1996), global satisfaction measures may obscure which elements of combined treatments contribute most to
the variance in satisfaction ratings. On the other hand, ratings of specific elements of combined or multimodal
treatments may not reflect their impact on satisfaction or acceptability when treatments are presented as a
package. Furthermore, ratings of specific treatment elements taken long after the fact (e.g., ratings of intake
taken at termination) may be influenced by client experiences throughout the intervention.

Finally, consumer satisfaction measures should be subjected to the same psychometric scrutiny afforded
other outcome measures. Thus, they should demonstrate test—retest reliability, internal consistency,
convergent and discriminant validity, and the like. This is particularly true if the investigator wishes to report
data for single items, which are notoriously unreliable.

Two other methodological issues are important to consider in interpreting the results of clients’ posttreatment
ratings of treatment acceptability. First, often only those who complete treatment participate in the
assessment; others may have "voted with their feet" by not initially accepting treatment or by dropping out
(McMahon & Forehand, 1983). Second, it is important to reduce possible demand characteristics by using
an assessor other than the therapist to conduct the assessment and by ensuring client anonymity.

**Additional Dimensions of Treatment Acceptability**

Although consumer ratings of various aspects of treatment acceptability provide some indication of the
viability of treatment, these are not the only data relevant to treatment acceptability. APA guidelines (Task
Force on Psychological Intervention Guidelines, 1995) nest acceptability within the broader domain of
treatment feasibility, which is the broader target of social validity assessment. These guidelines stipulate that
feasibility includes "factors such as patient acceptability of the intervention, patient compliance with the
requirements of the intervention, the ease of dissemination of the intervention, the ease of administration of
the intervention, and the cost of the intervention" (Task Force on Psychological Intervention Guidelines,
1995, p. 14). Other direct indicators of feasibility are relevant to this issue, including the number of eligible
clients who do not enroll in treatment when asked to do so, dropout and no-show rates, and data on
completion of therapy requirements (e.g., homework tasks). Schwartz and Baer (1991) described even
more indicators of poor program or treatment acceptance: "demands that program personnel leave ...
complain[ts] to friends, officials, and the media ... and negative affect and low enthusiasm during program
sessions" (p. 199). Gresham and Lopez (1996) focused on therapist indicators of treatment acceptability:
use of treatment and the fidelity with which the treatment is implemented. Although factors other than
treatment acceptability can contribute to all of these responses (Fuqua & Schwade, 1986), these behaviors
are important indicators of treatment viability and are particularly important pieces of information about
possible difficulties when exporting an intervention from research to community settings.

Finally, it is important to note that treatment acceptability is a dynamic and not a static variable. Advertising,
public relations campaigns, and consumer education programs are all predicated on the notion that individuals’ views of products can be improved or altered with information (see Schwartz, 1991, and Winett et al., 1991, for a more extended discussion of these issues). Research on the effects of education on treatment acceptability ratings in analogue tasks supports this as well (Miltenberger, 1990).

**Importance of Outcomes**

Demonstrating that a treatment produces a statistically significant change on some measure of outcome does not guarantee that the change is meaningful for the client or for the client's significant others. All of the articles in this series discuss methods for establishing the importance of treatment outcomes.

From a social validity standpoint, demonstrating meaningful change requires, first, that the investigator or clinician measure important aspects of the client's life (i.e., that the goals of treatment have some degree of importance). Selection of important goals is not sufficient for meaningful change, however. Second, the quantity or amount of change produced by treatment must be meaningful: Problems have declined or adaptive functioning has increased sufficiently to make a difference in some important aspect of the client's life.

As in other areas of social validity, normative comparisons provide one basis for examining level or degree of change after treatment. Earlier we described general concerns about the role of normative data in selecting treatment goals, and these apply here as well. Kendall et al.’s (1999) article addresses in detail additional issues in this area.

The second method of examining meaningful change has involved subjective evaluation by the client, by significant others in the client's environment, or by expert or representative judges. One way behavior analytic researchers have used this method involves filming or audiotaping a client's performance before and after treatment, then asking other individuals to provide global evaluations of the quality of the client's performance. Sometimes judges evaluate permanent products (e.g., writing samples to assess academic skill) instead of videotaped samples. Improved global judgments presumably provide evidence that important changes have occurred. For example, Kelly, Wildman, and Berler (1980) asked personnel managers to rate audiotapes of job interviews made by men with mild retardation before and after training in job interview skills. They found that the raters' evaluations mirrored observational indicators of behavior change and that both showed change from pre- to postintervention.

This approach is most appropriate for interventions that address highly specific behavioral skills (e.g., work-related, interactional, and self-help skills; see Storey & Horner, 1991, for a review of research with individuals with disabilities), when global performance evaluations are important indicators of performance quality, and when objective levels of acceptable skill performance are difficult to establish. Determining acceptable skill levels is particularly challenging when many different types or combinations of behaviors can be acceptable or when a myriad of factors interact to influence a judge's evaluation. For example, peers' global judgments of whether they would like to spend time with a client before and after social skills training may depend on a host of different factors, including voice tone, facial expression, eye contact, reciprocity of comments, and topical content of the client's remarks. In this case, subjective evaluation would provide an important supplement to observational data that indicated changes in conversational behavior.

A number of methodological issues are important to consider when using this approach. First, the behavior viewed on videotape must provide an adequate sample of the client's behavior. Specifically, the segments
should be long enough for the judge to make an informed evaluation (Storey & Horner, 1991) and should be representative of the client's typical performance. The latter could be enhanced by providing multiple examples of the client's performance, although this might strain the demands on those asked to judge the client. In addition, pre- and posttest tapes should be counterbalanced or randomly ordered, and raters should not be provided with potentially biasing information: Global evaluations of brief segments of observed performance are quite susceptible to the influence of information about client diagnoses, expected outcomes, and the like (Kent, O'Leary, Diament, & Dietz, 1974; Schuller & McNamara, 1976).

This approach is also limited for documenting the importance of outcomes when the goals of intervention are not directly observable (e.g., affective or cognitive), when the ultimate goal of treatment is broadly defined to encompass many aspects of the client's performance in a variety of settings (e.g., as with depression or antisocial behavior), or when the client declines the procedure. In these cases, questionnaires can be used to assess clients' (or others') judgments of the degree to which treatment has met its ultimate goals.

Regardless of the method used, several key issues are involved in using subjective evaluation to examine outcomes. These include questions about who should provide the evaluations, what the evaluators should be asked to assess, and how they should be asked to conduct their evaluation. As with evaluation of goals, the individual(s) chosen to evaluate outcomes should be selected on the basis of the relevance of their judgment to the type of outcome being assessed. Clients, for example, may be the best judges of their ultimate goals, assuming that the client entered therapy voluntarily and that the problems for which they sought treatment were distressing to the client. On the other hand, a significant other or some other consumer of the client's behavior (e.g., teacher or staff member) may be a good evaluator of ultimate goals when the client's behavior is troubling to others, as with many externalizing behavior disorders.

Social validation of instrumental outcomes is less clear-cut. Because instrumental goals should be selected on the basis of presumed or empirically established relationships with ultimate outcomes, theoretically client satisfaction with these outcomes is irrelevant. This is not to say that these outcomes should not be assessed, however. Indeed, this assessment is crucial for testing the linkages between ultimate outcomes and the processes by which treatment was presumed to work. For example, showing that ultimate goals improve when instrumental goals do not (or vice versa) should prompt an investigator to rethink the assumptions underlying the intervention. Improvements in both sets of goals, together with results to show that improvements in instrumental goals correlate with improvement in ultimate goals, support the supposed linkages between types of outcomes.

Evaluators can also be asked to judge a number of different dimensions of change. Clients or their significant others, for example, are often asked to rate client performance before and after treatment, usually using questionnaires that assess perceptions of symptoms or behaviors related to ultimate goals. These assess whether the client's behavior has changed. More relevant to social validation, however, are evaluations of whether the client (or significant other) is "satisfied" or "happy with" the outcome.

As with evaluations of treatment acceptability, evaluations of outcome should be specific to the focus of the evaluation. In other words, they should assess satisfaction with outcomes per se and not satisfaction with treatment procedures. Furthermore, when assessing significant others' evaluations of the client, investigators should remember that others' perceptions may lag behind genuine client behavior change. For example, Oden and Asher (1977) evaluated social skills training with children. Children's behavior improved significantly from pre- to posttreatment, but peer evaluations of how much they liked the client children (the social validity assessment) did not. By 1- and 6-month follow-ups, however, peer evaluations indicated that peers liked the treated children significantly more than the control children. Presumably, the peers needed a
good deal of experience with the changed behavior of the treated children before their global perceptions changed.

As in the evaluation of other aspects of treatment, the evaluation of satisfaction with treatment outcomes should include multiple informants whenever possible. From a conceptual standpoint, each informant has a unique viewpoint, and thus reports from different informants are not parallel forms of the same measure. Interinformant disagreement is not particularly problematic from this perspective: Differences in perceptions among various stakeholders are not measurement errors but need to be examined in their own right.

Multiple informants have methodological advantages as well, particularly for teasing out associations due to shared method (or informant) variance. This can be done when multiple informants each appraise many different aspects of treatment outcome. Lambert, Salzer, and Bickman's (1998) findings illustrate the potential value of this approach. They collected parent and adolescent ratings of (a) adolescent behavior problems (before and after treatment), (b) global ratings of satisfaction with treatment and clinic helpfulness (after treatment), and (c) global ratings of perceived improvement (after treatment). Interviewers also rated child functioning, and parents rated family difficulties they experienced because of the child's problems. Parent reports of symptom reduction correlated significantly with parent measures of satisfaction ($r = .38$), as did teen reports with teen satisfaction ($r = .28$). A multitrait—multimethod approach using structural equation modeling, however, showed that these constructs were unrelated when method variance was controlled. Furthermore, adding a method variance component significantly improved the fit of the model, indicating that correlations among measures of satisfaction and symptom change may be spuriously high when only a single informant's reports are considered. Finally, global parent appraisals of adolescent improvement made on a five-item scale related more closely to ratings of satisfaction than to other evaluations of symptom change.

As with ratings of treatment acceptability, consumer ratings of satisfaction with treatment outcome should be administered in ways that reduce demand characteristics. In addition, as in other areas, psychometrically sound measures are required for adequate assessment. Finally, consumer satisfaction with outcome is only one perspective on whether the outcomes are valuable. Some might argue that client satisfaction with outcomes may be necessary, but not sufficient, for showing successful treatment. Measures that are not subjective in nature, but nonetheless speak to important outcomes (e.g., recidivism or rehospitalization), could address this issue.

In addition, therapy can produce meaningful improvements even if it does not completely impact "important outcomes." For example, a couple may divorce after couples therapy but do so amicably—an important outcome in light of the voluminous literature documenting the negative relationship between parental conflict and children's adjustment (Grych & Fincham, 1990). In addition, a client may not be completely satisfied with treatment outcome but nonetheless be substantially improved in important ways. In spite of these caveats, data on client satisfaction with treatment, along with information on variables related to functional adaptation and impairment (such as the number of clients who continue therapy after a particular treatment ends or continued need for special services for children), are important pieces of outcome data that should supplement measures of ultimate and instrumental outcomes.

Future Directions

Consumers of mental health services, corporations, and governments have become increasingly concerned with the cost-effectiveness of treatment. Although many agree on the need to measure the efficacy and
effectiveness of services, limited consensus has emerged about how this should be done, particularly in real-world settings. Moving treatment methods and assessment procedures from controlled research settings to "messy" community sites is necessary, but this is easier said than done. The concept of social validity addresses many issues relevant to these concerns. Although forged by clinical researchers working from an applied behavior analysis perspective, the concepts and procedures have direct relevance today for evaluators who assess the effectiveness of psychological therapies in general, as well as for evaluations of psychological services provided by managed care plans or providers.

At the broadest level, the concept of social validity addresses these issues by reminding mental health professionals that interventions are applied within a sociopolitical, cultural, and historical context and that what clients think about procedures and outcomes is important, perhaps because what clients think is important in its own right or perhaps because what clients think will predict how they respond to interventions. Furthermore, social validity issues highlight the need to evaluate rather than to assume that treatment goals, procedures, and outcomes accord with the personal values of the client and the general values of that portion of the community in which the intervention is implemented. Finally, the consumer's (client's) participation in decision making, in selecting and evaluating intervention goals, and in planning and evaluating treatment—which is often a part of social validity approaches—is consistent with current models of client empowerment and personal responsibility (Salzer, 1997).

The general ease with which the aims of social validity can be described masks a host of ambiguities, however. Incorporating views of important consumers of mental health services requires that psychologists determine which groups constitute important stakeholders and recognize that different subgroups may differ markedly in their perspectives and views. Furthermore, not all stakeholders have equal power in determining the importance and acceptability of treatment goals, procedures, and outcomes. Client perspectives are clearly necessary to assess, as clients are the ultimate beneficiaries of psychological interventions. One could argue, however, that the perspectives of those who decide whether goals and outcomes are worth their cost for service delivery are also crucial. This, in turn, leads to the issue of how these individuals make decisions about what degree of dysfunction warrants intervention—and whether the bases for these decisions are justifiable.

In addition, importance and acceptability are not static concepts. Because many dimensions of these constructs are contextually embedded, they can shift as social norms and views change. Education, too, can shift community views of what is and what is not appropriate in providing mental health services. Thus, generalizability of social validity indicators over time, as well as across communities and consumer groups, should be investigated rather than assumed.

The concept of social validity is also complex. It encompasses the global concepts of treatment importance and acceptability, each of which is multidimensional. Limited explicit discussion exists about how these concepts should be defined and operationalized. Furthermore, social validation as a concept and social validation as a set of procedures need to be distinguished more completely. Discussions of social validity also need to differentiate more clearly between social importance and acceptability of goals and outcomes for the individuals involved in treatment versus importance and acceptability in larger social groups. The development of an integrative conceptual model to guide research would greatly help in making these distinctions. In addition, social validity should be treated as a multidimensional construct described by continua rather than by dichotomies.

Despite these limitations, a few general recommendations emerge from this body of literature. First, clinical investigators justify the importance of their goals. Social validity approaches have little to say about the
importance of instrumental goals, which should be selected on the basis of their empirical relationship to the ultimate goals of treatment. The importance of ultimate goals can be justified in any or all of several ways: (a) the extent to which the goal (or its absence) negatively affects a person's life or relationships or interferes with living and acting in the community and learning adaptive skills; (b) the potential physical harm, discomfort, or sanctions a behavior causes for the individual and for others; and (c) legal or ethical rules that support the goal. A fourth method, normative or local norm comparisons, may also be useful for demonstrating the need for treatment but will be most believable either when one of the previous three criteria is also met or when nonnormative levels have been previously shown to be problematic empirically.

Second, clinical researchers should explicitly address the clinical significance of the outcomes of treatment. Criteria similar to those just suggested for evaluating goals can be used to assess the importance of outcomes by examining the degree to which the client's status on any of the variables just described has shifted as a function of treatment and whether the degree of change constitutes meaningful improvement. This requires that some sort of social validity assessment plan be built into treatment protocols from the outset, that the measures used to assess the importance of goals and outcomes meet contemporary psychometric standards, and that the measures address aspects of social validity particularly relevant to the target population and the goals of treatment. Setting appropriate ranges for scores on various measures that indicate what constitute levels of "adaptive functioning," "significantly improved performance," or "normative levels" will be a particular challenge in this regard and involves establishing cutoffs for variables that are generally continuous rather than categorical. This complex issue is further complicated by the need to take measurement error and possible regression toward the mean into account in setting cutoff scores (Hsu, 1995).

Third, clinical researchers should routinely assess variables relevant to treatment feasibility. These should include, but not be limited to, measures of how acceptable different aspects of treatment (including both instrumental and ultimate goals) are to clients. Ideally, these data could be collected throughout treatment (and not just at the beginning and at the end) and examined in relation to key process and outcome variables in treatment (e.g., enrollment, participation, compliance, and improvement). Treatment acceptability issues should also be examined among consumers of treatment other than clients, particularly with an eye toward identifying and removing barriers to implementation of treatments with demonstrated efficacy for the target problem and population. Ideally, this should be done during development, effectiveness trial, and dissemination phases of treatment.

It will also be important to move social validity research beyond investigations of behavioral treatments with children and individuals with disabilities; issues of importance and feasibility are clearly relevant to studies with adult populations as well. In so doing, investigators should also turn greater attention to the reliability and validity of measures used to assess various aspects of social validity. Reviews of the literature have routinely lamented the lack of psychometric scrutiny of measures used in this area (e.g., Fuqua & Schwade, 1986; Schwartz & Baer, 1991; Storey & Horner, 1991). Content validity is particularly important to consider: Are the measures asking the right questions about the right topics of the right respondents about the right stimuli? Information on interrater agreement or evaluator consensus, test—retest reliability, and convergent and discriminant validity are also important to establish for social validity measures.

Finally, when investigators use subjective evaluations to evaluate acceptability or importance of treatment goals, procedures, and outcomes, they should specify clearly who served as judges, how the judges were selected or recruited, and which communities they represent. Investigators should also collect information from a sufficient number of judges to provide a reasonable sample of judge opinion. Finally, investigators should clearly indicate the bases for selecting the particular population of judges: What is the judges' stake in treatment? Why are their views important to assess?
Although social validity assessment began with behavior analysts, the issues raised by Wolf, Kazdin, and others 20 years ago still resonate in contemporary treatment—outcome research. Routinely struggling with issues involving the acceptability and importance of treatment goals, procedures, and outcomes in clinical outcome research will help investigators both to explore how various aspects of social validity of interventions can be enhanced and to narrow the gap between research information and practice needs.

References


Wolery, M. & Gast, D. L. (1990). Re-framing the debate: Finding middle ground and defining the role of social validity. (In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 129—143). Sycamore, IL:

Sycamore.)


1

There are exceptions, of course. An individual may be mandated to participate in treatment, an individual with cognitive disabilities may not be capable of formulating goals, or a parent or teacher may set goals for a child. In these cases, the person formulating the goals is sometimes considered the client and often participates in determining the ultimate goals of treatment.

2

The importance of an ultimate goal to direct consumers can ordinarily be assumed if the investigator documents that the client presents with the particular problem that defines the ultimate goal of treatment (and therefore, by definition, considers it important enough for treatment).