JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1991, 24, 189–204 NUMBER 2 (SUMMER 1991)

SOCIAL VALIDITY ASSESSMENTS: IS CURRENT PRACTICE STATE OF THE ART?

ILENE S. SCHWARTZ AND DONALD M. BAER
UNIVERSITY OF KANSAS

The use of evaluative feedback from consumers to guide program planning and evaluation is often referred to as the assessment of social validity. Differing views of its role and value in applied behavior analysis have emerged, and increasingly stereotyped assessments of social validity are becoming commonplace. This paper argues that current applications of social validity assessments are straying from the point originally proposed for them. Thus, several suggestions for improving current social validity assessment are proposed, including (a) expanding the definition of consumers to acknowledge the variety of community members able and likely to affect a program's survival, (b) increasing the psychometric rigor of social validity assessments, (c) extending assessment to heretofore underrepresented populations, (d) implementing widespread application of well-designed social validity assessments, (e) increasing meaningful consumer involvement in the planning and evaluation of behavioral programs, and (f) educating consumers to make better informed programming decisions.

DESCRIPTORS: social validation, assessment, methodology, dissemination

In 1978, Wolf formally introduced the issue of social validity to the field of applied behavior analysis. Essentially, he noted how rarely the consumers of behavior-analytic programs had been queried about their acceptance of a program's procedures, goals, and personnel; he warned from experience that nonacceptance could precede disastrous consumer rejection of the programs; and he recommended careful future assessment of consumer satisfaction from that point of view (Wolf, 1978). Applied behavior analysis has progressed extensively since 1978 and, indeed, social validity has often been of paramount interest to scholars tracing the field's history and predicting its future.

Opinion about the contribution of social validity to the discipline, however, is far from unanimous. Social validity has been heralded by some as an important guide for the future of the field (e.g., Baer, Wolf, & Risley, 1987) and denounced by others as a detraction from the scientific nature of our research practice (e.g., Barrett, 1987). Despite these differences, social validity measures are becoming almost commonplace in the behavioral literature.

This paper discusses the current practice of social validity assessment, proposes elements that should be included in any state-of-the-art procedure for assessing social validity, offers a propaedeutic taxonomy of consumers to guide the construction and circulation of these assessments, and suggests strategies that may help shape current practice to produce more accurate assessments of the behaviors associated with social validity.

Preparation of this article was supported in part by grants from the U.S. Department of Education (HO24J80003) and the National Institute of Child Health and Human Development (HD03144).

Appreciation is extended to Julie Cross Hoko and Mary Todd for their assistance in preparation of this manuscript, and to Susan A. Fowler, Kathryn K. Ramp, Montrose M. Wolf, Edward K. Morris, and Deborah Altus for their helpful comments on earlier versions of this paper. Requests for reprints should be addressed to Ilene S. Schwartz, Juniper Gardens Children's Project, 1614 Washington Blvd., Kansas City, Kansas 66102, or to Donald M. Baer, Department of Human Development, University of Kansas, Lawrence, Kansas 66045.

WHAT IS SOCIAL VALIDITY?

The purpose of social validity assessments is to evaluate the acceptability or viability of a programmed intervention. Most often, social validity assessment is accomplished by asking people other than the program planners or experimenters to complete some type of questionnaire (Kazdin, 1977; Wolf, 1978). The point of these assessments is to anticipate rejection of a program before that happens; therefore, the assessments should involve all

relevant consumers of the program. The consumers should be queried about the acceptability of the program goals, methods, personnel, outcomes, and ease of integration of program components into the consumers' current life-style. Then, this information should be used immediately, as well as in future program planning, implementation, and evaluation. Thus, social validity assessments are not meant to be, and indeed are not, primary dependent measures; logically, they can only supplement the objective measures of behavior that are the primary dependent measures (cf. Barrett, 1987; Dietz, 1978; Michael, 1980; Pierce & Epling, 1980; Scheirer, 1978).

Social validity assessment is ideally a two-part process: first an accurate and representative sample of the consumers' opinions is collected; then that information is used to sustain satisfactory practices or effect changes in the program to enhance its viability in the community. Failure to implement either part of this process undermines the social validity concept, and often will also discredit the social validity assessment. A useful social validity assessment should not (a) assess the opinions of a limited (or the wrong) community, (b) incorrectly assess the opinions of the relevant community, or (c) assess the opinions of the relevant community correctly, but not use that information to change the program and/or its consumers' opinion about it. In short, it is as inappropriate to sample criticism and praise and not respond to it as it is to respond to it without investigating its sources and its correlates (i.e., its accuracy).

Canvassing consumers for feedback on proposed and ongoing programs may establish closer ties between the community and the program; this may give a program's audience a sense of shared control, which in turn may circumvent complaints and program attrition (Baer et al., 1987; Giordano, 1977; Lebow, 1982). It is vital, though, to make this input functional by incorporating the feedback in future programming and telling consumers how the feedback process works (Warfel, Maloney, & Blase, 1981). Repeatedly soliciting feedback without explaining how it will be used, and without implementing any of the suggestions, may teach consumers that their feedback is not useful and the solicitors are deceptive. This may invalidate future attempts to assess social validity and in turn threaten program viability (i.e., consumers' program participation and adoption).

What is Social Invalidity?

Social invalidity is not simply the absence of positive evaluations by the consumers and relevant community when they are asked; it is not simply the failure to ask; it is not simply asking and then ignoring what you hear; and it is not simply the inverse of social validity. Baer (1987) described social invalidity as the behaviors of consumers who not only disapprove of some component in the ongoing program but are going to do something about their disapproval. That something may include withdrawing from the program, encouraging others to do the same, complaining to community officials and the media, or, more subtly, not implementing some or all of the program's procedures after the program consultant leaves, despite positive responses on questionnaires.

Discontented consumers are not members of a homogeneous group; there is great variation in the causes and displays of discontent (Aaker & Day, 1971). That is, some will do something more about their displeasure, some will not; and, of those who do something more, their demonstrations of disapproval can vary widely. Therefore, cases of social invalidity should be evaluated individually to trace the source to specific program or environmental variables and to explore what will happen next, if anything, if the program is changed in response to that feedback and also if it is not.

Social invalidity is difficult to predict from the results of typical social validity assessments. The most consistent finding across reviews of the consumer satisfaction literature in mental health services, medical services, and behavior therapy is that consumers generally rate treatment programs in an overall positive manner (Bornstein & Rychtarik, 1983; Fuqua & Schwade, 1986; Lebow, 1982; McMahon & Forehand, 1983; Ware, Davies-Avery, & Stewart, 1978). Consumers, even when given the opportunity to complain, seem acquiescent, es-

pecially with medical and mental health services (Lebow, 1982; Ware, 1978). Yet, common experience suggests that despite positive ratings, some of those programs are rejected. Some are rejected crudely, and many more subtly, especially in the form of discontinuing program procedures when program personnel leave the scene. Thus, that verbal acquiescence was less than accurate and, hence, less than valid; calling it social validity is frivolous, and relying on it handicaps the prediction of social invalidity.

The prediction of social validity and invalidity requires assessments that evoke truthful reporting of what consumers like and dislike about a program. The first goal of social validity assessments should be to gather accurate and useful information about possible trouble, not to encourage false praise from consumers. Discontented consumers should be urged to complain, and complain early, to key program personnel. Indeed, it may be that consumers who make their complaints public are often more satisfied with the resolution of the problem than consumers who complain privately or not at all (Bearden & Oliver, 1985). Positive responses to social validity assessments do reinforce the program personnel and sometimes please or appease regulatory or funding agencies, but if they do so at the cost of leaving silent a discontented consumer, the initial reinforcement implicit in noncomplaining consumers may be replaced with the punishment of program failure (Bornstein & Rychtarik, 1983).

What Social Validity and Invalidity Are Not

The variety of meanings attributed to social validity has burgeoned beyond usefulness and into confusion. Its original meaning was, and still is, simple: When applying programs in real-life settings, assess early how acceptable those programs will be to their relevant audience. Social validity assessment is a defensive technique. It is oriented toward detecting unacceptability in any of the three major sources—the program's goals, its methods, and its personnel.

But almost immediately after the introduction of this concept, quite different applications of social validity assessments emerged, and in recent years, their variety has flourished. The fear that behavior analysts would substitute the obviously subjective social validity measures for the obviously objective measures of program outcome is one of the oldest criticisms of this methodology. But, this was never intended to be a purpose of social validity assessments. In fact, social validity assessments were always proposed on the assumption that a behavioral program had target behaviors other than statements of liking or disliking the program's goals, procedures, and personnel. It was always supposed that those target behaviors would be measured as directly, behaviorally, objectively, and reliably as possible, and that social validity would be assessed as an important second issue, one relevant not to the program's effectiveness but to its viability (cf. Kazdin, 1977; Wolf, 1978).

Yet, ironically, an implicit contingency to produce positive measures of social validity, as if they were target behaviors, apparently did arise almost immediately. Program developers seemed to assume that an important criterion for publication of their program was a positive social validity assessment. It is not clear what this imputed contingency may have done to the accuracy of those social validity assessment attempts. Of course, a program that accomplishes its goals thoroughly and cost effectively, and is also liked by its consumers, is indeed the ultimate goal of applied behavior analysis. However, the point of social validity assessment is to identify, from all the programs accomplishing their target behaviors thoroughly and cost effectively and also from all the programs not that successful, which ones are liked by their consumers less than others. The dual points of social validity assessment are:

1. It is important to the advancement and survival of applied behavior analysis to know in advance which programs are liked and which are disliked, and thus publication of negative social validity assessments is certainly as valuable and important as publication of positive ones; and

2. It is important to begin the analysis of *why* some programs are liked and others disliked, so eventually social validity assessment can become a

calculated prediction rather than an empirically assessed early warning or endorsement. Without publication of a rich sample of the negative instances to compare with the positive ones, this second goal is not likely to be achieved.

If the experience of behavior analysis is indeed that all its programs seem socially valid, which is what could be concluded from its literature so far, then almost certainly social validity is often being assessed in a spurious manner. Applied behavior analysis simply cannot be that good. It is more likely that current techniques of assessing social validity are too often themselves socially and psychometrically invalid. They are implicitly demanding their consumers to answer their short, simple, bland, undemanding 7-point scales positively. We are encouraging consumers to "fake good," in the terminology of the test-development researchers who have been plagued by similar problems for at least five decades.

In a recent ABA panel discussion of current uses of social validity assessments, Cataldo (personal communication, May 1990) noted a quite similar contingency: The need to rebut for program sponsors, funders, and any other attentive audience the often-heard criticism that behavior-analytic programs are disliked by modern consumers. Indeed, both the media and our conceptual competitors often suggest that behavioral programs are not congruent with the contemporary Zeitgeist of mentalism, personal autonomy, and the need to analyze every social problem to find its origin in the failings of its clients. Confronted by such a need, positive social validity assessments may well seem to be a useful defense. But, as in the prior case, the nature of that contingency is to reinforce positive assessments and punish negative ones and thereby degrade the accuracy of those social validity techniques. Thus, it is a contingency destructive to the original function of yielding early warnings of program rejection by the program's consumers.

In the same ABA panel discussion, Geller (personal communication, May 1990) suggested that social validity assessments are sometimes seen now as techniques for discovering the new, important social goals that behavior-analytic (and other) programs ought to target and for reaffirming the social importance of certain past goals not yet achieved on a societal scale. If social validity assessments are indeed to be used for this purpose, two problems will arise immediately:

1. If the point of social validity assessments is to detect program rejection (by those capable of it and likely to), then sampling should be aimed at representing just those populations. So far that is a relatively easy and inexpensive thing to do in our current small scale of research and application. But, if social importance is the point, then any concept of social importance deriving from the wishes of the society itself will require a random sample of the society as a whole, and this is an extraordinarily difficult and expensive thing to do.

2. However, many concepts of social importance are orthogonal to, and sometimes antithetical to, the momentary wishes of the society as a whole. The true liberation of women and the functional equality of racial, ethnic, and religious minorities are examples of program goals that would have shown little social validity when they began and still show less than unanimous support. Yet, who could now declare them socially unimportant goals? Thus, social importance tends to be, and perhaps always ought to be, resolved by the usual political process of a democratic or even partially democratic society, rather than by anything akin to social validity assessments.

Assessing Social Validity

Identifying Program Consumers

What to ask your audience, who constitutes your audience, and how to assess your audience reliably are key questions in social validity assessments. Reviews of this literature show general agreement on the questions to be asked; however, the questions of who should be asked and how to ask them are not so clear (e.g., Fuqua & Schwade, 1986; McMahon & Forehand, 1983). Three types of questions to ask were summarized succinctly by Wolf (1978):

1. Are the goals of the procedures important and relevant to the desired life-style changes?

2. Are the techniques used acceptable to the consumers and the community, or do they cost too much (e.g., in terms of effort, time, discomfort, ethics, or the like)?

3. Are the consumers satisfied with the outcome, both with the predicted behavior change and with any unpredicted side effects?

Once the questions to assess social validity are chosen, the more difficult issue arises: whom to ask? There is general consensus that the "consumers of the program" should be asked (Bornstein & Rychtarik, 1983; Kazdin, 1977; Larsen, Attkisson, Hargreaves, & Nguyen, 1979; Lebow, 1982; Wolf, 1978). The problem, however, lies in identifying these consumers, especially the consumers who control the viability, either directly or indirectly, of the program. Currently, the consumers most often are simply the recipients of the program. But it could well be true that the persons or agencies who hired the program and certain members of the larger community are also consumers, especially if the criterion is control of the program's viability. For example, the next-door neighbors of a group home program for adults with developmental disabilities become its audience, and they can complain to the larger community if they do not like it. The spouses of factory workers can be a supportive audience of a factory safety program, if they like it, or can be destructive critics of the procedure if they do not. The peers of an aggressive or socially withdrawn child can become part of a program to change the child's behaviors or can sabotage it as soon as the program personnel are absent. Are not taxpayers consumers of all programs supported by public funding?

The key characteristics of these examples are that many people other than the program recipients are passive consumers of treatment programs, and if they decide to become active consumers they can be supportive or critical of the program. We have at present almost no analysis of what turns nonrecipients into active consumers and what makes these consumers program supporters or critics.

In the business world, a consumer is anyone who purchases goods or services or causes them to be purchased (Engel, Kollat, & Blackwell, 1973; Troelstrup, 1974). The breadth of this definition is appropriate to both business and the application of behavioral programs, but the terms *consumer* and *purchase* need more development before they can contribute to the increased accuracy of social validity assessments. For this purpose, it may be helpful to categorize consumers as direct consumers, indirect consumers, members of the immediate community, and members of the extended community. Membership might be determined by the following criteria:

Direct consumers. Direct consumers are the primary recipients of the program intervention. They may have purchased or "hired" the program or may have been referred by someone else. (Thus, a child with developmental disabilities is often the direct consumer, yet rarely "hires" the program.) These consumers can affect program viability directly and at any moment, by participating or by selective or generalized refusals to participate. Examples of direct consumers include parents in a parent-training program, peer monitors and point-earners in a peer-monitoring classroom program, motorists who are referred to a driver-safety program due to their record of traffic violations, and customers in a restaurant participating in a program to increase consumers' selection of low-fat entrees.

Indirect consumers. Indirect consumers purchase or hire the program for someone else or are strongly affected by the behavior change targeted in the intervention, but they are not its recipients. Indirect consumers may directly affect program viability through continuing to purchase more of it, or refusing to do so, and indirectly affect it by behaving as satisfied customers, spouses, or friends of the direct consumers and thus as an advertisement for the program. Examples of indirect consumers include the parents of a child with developmental disabilities who learns to dress independently, the administrators of a company that commissions a program to improve safety conditions in a factory, and the family members of participants in a home-weatherization program.

Members of the immediate community. The immediate community are those people who interact with the direct and indirect consumers on a

regular basis, usually through close proximity during work, school, or social situations. These consumers can affect program viability indirectly, through interaction (or lack of interaction) with the direct and indirect consumers. Examples of members of the immediate community include neighbors of a group home for juvenile offenders, the regular bus driver on the route used by adults with developmental disabilities, and children who occasionally play at the same playground as a child whose severe aggressive behavior was targeted for treatment.

Members of the extended community. The extended community includes those people who probably do not know or interact with the direct and indirect consumers but who live in the same community. Examples include newspaper editors who may find treatment programs newsworthy as either human interest or bad social policy and their swayed or angry readers; supermarket cashiers who may find adults with developmental disabilities using newly acquired shopping skills either heartwarming or troublesome; taxpayers who may protest paying for a program or demand more of it and less of another; Mothers Against Drunk Driving and similar organizations; and legislators and bureaucrats who may regulate the program into or out of existence.

The following example was designed to illustrate these abstract classes of consumers, their roles as purchasers, and the interrelationships among the different roles:

An insurance company contacts one of its clients, the owners of an open-pit mining company, and tells them that unless they implement a safety program for their employees in the mine and document a decrease in the rate of industrial accidents, their insurance rates will quadruple and their policy may be canceled. The mine owners respond by contracting with safety consultants to implement and document an ongoing safety program for the miners (see Fox, Hopkins, & Anger, 1987). The direct consumers are the employees who work at the mine, including miners, janitors, and clerical staff. These consumers, however, did not purchase the program. The program purchasers are the mine owners; therefore, the mine owners are one group of indirect consumers. Another group of indirect consumers consists of the families and friends of the direct consumers who benefit indirectly from the increased safety of their loved ones and breadwinners and may also benefit from any incentive system the safety program uses. The insurance company, although it provided the impetus for the program, is neither a direct nor an indirect consumer. It is a member of the immediate community; it is affected only monetarily by the outcome of the program and may have no other involvement with the program or its participants. The extended community includes taxpayers, who may benefit from fewer workers collecting disability payments, and the merchants who benefit from workers spending both their extra pay and their cash incentives for following the safety guidelines.

This example of consumer classes, despite their somewhat indefinite membership, provides a logic of determining whom to include when assessing social validity. Simply recognizing the existence of different groups of consumers and their possible stakes in program outcomes shows that program adoption or program rejection cannot be predicted safely from a restricted sample of only direct consumers (cf. Mathews & Fawcett, 1979). In this case, the immediate and extended community represent consumers who can truly control program survival, and this case is not unrepresentative.

Conducting Social Validity Assessments

Once the questions of whom and what to ask in a social validity assessment are decided, the next difficult question is how to collect information in a valid, reliable, and cost-efficient manner. This question poses special problems for behavior analysts. Most social validity assessments rely on the use of interviews, questionnaires, or surveys administered by the experimenter (e.g., Fuqua & Schwade, 1986; McMahon & Forehand, 1983; Schwartz, 1991). The subjective nature of this type of information, paired with the possible confounding variables of social contingencies provided by the experimenter (often referred to in other disci-

plines as "demand characteristics"), often make these data difficult to interpret (see Azrin, Holz, & Goldiamond, 1961, for an empirical case in point).

Subjective data can be of value to behavior analysis. Skinner (1953, 1963) acknowledged the importance of private behavior in the study of human behavior; indeed, he differentiated public and private behaviors by their accessibility to observation, not by the extent to which they are lawful phenomena. Even so, any predictable or uniform labeling of feelings is unlikely. For example, many social validity assessments ask respondents to answer questions by describing (e.g., on a 7-point scale) how satisfied they are. "Satisfaction" describes a state controlled by a vast range of different stimulus conditions for different people. For some individuals, "satisfied" may be an average rating, for others it may be extraordinary. The challenge, as Skinner noted long ago, is to seek the public accompaniments of private behavior.

Answers on social validity questionnaires are certainly public; the question becomes one of determining when they are early accompaniments of the behaviors that truly determine program viability and when are they the false positives that may endanger it. Some methods to assess social validity have emerged that cite a wider range of observable behavior but ask the same types of questions as the more commonly used subjective measures. These techniques include (a) use of operationally defined affect rating scales to assess a child's "emotional" behavior while engaged in the target intervention (e.g., Dunlap, 1984; Dunlap & Koegel, 1980), (b) allowing consumers to choose the intervention after being exposed to two or more interventions designed to address the same target behavior (e.g., Harris, 1986), (c) experimentally assessing different rates of the target behavior in the natural environment to determine the optimal rate and using this information to determine the goal of intervention (e.g., Jones & Azrin, 1969; Warren, Rogers-Warren, & Baer, 1976), (d) asking competent performers of the target behavior to judge the adequacy of permanent-product examples produced by the direct consumers of the intervention (e.g., Schepis, Reid, & Fitzgerald, 1987), and (e) asking peer and expert judges to compare photographed or videotaped pre- and postintervention behavior samples (e.g., Friman & Hove, 1987).

Although these procedures encourage applied behavior analysts to use social validity measures that rely on wider sampling of observable and probably relevant behaviors, they represent the minority of procedures used. Sound social validity assessment consists of asking the right questions, to the right people, in an appropriate manner. Before attempting to chart a course for future directions in the area of social validity assessment, we should assess the current practices of the field. We should know (a) who is assessing social validity; (b) who, what, and how they are asking; and (c) what they are doing with the information. Schwartz (1991) attempted to answer these questions in a review of the frequency and variety of social validity assessments reported in one year's output of seven behavioral journals. She found that 29% of the 139 articles reviewed reported some form of social validity assessment (but note that in *JABA* 41% of the 34 articles reported some measure of social validity). However, these articles used inconsistent methods to collect their social validity data and inconsistent conceptual language to describe them. Although an encouraging number of interesting and innovative methods to assess social validity were reported, most of the studies questioned only the direct or indirect consumers and relied solely on the use of questionnaire data to assess social validity.

Future Directions

These results suggest three future directions for improving social validity assessments: (a) More researchers and practitioners should conduct and respond to social validity assessments and begin reporting the apparent results of doing so, negative as well as positive, even if only as anecdotes; (b) a greater breadth of consumers and community members should be sampled, and again the apparent results of doing so should be reported, even if only as anecdotes; and (c) research should target the discovery of more objective, more reliable pre-

cursors of social invalidity than are available at present—perhaps at first by scanning those emerging anecdotes for interesting commonalities.

Each of these areas will be discussed individually, even though improvements in each area may well affect the others, and even though improvements in either of the first two will probably not substantially reduce the present risk of proceeding inappropriately. For example, if researchers currently reporting social validity assessments found more reliable ways to predict social validity and social invalidity, applied behavior analysis would be a better discipline for the knowledge, but only a small proportion of current programs would benefit. On the other hand, if all behavioral researchers began to use the current assessments of unknown validity, the field would be no closer to its goal of solving important problems with effective *and* socially acceptable methods.

Increasing Reports of Social Validity

The first problem is the current rate of using social validity assessments. Is there a case for making social validity assessments as routine in program application and evaluation as measurement reliability assessment is in our current research practices? We believe there is. The results of research programs are not considered credible or replicable without assessing the reliability and validity of their measurement procedures. In the case of application programs, it is not the credibility of their outcomes but the programs themselves that are at risk when they proceed without a reliable assessment of social validity. Perhaps it is just that shift in where to look for validity and reliability that accounts for present consistent attention to measurement reliability and minimal attention to social validity. So far, most of the field's practitioners who report in journals were trained as researchers rather than as practitioners; most of them learned application on the job. If most of their applications have survived so far, at least as long as they were directly involved, they may not even question the viability of their programs. Thus, neither their professional socialization nor their current experience has consistently punished an absence of social validity assessments—at least not with the same vigor as any absence of measurement reliability was punished.

We may solve this disparity by waiting for the inevitabilities of the future, or, perhaps, we may use verbal behavior now to change enough professional behavior to avoid some of them. One technique that may facilitate this change in our professional behavior is to establish criteria for conducting and reporting social validity measures that must be met for publication in behavioral journals. These criteria exist for other methodological issues; perhaps it is time to extend them to social validity assessment.

The field of applied behavior analysis has done much toward demonstrating the lawfulness of human behavior and developing a technology for behavior change. It has been considerably less successful in marketing and disseminating itself and its technology (e.g., Geller, 1989, 1990; Morris, 1985; Pennypacker, 1986; Seekins & Fawcett, 1984). It is at least possible that the widespread use of *accurate* social validity assessments would improve some of that condition, if only by showing program consumers that behaviorists care about what they think rather than about controlling their thoughts (as suggested by the opposition).

Perhaps social validity assessments will become standard procedure when practitioners and researchers stop using social validity assessments as nothing more than compliance with the law (Community Mental Health Centers Amendment, PL 94-63) and as an inexpensive method of assuaging the concerns of vocal program participants and community members. However, the problem may prove to be remarkably similar to that of many health-education programs: Students learn what is taught about preserving or maintaining their health but do not begin to practice what they have learned until they have seen enough peers suffer the consequences of not doing so. Unfortunately for that analysis, program invalidity and its subsequent disasters are rarely reported as journal, convention, or media reports, and so program developers learn about them, and their rates, only through chance anecdotes. Would a program obituary column in every issue of every relevant journal prove salutary?

Even accurate social validity assessments will be of little use to program viability unless they are conducted prescriptively rather than remedially. The social validity of program goals, methods, and anticipated outcomes needs to be known prior to the beginning of the program, and should be assessed periodically throughout implementation. However, if consumers are asked to complete frequent satisfaction questionnaires, doing so had better benefit the consumer in tangible ways: Either any disliked components should be made better, or an excellent and credible reason for not following consumer recommendations should be offered. Otherwise the consumer learns that the social validity assessments are not only useless but in fact are in some sense fraudulent, and thereby the assessment procedure becomes one more element in the program's social invalidity. The evaluation component of the Teaching-Family Model offers an excellent example of the positive side of that relation: It includes an ongoing consumer evaluation system that is steadily responsive to consumers, either with changes or explanations (Blase, Fixsen, & Phillips, 1984; Phillips, Phillips, Fixsen, & Wolf, 1972; Warfel et al., 1981).

Expanding Consumer Participation

Just as it may be difficult to encourage practitioners to conduct social validity assessments more frequently, it may also be difficult to move program developers to sample their consumers more broadly. The extended community has rarely been represented in social validity assessments, and when they have been included they are most likely asked about the acceptability of program objectives, although they are likely to have equally strong interests in the program's methods and outcomes (Schwartz, 1991). Applied behavior analysts, who are usually not in direct control of the viability of their programs, can hardly be considered *applied* without soliciting and using community feedback (Baer, Wolf, & Risley, 1968).

Some researchers are already meeting this challenge. The Teaching-Family Model offers an example for conducting consumer-satisfaction interviews with people with developmental disabilities (Strouse, 1988). Fawcett and his colleagues (e.g., Seekins, Fawcett, & Mathews, 1987; Seekins, Mathews, & Fawcett, 1984; Whang, Fawcett, & Mathews, 1984) validated community-based interventions with panels of experts from the community. For example, the judges of a leadership-training program were two political scientists, an organizational development consultant, an urban planner, and a communications consultant (Seekins et al., 1984); the judges for a job-related social-skills training program were local business people (Whang et al., 1984); and judges of a program teaching consumer-advocacy skills were a local politician and a public administrator (Seekins et al., 1987).

As behavior analysts begin to include a broader spectrum of consumers in social validity assessments, the issue of differentially weighting feedback from various groups of consumers, especially when that feedback is conflicting, must be addressed. The purpose of social validity assessments is to provide information to help ensure program survival. Therefore, the information from consumers most directly related to program viability should be given the most weight. The question of what group of consumers most directly affects program viability is empirical; however, the data necessary to answer these questions are not yet available. So, behavior analysts should begin to collect these data by including members of all four classes of consumers (i.e., direct, indirect, members of the immediate community, and members of the extended community) in social validity assessments and by relating consumer satisfaction and dissatisfaction in each group to program longevity.

The Accuracy of Social Validity Assessments

The accuracy of social validity assessments is the core issue for future work. Social validity is intrinsically an adjunctive measure; its function is not to evaluate program effectiveness but program acceptability and viability. Similarly, its purpose is not to compare programs but to safeguard pro-

grams against rejection or sabotage. Rather than attempting to compare the acceptability of programs or program components, perhaps we should review the assessed components of programs and compile a menu of elements of different programs rated favorably by consumers and a parallel menu of those rated unfavorably. Then we can examine these program components to determine whether they have anything in common that would explain their similar ratings, despite the dissimilarity of the programs in which they operated. It is the similarities in the highly regarded program components (and in the unfavorably rated ones) that may provide answers of how to provide acceptable and effective behavior-change interventions.

Improved construction of social validity assessments is an important goal for applied behavior analysis. Researchers should consider basic rules of test construction and statistics when developing social validity assessments (Baer, 1987). Questionnaires about program acceptability should (a) use scales that invite a workably wide variation in consumers' responses (e.g., a 7-point rather than a 3-point scale), (b) require differential responding by the consumer (i.e., they should require even the most satisfied consumer to use the entire range of the rating scale) (Cone, 1981), (c) specify the period of time being rated (e.g., ask consumers to rate only the services provided during the last 3 months, rather than all services) (McMahon, 1984), (d) address all the dimensions pertinent to the acceptability and viability of a program (e.g., if you are interested in the consumers' opinions about observers in the classroom, ask directly), and (e) be as specific as possible, because increased specificity may increase the usefulness of information collected from social validity questionnaires (Mash & Terdal, 1981).

If, in addition, more objective techniques for assessing social validity are developed, these assessments might well prove even more useful. Four classes of techniques could be expanded to assess social validity more functionally: (a) proof that a program's goals and outcomes are themselves valid, (b) unobtrusive measurement, (c) identification and measurement of the behavioral precursors of the kinds of satisfaction and dissatisfaction culminating in program rejection, and (d) providing the consumers with experiences with different program options, then allowing them to choose the most satisfactory option.

Goals and outcomes. Asking about a program's goals and outcomes is the most frequently assessed form of social validity (Schwartz, 1991). Van Houten (1979) suggested that the behavior of normal models and competent performers be used to help identify target behaviors and determine the standards of competent performance. Identifying competent performers, however, invokes difficult ethical issues. For example, who are the competent models for people living in a state institution? Should we set community standards for these consumers, because we assume they will get there, or do we pose lower standards appropriate to their present environment, because we assume they are staying? If the former, then future-environment surveys (Anderson & Schwartz, 1986; Fowler, 1982) can answer these questions. These surveys identify the clients' next environment and set behavioral goals to levels appropriate to the subsequent setting. This technique can ease clients' transitions across programs (e.g., from a residential facility to living at home, or from a specialized preschool to regular kindergarten).

The behavior of community members can be used as models to set intervention goals. Junior high school and college students can model the appropriate levels of conversational skills to teach delinquent and predelinquent teenagers (Minkin et al., 1976). Normal peers in preschool classrooms can be observed to determine desirable within-classroom transition skills and rates of verbal interaction (Osnes, Guevremont, & Stokes, 1986; Sainato, Strain, Lefebvre, & Rapp, 1987). Judges can score the fluency, "naturalness," and speech rates of college students treated for stuttering (Jones & Azrin, 1969). The optimum rate for preschoolers to make share offers to peers is taken to be the rate most often accepted by those peers; the peers become the experts simply by consistently responding more often to some rates than to others (Kohler & Fowler, 1985; Warren et al., 1976). Teachers and peers

can rate the social status of their students or friends and then identify target skills for those rated lowest. Subsequent to the establishment of those skills, teachers or peers can rate the students or friends again, to see if those skills were indeed crucial to improving their status with those judges (Hoier & Cone, 1987; Plienis et al., 1987).

When selecting community members to serve as experts, however, we need to demonstrate that they are in fact appropriate experts for the target intervention (Greenwood, 1990). Community membership alone does not qualify someone to serve as a model for appropriate behavior or to judge the outcome of an intervention. In general, the original complainer or buyer (i.e., the indirect consumer who hired the program) can validate the goals and the behavioral targets as almost no one else can, especially if that person is the only complainer (Baer, 1986). If it is their complaint that defines the problem, then the program applied has social validity if their complaint is satisfied by it—they are the "community" defining validity for this problem. If, however, the problem is defined by a larger community, we must identify representative and appropriate members of that community to assist in assessing the social validity of the proposed solution.

Unobtrusive measurement. Obtrusive measurement is often its own threat to social validity, because so many people dislike having their behavior rated. Thus, the development of unobtrusive measures can occasionally be crucial to the social validity of a program acceptable in every other dimension. Unobtrusive measurement occurs when the subject is unaware of ongoing observations or target behavior (Kazdin, 1979). However, this type of measurement rarely is ethical and often is illegal (American Psychological Association, 1979). And when it is ethical and legal, it often generates some logistic issues of how to do it, especially outside of laboratory settings. Still, some forms of unobtrusive measurement can escape these handicaps: for example, the use of archival records. Documents can be searched to retrieve data on rates of recidivism, complaints (see Baer, 1988), industrial accidents, energy consumption (Winett, Neale, & Grier, 1979), and the like. It is important to remember, however, that although data from archival sources escape many difficulties in the realm of social validity, their validity as outcome data is often doubtful. For example, reduced recidivism of juvenile offenders may reflect increased criminal skill, decreased police behavior, changed judicial standards, or overloaded record-keeping instead of decreased criminal activity (Kazdin, 1979). Similarly, increased recidivism by children with autism in residential treatment facilities may reflect a breakdown in family systems rather than a lack of generalization and maintenance of children's behavioral gains (Anderson, Christian, & Luce, 1986).

Behavioral correlates of satisfaction. The most basic technique crucial to accurate social validity assessments, however, remains the identification and measurement of observable behaviors that correlate with program satisfaction or rejection. The behaviors that clearly show support for a program include continued participation in it, regular attendance and prompt arrival for program sessions, recommending it to friends, earning many of whatever points its incentive system may offer, defense of the program against attack by others, and pleasant affect and high enthusiasm during program sessions. The behaviors that clearly show the social invalidity of a program include withdrawal, demands that the program personnel leave (e.g., the program buyer "firing" the program), complaining to friends, officials, and the media, poor attendance and/or chronic tardiness at program sessions, refusal to participate or poor performance in whatever incentive system the program may offer, generalized or selective noncompliance with program routines, and negative affect and low enthusiasm during program sessions.

Many of these indices of social validity are relatively easy to measure and, in fact, may already be included in ongoing data collection systems. Their disadvantage, which can often outweigh their validity, is their lateness in emerging. By the time these signs of social validity or invalidity appear, especially in the case of social invalidity, it is often too late to respond adaptively to them.

Even so, these measures may have some utility

in programs for people with developmental disabilities, because many do not require much verbal ability, and people with disabilities often express their displeasure with a program quite early in its development, often in just these ways. Indeed, it is more often the developmentally able consumers who are burdened with a courtesy that keeps them from troubling program appliers with their complaints until those complaints are quite severe. For example, Koegel and his colleagues (e.g., Dunlap & Koegel, 1980; Koegel & Egel, 1979) have developed a reliable coding system for assessing affect and enthusiasm in children with autism. These overt measures of satisfaction are potentially relevant to measuring the social validity of programs serving people with severe developmental disabilities. Note again that this measurement system is an adjunctive measure, not an evaluation of program effectiveness.

A prerequisite for measuring behaviors correlated with program satisfaction or dissatisfaction is the identification of these behaviors. Researchers may be able to identify behavioral correlates of satisfaction by studying successful treatment programs (see Fuoco & Christian, 1986; Paine, Bellamy, & Wilcox, 1984, for descriptions of successful programs). Identifying correlates of dissatisfaction, especially early and perhaps predictive correlates, is more elusive. One reason may be that researchers do not disseminate information about program failures or dissatisfied consumers. Perhaps applied behavior analysts should act as epidemiologists and conduct careful post mortem analyses on unsuccessful programs. Prior to program implementation, behavior analysts conduct thorough need assessments, and throughout an intervention they systematically evaluate program effectiveness. This type of thorough assessment should be extended to dissatisfied consumers and program failures, and the results of these assessments should be disseminated through professional journals and conference presentations. Careful study of elements common to program failures may lead to more effective program planning and more sensitive social validity assessments.

Choice offering. The ultimate measure of a program's social validity is the range of alternatives a consumer will reject in order to choose the program. Choosing a program over a number of alternatives is the fundamental behavioral definition of preference, and what is more socially valid than that? Arranging such a measure of social validity, however, is simple only in logic: Give the consumers experience with two or more programs and let them choose. The probability of their later rejecting what they have chosen against a range of alternatives should be low, to the extent that the range is wide. However, as new alternatives to their current choice become available—as the range widens—the choice must be offered again to ensure the stability of their preference and the durability of the social validity assessed earlier.

The validity of choice is strong and the logic of choice is simple, but the implementation of choice is not. Valid choice requires that the consumer have extensive, nearly concurrent experience with all the alternatives, that all the alternatives are equally effective, and that all alternatives are easily available. Additionally, evaluators must carefully select the treatment options, so not to bias the assessment by offering the target program posed against much less desirable alternatives. Arranging those conditions will prove expensive in time, effort, resources, and sophistication, especially if the number of choices exceeds two. And although all of these barriers are formidable, time may be especially problematic. The function of social validity assessments is to provide an *early* warning of program rejection in any of its forms; thus, any measure requiring a great deal of time to be valid may be useless.

As a consequence, we are always tempted to search instead for quick, cheap, easy, and gracefully repetitive measures that will predict social validity or invalidity early, and our intrinsic problem is the suspicion that the quicker, cheaper, easier, earlier, and more graciously repetitive these measures are, the lower their correlation with the actual events that make up social validity and invalidity. In other words, easily and early-stated verbal preference and later, actual behavioral choices do not always correlate highly, even with nonimpaired adults (e.g., Lockhart, 1979), and perhaps especially with nonimpaired adults. Still, it may be worth recom-

mending studies of the correlation between various forms of verbal preference, some of them much like the methods currently in use, to see if any of them correlate especially well with the very expensive choice techniques.

Offering choices, however, is not an impossible technique, especially if we do not insist on perfection in its execution. It has been used with college students to assess different teaching methods (Lockhart, 1979), among adults with severe mental retardation to assess task preference in vocational settings (Mithaug & Hanawalt, 1978; Mithaug & Mar, 1980), among students with learning disabilities to select preferred instructional strategies (Harris, 1986), and with students demonstrating high rates of self-stimulatory behavior to assess their preference for more appropriate forms of stimulation (Buyer, Berkson, Winnega, & Morton, 1987). So far, we have not learned that any of those less-than-perfect applications proved disastrously misleading.

Although the logic of choice as the most valid method of making social validity assessment is well developed (e.g., Fuqua & Schwade, 1986; McMahon & Forehand, 1983), methods for implementing choice in more natural settings need work to facilitate integration into program evaluations. Additionally, choice making is not an appropriate assessment technique with all programs. Not all programs permit side-by-side comparisons with alternative programs, but for selected programs this technique may be an efficient and objective technique to assess social validity.

Conclusions

The concept of assessing the social validity of behavioral programs, and then using these data in programmatic decisions, has been controversial since it was introduced (Wolf, 1978). Yet, rather than disappearing, social validity assessments are becoming more common in behavioral research and practice. Changes in the current practice of social validity assessment are needed, however, before these assessments can be used as accurate and reliable predictors of program success or failure. As behavior analysts, we require procedures relevant to behavior-change programs to be thoroughly specified and defined (Baer et al., 1968); we should judge techniques to assess social validity by the same criteria. Most current social validity assessment procedures do not meet those criteria fully enough to be considered a technological tool used to secure the maintenance of applied programs and advance the state of our science.

The procedures for conducting social validity assessments should include the following guidelines:

1. Social validity assessments should be a standard part of program application and applied research methodology. Inclusion of a thorough social validity assessment should be a minimum requirement, similar to assessment of interobserver reliability, in all applied behavioral research and practice.

2. Social validity assessments should be conducted prospectively and throughout an intervention, as well as at the end. Otherwise, consumers' concerns about the program cannot be answered in ways that defend the consumers, the program, and the discipline.

3. Applied behavior analysts need to recognize that there are more relevant and powerful consumers than they have identified and queried so far. Thus, a wider spectrum of consumers should be included in ongoing social validity assessment, with special consideration to consumers who control the viability of community programs. When consumers cannot respond to standard forms of social validity assessments (e.g., people with developmental disabilities), special techniques must be developed for them.

4. Social validity assessments, though not program-outcome measures, nevertheless deserve the same psychometric rigor as any behavioral measure. Rigor is not the same as standardization, but striving for valid and reliable measurement would probably lead to a more uniform methodology than characterizes the current practice of social validity assessment. Better attention to validity and reliability would make these measures better predictors of program acceptance or rejection.

5. If the fourth recommendation is to be achieved,

the field will require the development of some objective, clearly valid measures of social validity and social invalidity (i.e., program acceptance and rejection). It follows that we shall have to study actual program acceptance and rejection, painful as that will be, in order to know something about its truly reliable and valid antecedents. When we have identified valid and reliable precursors of program acceptance and rejection, we can then study how well these questionnaire techniques relate to them, or, more characteristic of us, how well they *can be made to* relate to them. Objective procedures to measure social validity (e.g., observations, reports of program adoption rates) will invariably be more expensive to conduct than the standard subjective measures (e.g., questionnaires). This expense, however, is minimized if viewed in the context of accurately predicting program viability.

6. Finally, applied behavior analysts should offer their consumers and relevant community members more education about the ongoing treatment programs, potential treatment programs, and other treatment options. This information will probably represent a mix of what we discover our consumers want to know and what we believe they need to know. This information should not be used as propaganda. We are not usually and not properly in the business of shaping our consumers' values; instead, we need to know what information is required to enable the consumers to make informed decisions. In behavior analysis, this task is especially important. Behavior analysis is often misrepresented in the popular and scientific media (Morris, 1985; Todd & Morris, 1983); we had better not intensify the problem. Consumers are entitled to accurate information presented in a clear and intelligible manner. They should not be asked to make choices about any services without adequate information about them and about all relevant, available treatment options.

Since the introduction of social validity, applied behavior analysis has moved further toward Wolf's (1978) implicit prescription—the search for its heart. The use of social validity assessments in behavioral research has increased, and researchers and practitioners alike (at least sometimes) try to improve its technique. Better social validity assessments are vital to the survival of applied behavior analysis; however, they are not the discipline, only a defensive tool of it. We still need to be applied, behavioral, analytic, technological, conceptual, effective, and in as generalized a way as the problem requires (Baer et al., 1968, 1987). State-of-the-art social validity assessment should address all seven of those dimensions, and when it does, the discipline will be another step closer to finding its heart.

REFERENCES

Aaker, D. A., & Day, G. S. (1971). Introduction: A guide to consumerism. In D. A. Aaker & G. S. Day (Eds.), *Consumerism: Search for the consumer interest* (pp. 1–19). New York: Free Press.

American Psychological Association. (1979). *Ethical standards of psychologists.* Washington, DC: American Psychological Association.

Anderson, S. R., Christian, W. P., & Luce, S. C. (1986). Transitional programming for autistic individuals. *Behavior Therapist,* **9,** 205–211.

Anderson, S. R., & Schwartz, I. S. (1986). Transitional programming. In F. J. Fuoco & W. P. Christian (Eds.), *Behavior analysis and therapy in residential programming* (pp. 76–100). New York: Van Norstrand Reinhold.

Azrin, N. H., Holz, W., & Goldiamond, I. (1961). Response bias in questionnaire reports. *Journal of Consulting Psychology,* **25,** 324–326.

Baer, D. M. (1986). In application, frequency is not the only estimate of the probability of behavioral units. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 117–136). Hillsdale, NJ: Erlbaum.

Baer, D. M. (1987, March). *A behavior-analytic query into early intervention.* Paper presented at the Banff International Conference on Behavioral Science, Banff, Canada.

Baer, D. M. (1988). If you know why you're changing a behavior, you'll know when you've changed it enough. *Behavioral Assessment,* **10,** 219–223.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* **1,** 91–97.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* **20,** 313–327.

Barrett, B. H. (1987). Drifting? Course? Destination?: A review of *Research in applied behavior analysis: Issues and advances. The Behavior Analyst,* **10,** 253–276.

Bearden, W. O., & Oliver, R. L. (1985). The role of public and private complaining in satisfaction with problem resolution. *Journal of Consumer Affairs,* **19,** 222–240.

Blase, K., Fixsen, D., & Phillips, E. (1984). Residential treatment for troubled children: Developing service delivery systems. In S. C. Paine, G. T. Bellamy, & B. Wilcox (Eds.), *Human services that work: From innovation to standard practice* (pp. 149–166). Baltimore: Paul Brookes.

Bornstein, P. H., & Rychtarik, R. G. (1983). Consumer satisfaction in adult behavior therapy: Procedures, problems, and future perspectives. *Behavior Therapy,* **14,** 191–208.

Buyer, L. S., Berkson, G., Winnega, M. A., & Morton, L. (1987). Stimulation and control as components of stereotyped body rocking. *American Journal of Mental Deficiency,* **91,** 543–547.

Cone, J. (1981). Psychometric considerations. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (pp. 38–68). New York: Pergamon.

Deitz, S. M. (1978). Current status of applied behavior analysis: Science vs. technology. *American Psychologist,* **33,** 805–814.

Dunlap, G. (1984). The influence of task variation and maintenance tasks on the learning and affect of autistic children. *Journal of Experimental Child Psychology,* **37,** 41–64.

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis,* **13,** 619–627.

Engel, J. F., Kollat, D. T., & Blackwell, R. D. (1973). *Consumer behavior.* New York: Holt, Rinehart, & Winston.

Fowler, S. A. (1982). Transition from preschool to kindergarten for children with special needs. In K. E. Allen & E. M. Goetz (Eds.), *Early childhood education: Special problems, special solutions* (pp. 309–334). Rockville, MD: Aspen.

Fox, D. K., Hopkins, B. L., & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behavior Analysis,* **20,** 215–224.

Friman, P. C., & Hove, G. (1987). Apparent covariations between child habit disorders: Effects of successful treatment for thumb sucking on untargeted chronic hair pulling. *Journal of Applied Behavior Analysis,* **20,** 421–425.

Fuoco, F. J., & Christian, W. P. (Eds.). (1986). *Behavior analysis and therapy in residential programs.* New York: Van Nostrand Reinhold.

Fuqua, R. W., & Schwade, J. (1986). Social validation of applied behavioral research: A selective review and critique. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 265–292). New York: Plenum.

Geller, E. S. (1989). Applied behavior analysis and social marketing: An integration for environmental preservation. *Journal of Social Issues,* **45,** 17–36.

Geller, E. S. (1990). Behavior analysis and environmental protection: Where have all the flowers gone? *Journal of Applied Behavior Analysis,* **23,** 269–273.

Giordano, P. (1977). The client's perspective in agency evaluation. *Social Work,* **22,** 34–39.

Greenwood, C. R. (1990, October). Social validity. In S. Warren (Chair), *Innovations in research standards and methods.* Symposium presented at the Annual Meeting of the Division of Early Childhood, Council for Exceptional Children, Albuquerque, NM.

Harris, K. R. (1986). Self-monitoring of attentional behavior versus self-monitoring of productivity: Effects on on-task behavior and academic response rate among learning disabled children. *Journal of Applied Behavior Analysis,* **19,** 417–423.

Hoier, T. S., & Cone, J. D. (1987). Target selection of social skills for children: The template-matching procedure. *Behavior Modification,* **11,** 137–163.

Jones, R. J., & Azrin, N. H. (1969). Behavioral engineering: Stuttering as a function of stimulus duration during speech synchronization. *Journal of Applied Behavior Analysis,* **2,** 223–229.

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification,* **1,** 427–452.

Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis,* **12,** 713–724.

Koegel, R. L., & Egel, A. L. (1979). Motivating autistic children. *Journal of Abnormal Psychology,* **88,** 418–426.

Kohler, F. W., & Fowler, S. A. (1985). Training prosocial behaviors to young children: An analysis of reciprocity with untrained peers. *Journal of Applied Behavior Analysis,* **18,** 187–200.

Larsen, D. L., Attkisson, C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning,* **2,** 197–207.

Lebow, J. (1982). Consumer satisfaction with mental health treatment. *Psychological Bulletin,* **91,** 244–259.

Lockhart, K. A. (1979). Behavioral assessment of human preference. *The Behavior Analyst,* **2,** 20–29.

Mash, E. J., & Terdal, L. G. (1981). Behavioral assessment of childhood disturbances. In E. J. Mash & L. G. Terdal (Eds.), *Behavioral assessment of childhood disorders* (pp. 3–76). New York: Guilford.

Mathews, R. M., & Fawcett, S. B. (1979). Assessing dissemination capability: An evaluation of an exportable training package. *Behavior Modification,* **3,** 49–62.

McMahon, R. J. (1984). Behavioral checklists and rating scales. In T. H. Ollendick & M. Hersen (Eds.), *Child behavioral assessment: Principles and procedures* (pp. 80–105). New York: Pergamon.

McMahon, R. J., & Forehand, R. L. (1983). Consumer satisfaction in behavioral treatment of children: Types, issues, and recommendations. *Behavior Therapy,* **14,** 209–225.

Michael, J. (1980). Flight from behavior analysis. *The Behavior Analyst,* **3,** 1–22.

Minkin, N., Braukman, C. J., Minkin, B. L., Timbers, G. D., Timbers, B. J., Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1976). The social validation and training of conversational skills. *Journal of Applied Behavior Analysis,* **9,** 127–139.

Mithaug, D. E., & Hanawalt, D. A. (1978). The validation of procedures to assess prevocational task preferences in retarded adults. *Journal of Applied Behavior Analysis,* **11,** 153–162.

Mithaug, D. E., & Mar, D. K. (1980). The relation between choosing and working prevocational tasks in two severely retarded young adults. *Journal of Applied Behavior Analysis,* **13,** 177–182.

Morris, E. K. (1985). Public information, dissemination, and behavior analysis. *The Behavior Analyst,* **8,** 95–110.

Osnes, P. G., Guevremont, D. C., & Stokes, T. F. (1986). If I say I'll talk more, then I will: Correspondence training to increase peer-directed talk by socially withdrawn children. *Behavior Modification,* **10,** 287–299.

Paine, S. C., Bellamy, G. T., & Wilcox, B. (Eds.). (1984). *Human services that work: From innovation to standard practice.* Baltimore: Paul Brookes.

Pennypacker, H. S. (1986). The challenge of technology transfer: Buying in without selling out. *The Behavior Analyst,* **9,** 147–156.

Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1972). *The teaching-family handbook.* Lawrence, KS: Bureau of Child Research.

Pierce, W. D., & Epling, W. F. (1980). What happened to analysis in applied behavior analysis. *The Behavior Analyst,* **3,** 1–9.

Plienis, A. J., Hansen, D. J., Ford, F., Smith, S., Stark, L. J., & Kelly, J. A. (1987). Behavioral small group training to improve the social skills of emotionally-disordered adolescents. *Behavior Therapy,* **18,** 17–32.

Sainato, D. M., Strain, P. D., Lefebvre, D., & Rapp, N. (1987). Facilitating transition times with handicapped preschool children: A comparison between peer-mediated and antecedent prompt procedures. *Journal of Applied Behavior Analysis,* **20,** 285–292.

Scheirer, M. (1978). Program participants' positive perceptions: Psychological conflict of interest in social program evaluation. *Evaluation Quarterly,* **2,** 53–70.

Schepis, M. M., Reid, D. H., & Fitzgerald, J. R. (1987). Group instruction with profoundly retarded persons: Acquisition, generalization, and maintenance of a remunerative work skill. *Journal of Applied Behavior Analysis,* **20,** 97–105.

Schwartz, I. S. (1991). *A review of current practice of methods used to assess social validity.* Manuscript submitted for publication.

Seekins, T., & Fawcett, S. B. (1984). Planned diffusion of social technologies for community groups. In S. C. Paine, G. T. Bellamy, & B. Wilcox (Eds.), *Human services that work: From innovation to standard practice* (pp. 247–260). Baltimore: Paul Brookes.

Seekins, T., Fawcett, S. B., & Mathews, R. M. (1987). Effects of self-help guides on three consumer advocacy skills: Using personal experiences to influence public policy. *Rehabilitation Psychology,* **32,** 29–38.

Seekins, T., Mathews, R. M., & Fawcett, S. B. (1984). Enhancing leadership skills for community self-help organizations through behavioral instruction. *Journal of Community Psychology,* **12,** 155–163.

Skinner, B. F. (1953). *Science and human behavior.* New York: Free Press.

Skinner, B. F. (1963). Behaviorism at fifty. *Science,* **140,** 951–959.

Strouse, M. C. (1988). The client satisfaction evaluation. In J. A. Sherman, J. B. Sheldon, & M. C. Strouse (Eds.), *Evaluation programs using the Teaching-Family Model for people with developmental disabilities* (pp. 1–36). Topeka, KS: Kansas Planning Council on Developmental Disabilities.

Todd, J. T., & Morris, E. K. (1983). Misconceptions and miseducations: Presentations of radical behaviorism in psychology textbooks. *The Behavior Analyst,* **6,** 153–160.

Troelstrup, A. W. (1974). *The consumer in American society.* New York: McGraw Hill.

Van Houten, R. (1979). Social validation: The evolution of standards of competency for targets. *Journal of Applied Behavior Analysis,* **12,** 581–591.

Ware, J. E. (1978). Effects of acquiescent response set on patient satisfaction ratings. *Medical Care,* **16,** 327–336.

Ware, J. E., Davies-Avery, A., & Stewart, A. L. (1978). The measurement and meaning of patient satisfaction. *Health and Medical Care Services Review,* **1,** 1–15.

Warfel, D. J., Maloney, D. M., & Blase, K. (1981). Consumer feedback in human service programs. *Social Work,* **26,** 151–156.

Warren, S. F., Rogers-Warren, A., & Baer, D. M. (1976). The role of offer rates in controlling sharing by young children. *Journal of Applied Behavior Analysis,* **9,** 491–497.

Whang, P. L., Fawcett, S. B., & Mathews, R. M. (1984). Teaching job-related social skills to learning disabled adolescents. *Analysis and Intervention in Developmental Disabilities,* **4,** 29–38.

Winett, R. A., Neale, M. S., & Grier, H. C. (1979). Effects of self-monitoring and feedback on residential electricity consumption. *Journal of Applied Behavior Analysis,* **12,** 173–184.

Wolf, M. M. (1978). Social validity: The case for subjective measurement, or how behavior analysis is finding its heart. *Journal of Applied Behavior Analysis,* **11,** 203–214.

Received August 8, 1990
Initial editorial decision January 7, 1991
Revision received February 12, 1991
Final acceptance February 17, 1991
Action Editor, E. Scott Geller