JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1991, 24, 231–234 NUMBER 2 (SUMMER 1991)

IF RELIANCE ON EPIDEMIOLOGY WERE TO BECOME EPIDEMIC, WE WOULD NEED TO ASSESS ITS SOCIAL VALIDITY

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

Behavior analysts often discuss social validity these days. Some see it as an essential element of the field's survival, others as a diversionary trap leading to the field's demise. Either argument is daring, in that we know rather little about the accurate and valid assessment of what gets called social validity—and we know a great deal less about the survival of fields.

Our most defensible argument may be only that until we can do social validity assessments well, we cannot determine their importance. Of course, if they have importance, it will be to help choose and guide program developments and applications. To the extent that we are unsure of the optimal way to choose and guide program developments and applications, we are not ready to address the needs or concerns of our community. This is not a simple problem. Any procedures developed to facilitate accurate social validity assessment must be applicable to the wide scope of research questions and consumers encountered by behavior analysts. Currently those range from teaching academic support skills to high-risk preschool students to implementing community-wide public health programs. In the future, if things go well, the necessary range will be even wider. The interesting question is whether the assessment of social validity is an important part of things going well in the future. One guess is that it ought to be not merely important, but crucial.

Winett, Moore, and Anderson (1991) address many of these issues and questions. Their concept of social validity is interesting, partly because it is somewhat different from our precedents; however, considering their examples, it is certainly applicable. Winett's group looks beyond the immediate results of their work; they guess about the long-term implications of both the programmed and unexpected behavior changes, even if the length of that term requires the guesses to become increasingly diffuse and nonspecific. Certainly that is the central strategy of social validity assessment. But their tactic in that strategy is distinctive, at least for our literature, at least so far. They offer first a primer of epidemiology and an adventure called "social marketing," and from that base suggest an intersection of these methodologies for social validity assessments and program evaluation.

Their definition and characterization create what might be called a *rational* social validity—the social validity that will be seen as such by rational people. Its essence is that a problem is not simply what causes people to complain about it; a problem has *verifiable* importance (e.g., at the level of epidemiological data). Furthermore, interventions into problems are to be evaluated first by objective, reliable measures of their effectiveness testifying in careful, complete experimental designs that yield unambiguous judgments about that effectiveness. When a relevant audience is not rational enough to adopt and maintain an intervention merely proven to solve what they ought to recognize as their problem, the intervention is socially marketed interactively with its development, either as it becomes clear that it must be, or routinely, like an immunization against some disease (in this case, the disease of social rejection).

We can only agree, but that is mainly because we are often members of that rational audience. (Please, all contradictions should be restricted to written notes sent to us in a plain brown wrapper.) That is, we, too, are reinforced by objective, reliable

We gratefully acknowledge Mary Todd for her assistance in preparing this manuscript. Requests for reprints should be addressed to Donald M. Baer, Department of Human Development, 4001 Dole Human Development Center, University of Kansas, Lawrence, Kansas 66045, or Ilene S. Schwartz, Juniper Gardens Children's Project, 1614 Washington Blvd., Kansas City, Kansas 66102.

(often epidemiological) data testifying correctly to the societal importance of a problem (which is often its size interacting with the severity of its consequences); perhaps that is because we see ourselves also as members of a society so much at risk for its survival that the research and intervention efforts of our discipline ought to be aimed mainly at the most urgent threats to that survival that our discipline knows how to approach. We are also members of a science subcommunity that defines proof in natural-science terms; thus we define effectiveness in terms of the representativeness of the experimental samples, the choices of the control conditions, the choices of the measures, the objectivity, reliability, and validity of those measures, the kind and size of changes in those measures achieved by the intervention, and the thoroughness of the experimental design in which those data were gathered—for starters. Finally, we are workers in a discipline called applied behavior analysis; because of that, we assume that the behaviors constituting social validity and invalidity are modifiable behaviors, and consequently might well be modified, perhaps primarily by the same techniques that make up social marketing.

But, quintessentially (at least for the moment), we are also members of a social-policy subcommunity of a philosophy-of-science community likely to label the modification of any current social validity or invalidity in a target population not only as behavior modification but also as the imposition of our social validity on theirs, apparently because we know we are right.

In addition, we are members of a skeptical community that acknowledges the objectivity, reliability, and validity that epidemiological data can have, but also notes there are many sets of such data; in effect, someone must choose which of them identifies the most imperative problem for our next intervention. Although epidemiological data are (or can be made) objective, reliable, and valid, choosing among them is personal, subjective, and behavioral. Therefore, it is subject to many contingencies and stimulus controls other than, and in addition to, rationality. So, when some of us respond to the range of epidemiological data on our current societal problems by choosing one of them for immediate intervention and others of us respond to the same range of data by choosing a different problem for immediate intervention, where is the epidemiology to show in advance which of us made the correct choice? More likely, we will engage in one of those behavioral processes usually labeled "politics" to see which choice will be declared officially correct and acted upon. And what else is the assessment of social validity if not the attempt to predict the outcome of exactly that process in advance—either to acknowledge and comply with it, or to acknowledge and then try to change it? Assessment does not commit us to either course.

Does the case described by Winett et al. (1991) create a "we" who are presumed more rational than "they"? We think so. If it does, what are the implications of doing so for the survival of a discipline trying to study and contribute to the survival of its society? We do not know; we can only guess. Thus, the approach taken by Winett and his colleagues is a strikingly profound one, and the resulting definition of social validity challenges current applied behavior analysis. True, this field has encountered many deliberate challenges these past two decades, and too many of them have taught us only to tend our garden. However, this one looks different—this one we had better meet.

If we are becoming dissatisfied with assessments of social validity that merely ask selected consumers of a program to complete a short questionnaire administered by the experimenter at the end of the intervention, and if we are becoming dissatisfied with assessments of treatment efficacy based on a small, single, hand-picked sample that received the intervention under ideal situations, then perhaps we are ready to become contextual. If so, then we shall soon learn how to assess the contextual variables governing the outcomes of our interventions, and to acknowledge that when any component of that context is altered, other components very likely will react to that change. Perhaps that is indeed a fair description of the paths taken earlier by both epidemiology and social marketing. Perhaps we have something to learn from them.

But, as always in disciplinary challenges, perhaps

we do not. However, as always, the prudent strategy in responding to strategy clash is to find out. We might well be cautious, though, in doing that. First, we should remember that welcoming new assessment methodologies from other disciplines does not require discarding our old methodologies, either for assessing or achieving the outcomes we target. This is good to remember, especially when the other disciplines from which we are borrowing seem to recommend just that. Most of what we do has been empirically documented as effective in accomplishing what it was aimed to do. These new recommendations, after all, are only to aim at more: in particular, to aim at a specific more, a kind of consumer behavior tentatively labeled "consumer satisfaction" or "social validity," that may prove crucial to achieving a large-scale effectiveness for what previously was small-scale effectiveness. It may often be true that the achievement of small-scale effectiveness does not demand as much consumer satisfaction as does large-scale effectiveness. On the other hand, perhaps that is not true very often. To find out which is the case, we shall need an accurate assessment of consumer satisfaction. If we then find that we do indeed often need a lot of consumer satisfaction for viable large-scale applications, how fortunate it will be for us that we already know (by then) how to assess it accurately and validly.

Thus, these are recommendations only to begin the study of the accurate and valid assessment of consumer satisfaction, so that we will have a crucial dependent variable for all that research. (Note how different is "a crucial dependent variable" from "the crucial dependent variable" or "the only dependent variable.") Thus, we need not alter our current conceptual approach in this adventure. We have proposed only an extension of behavioral measurement, starting with what we already have. The main ugliness in what we already have is only that it looks terribly vulnerable to invalid and misleading assessments of social validity, and if so, that could spell trouble for our clients, us, our discipline, and our discipline's chance of contributing to its society. If, in the process of finding out how to measure social validity accurately and validly, we discover that (a) we had been doing so all along and/or (b) it was not all that important to the survival of our applications and the health of our discipline, then we may celebrate, resolve not to heed foolish Cassandras ever again, and get on with the important development of better, more valuable, and more applicable interventions, either with those good old cheap and easy measures of social validity or without any. Some of us will also emit some tacts about surprise—after all, Cassandra was correct. But in that future, this form of surprise will hardly constitute a literature event, let alone a journal symposium.

The arguments of Winett and his colleagues require only caution, not criticism. Those arguments recommend some techniques to extend our current assessments of social validity, and so do our arguments. But neither team is yet in a position to urge all researchers and practitioners to discard their current practices in favor of what now are only promising, still not thoroughly developed, and very expensive techniques of unknown generality. We are in a position only to recommend their intensive investigation because of their attractive logic and their usefulness in some initial trials.

For example, Winett et al. (1991) suggest that social validity "is established through a number of interactive a priori steps" (p. 215). They offer some examples to make these abstractions real. In these examples, they do not seek expert views of the problem, do not let the researcher identify the problem, and do not wait for someone in the community to generate specific complaints. Instead they choose some class of epidemiological data to inspect, and from it, to identify problems and the target groups most at risk. That determines the nature of their interventions.

The original Cassandra was the one person who knew with certitude that her Trojans would soon make a fatal error and later knew with equal certitude exactly which of their numerous tactics it was to be. Her problem perhaps was that the principle of crying wolf too often had not yet been disseminated, and so she overdid and consequently was not believed. Yet clearly, half of social validity is to be a Cassandra. The trouble with trying to find a Cassandra is the extraordinary abundance of

applicants, all quite certain about future disasters, ready to caution anyone seen to be puzzling over how behavior works: You never know until too late which, if any, of the numerous competitors for Cassandra's job was correct. In this context, the nomination by the Winett team of independent, objective, widely sampled, usually reliable epidemiological data seems to accomplish the standard aim of science—to replace apparently magic personal skills with valid objective methods that anyone can apply (given the proper graduate training).

These procedures may be an improvement over current practice for some features of social validity, with some research questions and populations. These procedures are very appropriate to their example of reducing the risk of HIV infection in adolescents. However, they are not appropriate for many of the problem behaviors addressed by behavior analysts. Is it necessary (or possible) to collect epidemiological data to establish the importance of decreasing self-injury in people with developmental disabilities? If concerned parents seek intervention to teach independent play skills to their young child, is their concern enough to constitute importance? Winett et al. (1991) clearly state that the lack of epidemiological data does not preclude a problem from being socially important; however, overreliance on epidemiology may dissuade behavior analysts from developing new, and perhaps more effective, methods to assess the social importance of the problems targeted by behavior-analytic intervention programs. These interactive steps may be appropriate in assessing certain components of social validity, but any a priori procedure will be inadequate to assess a comprehensive view of social validity.

The question of whether or not to conduct social validity assessments was answered most eloquently by Wolf (1978). The important questions currently facing the field of applied behavior analysis are how to conduct these assessments and how to apply these data to improve behavioral interventions. Winett and his colleagues describe a methodology that may be useful in answering these questions; in particular, the framework they provide suggests questions and challenges that individual behavior analysts can use to further their own work in this area.

REFERENCES

Winett, R. A., Moore, J. F., & Anderson, E. S. (1991). Extending the concept of social validity: Behavior analysis for disease prevention and health promotion. *Journal of Applied Behavior Analysis,* **24,** 215–230.

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 January 15, 1991
Initial editorial decision February 19, 1991
Final acceptance March 1, 1991
Action Editor, E. Scott Geller