6

BEHAVIOR: ITS UNITS, DIMENSIONS, AND MEASUREMENT

BEATRICE H. BARRETT
W. E. Fernald State School
JAMES M. JOHNSTON
Auburn University
H. S. PENNYPACKER
University of Florida

Descriptions of behavioral assessment have become increasingly diffuse over the past few years (Cone & Hawkins, 1977; Mash, 1979; Mash & Terdal, 1981a; Nelson & Hayes, 1979; O'Leary, 1979). A perusal of recent publications may, for some, provide a déjà vu experience. Has this new field redefined behavior? Redefined it to the extent that behavior now encompasses the hypothetical metalistic activities that behaviorism eschewed decades ago (Skinner, 1938; Watson, 1924)? Or has there emerged a new brand of assessor schooled in traditional procedures and merely applying the popularized label "behavioral" to procedures once called "psychological?" If the latter, then the definition of behavior has, perhaps, remained intact but has merely been overlooked or not thoroughly understood.

Relabeling or creating new or slightly different descriptive vocabulary is not an uncommon way of attracting the professional community. "Schools" of psychotherapy in the 1940s and 1950s were created by, or at least reflected, this conceptual pastime. What practitioners said and wrote about what they did was more different from "school" to "school" than what they really did. From Horney and Alexander to Mowrer, Dollard, and Miller, the major differences were in verbiage. Operations of differently affiliated practitioners were more similar than they were different (Barrett, 1958). But the verbal-conceptual differences became the banners of camp followers whose professional identities became secure in conceptual distinctions often tenuously related to different practices

"Behavioral assessment" today describes procedures that, by and large, were once called "psychological assessment." A quick glance at a mental testing textbook popular in the 1940s and 1950s (Greene, 1941) provides such a comparison. Although operations listed in today's books and journals (Cone & Hawkins, 1977; Hersen & Bellack, 1976; Mash & Terdal, 1981b) are couched in the language of "specificity," the contents of procedures and concerns have changed so slightly that the oft-reiterated differences seem more apparent than real. Distinguishing characteristics are now said to be more conceptual than methodological (Cone, 1977; Goldfried & Kent, 1972; Mash & Terdal, 1981a; O'Leary, 1979). A recent survey of practitioners is not inconsistent with this interpretation (Swan & MacDonald, 1978).

Amalgamation and the cognitive revolution are fashions of the times—decidedly concordant with "behavioral"-"psychological" equivalence and the conceptual incorporation of some brand of "behavior" into the mainstream of current ideology. The implications of such an amalgam are critical to the viability of behavior science methodology as a major investigative medium in the design of human-service technology. Adoption of "behavioral"-"psychological" equivalence forfeits both the heritage of behavior analysis and the advantages that accrue from its natural science methodology (Johnston & Pennypacker, 1980). The forfeiture returns assessment of human behavior to its prebehavioral status. Readoption of psychometric methods and concerns, so prevalent in today's literature, indicates that the process is well under way. What will be sacrificed if a social science approach to behavior assessment regains its former status?

CURRENT CONCEPTS AND PROCEDURES

Assessment of human behavior purports to examine the functional relations between "meaningful response units and their controlling variables (both current environmental and organismic) for the purpose of understanding and altering behavior" (Nelson & Hayes, 1979, p. 1). Other than the distinction between "altering" and "improving," and the parenthetic recognition of organismic variables, this definition is essentially similar to that of applied behavior analysis by Baer, Wolf, and Risley (1968) in their standard-bearing article. The latter authors sought to distinguish the fledgling field from its laboratory-based heritage. Similarly, the recent literature on behavioral assessment deals repeatedly with its identity as a new field, distinct from traditional "psychological" assessment.

Instead of assuming that behavior is a product of an underlying personality structure resulting from early childhood experiences, behavior assessors treat observed behavior samples as the primary datum, eschewing traits or other unobservables that might encourage a search for consistency of responding in the presence of differing environmental conditions. Absence of interpretive inference confers the epithet "behavioral." Current measurable variables determine behavior. Therefore, assessment should focus directly and objectively on behavior itself in relation to the contemporary observable conditions that account for its maintenance and change.

It follows from these descriptions that behavioral assessment might be expected to approach its subject matter with procedures as distinct as its identity characteristics—procedures designed specifically for functional analysis of human behavior. But that does not appear to be the case. Although direct observation appears in all lists, current literature indicates that interviews are the most frequently used of all behavioral assessment procedures (Haynes & Jensen, 1979; Linehan, 1977; Swan & MacDonald, 1978). Checklists, ratings and rankings by clients and/or others, and self-report questionnaires and inventories are popular methods of quantifying behavior descriptions. Despite its emphasis on specific observable responses and the deterministic role of current environmental variables, behavioral assessment methodology may also include personality inventories (Goldfried, 1977). Moreover, standardized psychometric tests of achievement and intelligence, although routinely classified as nomothetically rather than idiographically oriented, are acceptable (Cone & Hawkins, 1977) if not desirable tools of the behavior assessor (Nelson, 1980).

As long as the obtained responses are treated as behavior samples rather than as signs of inferred causal conflicts or constructs, these conventional tests are admissable in the domain of behavioral assessment (Goldfried, 1977; Nelson, 1980). But behavioral constructs, likened to MacCorquodale and Meehl's (1948) intervening variables, are also admissible, though not considered causal. Behaviors observed in specific situations now define "behavioral" constructs. Their rated presence or absence yields values often summarized in the form of a total score (Linehan, 1980).

Such practices are reminiscent of psychometric efforts to evaluate personality traits or predispositions. Yet behavioral assessors vociferously deny trait theory on the grounds that generality has yet to find convincing empirical support (Cone & Hawkins, 1977). Although a more comprehensive temporal context has recently been argued (Mash & Terdal, 1981a), current measurable variables constitute the general focus of behavioral assessment. Intraindividual variability of responding across different situations is not only the subject matter but a methodologic imperative of behavioral treatment evaluation.

The stated focus of behavioral assessment appears to be inconsistent with the majority of its procedures. To obtain the objective samples of behavior that might reveal the variations of interest requires repeated assessment with instruments that retain sensitivity across repeated applications. Normatively derived tests and devices such as questionnaires, rating scales, and the like are designed for global single-sample inferential assessment of status relative to a standardization or comparison sample. Employed in traditional psychological assessment and associated diagnostic classification, such instruments are not designed to assist in formulation of individualized treatment or ongoing evaluation of its effects. None facilitate identification or analysis of the variables that could be manipulated to alter the behaviors they purport to describe (Skinner, 1953). None are conceived to track the course of individual behavior changes in response to treatment interventions. In short, none are suitable for integration into a self-corrective process of ongoing behavioral treatment evaluation. More importantly, none derive from or are consistent with either the concept of behavior or the methodology that gave rise to behavioral treatment.

ROOTS AND REMNANTS

Behavior analysis brought a new and different methodology to human service—the methodology of the natural sciences. Rooted in the interactions of individual organisms and their environments, its methods are sensitive enough to record individual behavior changes through time and powerful enough to verify treatment effects without recourse to the group statistical tests that had previously masked their presence.

Offering automatic definition of responses, automatically programmed stimuli, and a wealth of contingencies that produce rapid changes in the behavior of individuals, it bypasses human observer error and permits behavior to speak for itself. Its continuous automatic recording shows moment-to-moment fluctuations and the temporal dimensions of treatment effects. Standard response definition and recording units complemented with replicable environmental conditions facilitate comparisons of different effects not only for a given individual but across individuals as well.

Careful design of the behavior-analytic environment provides functional description of deficits in the retarded (Barrett & Lindsley, 1962; Barrett, 1965, 1969, 1975), prosthetics for the handicapped (Lindsley, 1964), analysis of psychotic behavior (Lindsley, 1956, 1960), autistic behavior (Ferster & DeMyer, 1961, 1962), and of functionally defined leadership, cooperation, and competition (Azrin & Lindsley, 1956; Cohen, 1962; Cohen & Lindsley, 1964; Hake & Vukelich, 1973; Hake, Vukelich, & Kaplan, 1973; Hake, Vukelich, & Olvera, 1975; Lindsley, 1966), conditioning history (Weiner, 1964), effects of drugs (Lindsley, 1962a) and depth of anesthesia (Lindsley, Hobika, & Etsten, 1961), and demonstration of contingency control of neurogenic tics (Barrett, 1962) and stuttering (Flanagan, Goldiamond, & Azrin, 1958), fine-tuning of the environment to achieve errorless stimulus

shaping (Sidman & Stoddard, 1966, 1967), and the content-related communication changes during psychiatric interviews (Lindsley, 1969; Nathan, Schneller, & Lindsley, 1964). The differential reinforcing value of television program content (Lindsley, 1962b) and the illusive domain of human preference also yield to analysis by these methods (Lockhart, 1979; Morgan & Lindsley, 1966; Schroeder & Holland, 1969). The latter references afford ample illustration of the value of functionally defined response classes and automatic programming over conventional verbal reports (Azrin, Holz, Ulrich, & Goldiamond, 1961) in assessment and analysis of human behavior. In more recent application, automatic response definition and recording reveals when palpation of a human breast model detects individual lumps as small as 2 mm in diameter—a standard for training detection of lumps *in vivo* (Adams *et al.*, 1976; Hall, Goldstein, & Stein, 1977).

Sensitive design of the subject's environment makes complex aspects of individual behavior accessible to analysis of the variables that compose and alter them. Temporal dimensions of their natural fluctuations emerge in permanent tracings drawn directly by response emissions through time. Wide ranges of individually different behavior patterns demonstrate the sensitivity of these methods. Yet their basic functionally defined response classes show lawful variations of human behavior in response to alterations of its environment and thus affirm the universality of operant principles revealed by analysis of infrahuman behavior.

The malleability of human behavior demonstrated by these methods lured clinicians from their psychometrics and group studies. Applications of procedures derived from the methodology of experimental behavior analysis became so successful with such a wide-ranging clientele as to change the complexion of the human service field. But the translation continues to sacrifice much of both the core methodology and its conceptual foundations (Birnbrauer, Burchard, & Burchard, 1970; Deitz, 1978; Fraley, 1981, Michael, 1980, 1985; Pennypacker, 1981; Pierce & Epling, 1980; Woods, 1980).

Today's literature refers to the remnants of the parent methodology as "direct observation"—certainly not attributable to behavior analysis per se. Distinguished from indirect verbal reports by clients or others, first-hand observation of clients' behavior in either contrived or natural environments requires little or no inference. It thus meets the exclusion criterion for a behavioral assessment approach. Substituting for the automatic response definers of the original system are people with coded forms to remind them what to look at and for Instead of direct, continuous recording automatically drawn by occurrences of electromechanically defined behaviors, observers look and record on intermittent schedules. Post hoc counting of observer tallies or the checked versus unchecked squares on a recording form now replaces automatic cumulative counters. Interval timers that once programmed contingencies for the behaver now signal observer behavior. Elapsed time of the observing period rarely appears in reported data.

Thus response defining is indirect, filtered through human bias and fatigue; recording is discontinuous; and timing applies not to the behaver's activity but to the observer's. Replacing online measures of response frequencies, the only measurement of what occurs during an observation period is a summary total, not of the client's behavior but of the proportion of observer looks and listens that were checked for seeing or hearing the behaviors of interest—the number of intervals marked. Reporting then further removes the available information from retrieval by relating these counts to the number of intervals used. Only by careful detective work can a reader sometimes discover how much time went into the observation process. "Goodness" of the observation products depends on another observer's tallies or checks of what *that* observer saw or heard. Such a standard of comparison reveals nothing about the accuracy of either observer's records.

Attempts to determine the relative frequency of the observed behaviors often appear in rating scales denoting levels such as "never," "occasionally," "usually," "frequently," and "always." Numerical values usually designate each position on the scale, and summation yields a quantified subjective estimate of whatever behavior is being rated (Walls, Werner, Bacon, & Zane, 1977). Today's rating scales purport to focus on "behavior" rather than on attributes, opinions, or feelings. Unfortunately, revising or relabeling the content of such scales does not improve the quality of their measured products. Nor does it bring us anything resembling adequate information for the stated mission of behavioral assessment. Subjectivity and insensitivity to anything but gross pre- and post-treatment behavior changes render such methods inappropriate for a functional analysis of change-inducing variables. These are some of the shortcomings that prototypical behavior assessment methods once put to rest. Alas, only temporarily.

The once-distinctive characteristics of behavior assessment methodology have succumbed to a melding process. From the residuals, there appear to be three recognizable imprints. First, "behavioral" content appears in clinical information (Angle, Hay, Hay, & Ellinwood, 1977). Rating scales and checklists tend to concern what people do rather than how they think or feel, and descriptive accounts of behavioral assessment reveal varying degrees of "behaviorese," routinely calling attention to the conceptual distinction between a behavioral versus a psychodynamic interpretive process. Second, consistent with a basic tenet of applied behavior analysis, self-monitoring now enjoys considerable status as an assessment technique (Hayes & Cavior, 1980; Mahoney, 1977; Nelson, 1977a, 1977b; Swan & MacDonald, 1978).

Counters and data forms help systematize and objectify its operations in the service of self-control. Third, the time-honored observation process itself is now more "systematic" by virtue of often exhaustively codified behavioral definitions and recording forms designed to increase accuracy and precision of behavior transduction by human observers.

That precision is now a buzzword rather than an accurate description of behavior assessment methods suggest that a cost-benefit analysis would raise major questions about the contribution of current methods to the knowledge base of the field. Expedience and convenience now accompany social validation and consumer decision making, recently recommended as an integral stage in formulating and evaluating human behavior research (Wolf, 1978). Assessment methods reflect growing distance from a natural science concept of behavior. As the discrepancy advances, the methods that emanated from the science of behavior give way to those it once replaced. This cyclic process, so evident in sociopolitical history, has no precedent in the domain of a successful science whose foundations should serve to support its advances. Without successive broadening and strengthening of its fundamental principles and methods, it falls prey to common sense, the source of its original questions. And "common sense . . . seldom is aware of the limits within which its beliefs are valid or its practices successful" (Nagel, 1961, p. 5).

This chapter returns to a natural science concept of behavior. It attempts to clarify a functional definition of behavior and its constituent response classes and to show how methods for quantifying behavior relate to its definition. It views assessment as measurement. And it adopts the position that clients in clinical settings should enjoy at least the methodologic privileges accorded small animals from whom we learned the principles of behavior.

BEHAVIOR: THE SUBJECT MATTER OF ASSESSMENT

While conceptual evanescence tends to reflect contemporary influences, the properties of behavior remain unchanged. Conceptual discrepancies and methodologic inconsistencies are common companions. Drift in one permits vagaries in the other, and failure to acknowledge their interdependence is a costly oversight. In the domain of human behavior assessment, development of maximally effective behavior remediation technology requires essentially the same basic information as does pursuit of its scientific analysis (Johnston, 1982). For this reason, and whatever their conceptual affinity, the investigative methodology of both behavior therapists and their behavior analyst brethren should address the universal properties of the phenomenon they seek to understand—behavior.

DEFINITION

An adequate definition of behavior includes both its interactive and its dynamic properties and, in addition, provides for its accessibility via standard measurement systems of natural science (Johnston & Pennypacker, 1980).

> The behavior of an organism is that portion of the organism's interaction with its environment that is characterized by detectable displacement in space through time of some part of the organism that results in a measurable change in at least one aspect of the environment. (p. 48)

IMPLICATIONS FOR MEASUREMENT

Each component of the definition specifies a characteristic of behavior that should find expression in its measurement. Each implies certain capabilities of a suitable measurement system that distinguishes behavioral from traditional psychological measurement.

Intraindividual Phenomenon.

Consistent with a natural science concept of behavior and its biological significance in adaptation and survival, the science of human behavior, like its progenitor, experimental medicine (Bernard, 1865/1957), views the behavior of individual organisms as the focus of analysis. Variability in an organism's interaction with its environment is the subject matter of the science. To observe it directly, the activities of other organisms are unnecessary. Ample supporting evidence demonstrates that lawful relations emerge from the study of individual organisms interacting with their environments. Recourse to groups cannot describe the behavior of an individual within them, but generalities can and do arise from aggregations of lawfully determined individual behavior.

Sensitivity to the nuances of individual behavior and its interaction with its environment is, then, a principal requirement of a suitable measurement system. Measures designed to describe large samples do not meet the sensitivity prerequisite. Especially exempt are measures derived from either assumed or obtained variability of large samples. Measures that reflect other sources of variation cannot accurately represent variability of individual

behavior. Psychometrically derived measures are thus inappropriate. Consistent with definitive statements about behavioral assessment, analysis of intraindividual phenomena provides the methodological approach shared by practitioners and researchers alike. It thus opens the search for lawfulness to participation by both clinic and laboratory.

Movement

"Detectable displacement in space through time" requires movement. Independent states of the behaver attributed to emotional or physiologic conditions such as being anxious or being hungry denote no movement component, nor do static postures such as eye contact, attending, or out of seat. All are capturable in a snapshot. All are prevalent examples potentially reconstructible with a cadaver and well within the skills of a sculptor.

Similarly, independent conditions or changes in the environment that happen to the behaver but require no interactive component fall outside the definitional limits. Examples include being pushed, falling downstairs, and being given a cookie. Being denied access to a meal may not function as a behavioral event until and unless the behaver engages in activity that avoids or escapes such denial.

The environment may be internal or external. The epidermis need not be a barrier to behavior. Thus both visible actions and intraorganism activities such as muscle potential changes and heartbeats qualify. The movement requirement provides for detection by means of public responses or by instruments which convert response occurrences into measurable changes in some aspect of the environment. Thus, gross displacement of objects by some act of the behaver as well as amplified transduction that deflects a needle or changes a display on a monitoring screen both exemplify the movement that makes various forms of behavior accessible to standard scientific measurement. Private events externalized via publicly observable responses are no exception. "The line between public and private is not fixed. The boundary shifts with every discovery of a technique for making private events public" (Skinner, 1953, p. 282).

Continuous Process

In addition to providing for detection, "displacement in space through time" reminds us that behavior is continuous in nature. Discrete static events and states without movement or interactive aspects are not behavioral events. Behavior not only takes time to occur and occurs in time; it also occurs through time. Time is thus a fundamental and universal parameter of behavior. Without temporal dimensions revealing trends and interactions, quantitated description is both incomplete and inappropriate for analysis of behavior and its controlling variables.

The continuous nature of behavior requires repeated observation and measurement. Procedures that sample many behaviors for a snapshot summary are shortcuts that prevent analysis of the continuing interaction of the organism and its environment. Those that focus on a few behaviors but sample them discontinuously also preclude tracking behavior as it changes through time in accord with environmental alterations. Lawfulness emerges not through restricting our range of vision but by expanding it to encompass nature's fullest and richest picture.

Measurable Change in Environment

Behavior and its environment are inseparable. Their relations are our subject matter. By definition, behavior affects the environment. The changes it produces afford our only entry to its measurement. Detection selectively scans the environmental traces of behavior; transduction converts them into measurable form. If we profess to study behavior in its own right, we deal directly with its environmental traces as they occur through time. Automatically transduced events that depress recording pens, or internal changes revealed on a cathode ray tube, or self-recorded private events registered by pushing a switch are examples.

Caveats are in order when the environmental changes produced by behavior become merely symbols of inferred conditions, past or present. Questionnaires, checklists, and the like are examples. In these cases, behavior-produced environmental changes themselves (check marks, encircled numerals) are usually not the subject matter of assessment (except in some educational contexts), even though the stimuli reputed to be controlling them—words on paper—may refer to what an individual did. Inferential measurement is neither conceived for nor designed to provide information about behavior itself.

Irrespective of its diverse conceptual-methodologic embellishments, the language of behavior assessment is the language of measurement. Description, comparison, and prediction underlie all behavior assessment whether undertaken for classification, placement, treatment formulation, prognosis, treatment evaluation, or collection of normative data. The degree to which measurement will accomplish these purposes depends on the fidelity of the correspondence between what is being assessed or analyzed and the measures chosen to represent it.

Variability of the *behaver's* interaction with the environment is the subject matter of both the natural science of behavior and behavior assessment. The spectrum of behavior-descriptive labels, their wide-ranging behavioral

referents, and the varying measurement procedures currently applied to them all introduce sources of variation that confound the behavioral phenomena of interest. From the once successful search for lawfulness, behavioral assessment now accepts a state of definitional-methodological anarchy (Cone & Hawkins, 1977). It is, after all, superficially consistent with individualized treatment and with personalized interpretation of the client's behavior. Unfortunately, it is also consistent with the assumption of inherent variability of human behavior—the assumption that originally gave rise to the social sciences' statistically derived concept of human behavior variation.

Practices that distort, mask, or otherwise misrepresent properties of behavior are counterproductive, not only to the development of effective clinical technology, but also to the advancement of the natural science of behavior on which it rests. Sources of such illusory variability, still unrecognized in much of the behavioral assessment literature, lie in decisions regarding the definition of behavior units to be studied, methods for quantitating their properties, and procedures by which they are observed and recorded.

UNITS OF BEHAVIOR ANALYSIS: RESPONSES AND THEIR FUNCTIONAL CLASSES

While various traditional social science approaches may assist in selecting the problem area to be addressed, behavior assessment begins with defining the behaviors selected for measurement during the intervention process. These may include behaviors targeted for development, behaviors designated for management, support behaviors that could serve as alternatives or replacements, and collateral behaviors that may be important to monitor for ancillary or secondary effects.

Measurement accessibility requires dissection of the stream of ongoing behavior-environment interaction into response classes appropriate for quantitative description. To fully represent the properties of behavior discussed earlier, response classes chosen for analysis must consist of movements that produce effects on the environment as they occur. To preserve the functional nature of behavior, response class definition should account for demonstrated controlling antecedent and subsequent environmental variables.

RESPONSE CLASSES IN CONTEMPORARY LITERATURE

Today's cadre of behavioral assessment techniques concerns a widely varied assortment of responses. The more sophisticated classes include physiologically or motorically activated switch operations associated with various replicable environmental conditions. Other less-refined versions include pencil slashes in printed boxes, encircled numerals, and check marks executed by the behaver in response to printed survey or scale items or by an observer/informant reporting on the behaver/assessee. One of the most prevalent classes of numeral encircling refers to verbal descriptions of remembered events, feelings, or states bereft of any responses. Levels of prompting required for the behaver to execute responses appear in such scales while the behaver's responses themselves remain unmeasured. Even further removed from behavior are observations of clients' stances such as "hands in lap," "in seat," and "on task"—again notated as checks or slashes in time-ruled coded squares. Summation then provides a score connoting an amount of such broadly conceived classes as "appropriate behavior," "assertiveness," "social interaction," or "independence," each defined by its own measures.

As units of behavior analysis, response class definitions must be stable and independent of the units used to measure them; otherwise, variability in the response-class definition will reflect itself as measured variability in the phenomenon under study. Moreover, response classes should represent the interactive property of behavior, specifying what responses do to affect the environment and, ideally, what aspects of the environment control their occurrence. In addition, definition of the classes selected for assessment should be sufficiently molecular to permit measurement of the variability necessary for sensitive analysis and successful treatment. Equating response classes and measurement units and failing to provide a properly sensitive level of analysis obscures the contextual properties of behavior-environment relations which are, after all, the given ingredients of intervention planning.

RESPONSE CLASS FUNCTIONS AND ASSESSMENT

When undertaken for clinical purposes, behavior assessment inventories both the desirable and the aberrant or deficient functional relations of a client's interactions with the environment. Remedial manipulations of the environment then broaden or otherwise modify the client's behavior by recomposing the existing array of selected functional response classes. Tactics to achieve this objective include: (1) altering the conditions under which selected responses occur, (2) providing more complex new functions for existing responses, (3) generating new movements that become functional responses by virtue of the environmental changes they provide, (4) changing some dimensional quantity of an already functional response class such as its frequency, and (5) modifying the topography of a functional response class to produce greater environmental effect. The nature of behavior intervention by definition deals with the functions of responses extant in the client's

repertoire and the manner in which the client's environment sustains presenting problematic behavior.

Response definition most commonly focuses on its form or *topography*, its most obvious aspect. Speech sounds and dance movements are examples for which specific criteria are legion. Success—reinforcement—accrues to the behaver for responses meeting these standards. But when extended to most response definitions, there appears to be little evidence to indicate an invariant relation between topographically similar or superficially classified behaviors and their functions for the behaver. Thus, on the deviant side of the inventory, "self-injurious behaviors" designates a socially labeled class selected for intensive treatment due to its aversive effects on most observers and its frequent life-threatening outcomes which society attempts to prevent. The class encompasses a range of topographies limited only by the interaction history and/or the creativity of the behaver: eye gouging, rectal digging, head hitting, nail biting, head banging, screaming, flesh and muscle chewing, and so on.

All may conceivably serve similar *functions* for the behaver. Moreover, their functions for the behaver may be quite different from the functions these behaviors serve for the observer-classifier. Yet neither the label of the class nor the interventions selected to treat it may reveal the effects such behaviors achieve for the behaver. Similarly, a single well-defined topography such as head hitting may serve multiple functions by which the behaver escapes or avoids aversive conditions or obtains social, sensory, or food reinforcement depending, perhaps, on the antecedent conditions that have become associated with each of these environmental effects. If topographic differences and/or societal reactions continue to remain the principal foci of assessment, the permutations and combinations of behaver-focused environmental effects that may exist across different individuals will present a melange that defies systematic inquiry.

Lawfulness is more likely to emerge from analysis of response classes defined in terms of the conditions under which they occur and the effects they achieve for the behaver than is the case with classes defined by the form of the movement in question or its effects on third parties (Skinner, 1935). The class of behaviors knows as imitative emerged from studies of widely varying topographies which resulted in reinforcement if they matched the topographies presented by a model. Systematic observation of covarying topographies during development of imitative behaviors revealed that some nonconsequated topographies covary with those meeting the matching (imitative) criteria. Thus commonalities of function demonstrated across constituent topographies under a similar set of conditions expanded the imitative response class (Baer, Peterson & Sherman, 1967).

As an environmental effect or possible treatment technique rather than as a response class emitted by the behaver, contingent imitation of behaver responses by other persons may eventually show lawful relations that define yet other response classes generated by the reciprocal imitations they produce. The different results of reciprocal imitation associated with differing topographies—vocal and motor—(Birnbrauer, 1979) opens the phenomenon to further analysis that should delineate classes associated with the accelerative versus decelerative functions of this environmental effect. In any event, similarities and differences obtained in response to a given environmental antecedent or subsequent event should further clarify the common functional relations of topographically disparate behaviors.

Clearly, a number of topographically different responses showing the same relation to common antecedent and/or subsequent conditions may constitute a functionally homogeneous response class. Such "functional clusters" raise the possibility, indeed the likelihood of altering multiple topographies by manipulating conditions for a single topography within a class (Wahler, 1975). By capitalizing on the empirically substantiated covariance of different response class members, maximum therapeutic gains might be achieved by minimal effort and cost. Moreover, all individuals with a given target sharing a common cluster or response class composition with other individuals might then be treated similarly with expectation of comparable cost-benefit ratios.

Treatment packages involving multiple manipulations as prescriptions for behaviors described only by their topographies preclude this type of assessment. Without empirical verification of the functions served by varying topographies for given individuals under the conditions of their everyday environments, the package risks targeting behaviors that, because they may not share homogeneous functions, could offset or mask whatever effects might otherwise be obtained. Thus an accelerative effect on one class of responses might accompany a decelerative or neutral effect on another class, yet a single topographically defined class may encompass the two functionally different classes.

Such combinations of effect are not predictable, and their likelihood seriously limits both the validity and the generality of package application. The topographical approach retards delineation of functional classes that underlie the vagaries of personal or social linguistic practices. More importantly and for reasons just mentioned, topographical definition may produce illusory variability that is not recognized and therefore not susceptible to either control or analysis.

The purpose of response definition is to derive units of analysis that are accessible to quantification by standard units of measurement. Functionally equivalent events are more likely to share common determinants than are those that resemble each other in superficial or irrelevant respects (Johnston & Pennypacker, 1980). If behavior assessment defines its unit, the response class, to account for its own relation to antecedent and/or subsequent events, associated measurement procedures will less likely confound irretrievable sources of variability. Orderliness is more likely to appear along the "natural lines of fracture" (Skinner, 1935).

FUNCTIONAL CLASSES, PROGNOSIS, AND GENERALITY OF TREATMENT

Response classes defined by empirical verification of functional homogeneity across their various topographies constitute the maximally efficient pretreatment assessment baselines from which to select and against which to evaluate the effects of remedial procedures. The likelihood of achieving generality of effect is enhanced by ongoing pretreatment measurement of pertinent functionally related response classes that, by definition, specify at least some of the conditions under which effects will be sustained when treatment terminates (see Birnbrauer, 1981). Pretreatment measurement of functionally defined response classes conducted over a sufficiently long period should display naturally occurring irregularities that may obscure or interact with treatment effects. Such information facilitates adjustment of the temporal dimensions of treatment to override or capitalize on existing response interactions. Optimal reliability and generality should result.

A body of knowledge permitting adequate prognostication of specific and secondary treatment effects would greatly simplify therapeutic endeavors. Treatment protocols could be matched to selected characteristics of presenting problems with a foundation of supporting evidence indicating likely effectiveness. The province of behavior intervention involves altering the functional relations between behaviors and the environment. For this reason, the functions of response classes become the focus of assessment. Given the overwhelming evidence that the environment is orderly and lawful, there is ample reason to postulate lawfulness in the interactive behaviors that it generates and sustains. What society considers aberrant behavior is no exception. When a procedure changes the behavior of an individual, it does so only by controlling the environment that supports such behavior. Therapeutic procedures do not change individuals in any permanent sense, only the environments with which they interact. This applies to pharmacologic interventions in the internal environment, manipulation of the environment outside the epidermis, or a combination. Commonalities in the relations between groups of topographically different responses and their supporting environmental effects will then determine generality of treatment effects across these behaviors. Commonalities in the response class constituents of individuals sharing a similar deviant form of behavior will determine whether cross-subject generality will result from a given environmental alteration. Similarly, commonalities in the cross-setting controlling stimuli of a deviant behavior and/or its targeted alternative(s) will determine the setting generality of the outcome and the suitability of selected alternative topographies.

It is the classification of such response–procedure relations, that is, the commonalities across topographies within response classes, that will facilitate formulation of criteria for optimal response class–treatment matches. In order for such commonalities to emerge, environmental modifications must themselves undergo analysis to sort out those components that are functional for specified response classes from those that are unnecessary or even counteractive to maximal effect.

Elucidating the functional relations shown by pretreatment behavior is a major mission of behavior assessment. It is these relations that are altered in the course of intervention. The interactions between pretreatment behavior and behavior changes during treatment constitute the basic information for development of prognostic indices. These, in turn, depend on evolution of effective, efficient matching of response classes with intervention procedures across clients with diverse behavior topographies, but sharing the common functions that orderly environmental reactions produce. Such an enterprise finds its parallel in bioassays that determine ingredients of a pharmacologic compound most likely to produce the desired effects and least likely to produce simultaneously undesirable side effects. The form of the pill is irrelevant. The job of the behavior assessor is quite similar to that of the bioassayst: to sort out the functions of existing topographies and the manner in which these functional relations change as various elements of a treatment (compound) package are introduced. However, without adequate pretreatment analysis of the functional relations to be altered, there is little hope of discovering the lawfulness of behavior that synthesis of more reliable and more general treatment effects requires.

QUANTIFYING THE DIMENSIONS OF BEHAVIOR UNITS

Determining the functional homogeneity of responses assigned to a class requires repeated measurement of their occurrences under different conditions. Moreover, methods chosen for this purpose should permit quantitative description and direct comparison across conditions and across responses as they occur in time. Selecting the most suitable measure is critical not only for reliability and generality of findings but also for description of the behaviors to which they apply.

Implicit in the choice of measurement is the distinction between sources of variability. As a dynamic organism–environment interaction, any behavior displays its own natural variability. Controlled alterations of environmental conditions during experimentation or treatment impose added variability. It is the business of behavior assessment and analysis to quantitatively distinguish one from the other and to explain the changes in measured behavior resulting from changes in experimenter-imposed conditions.

Sources of variability that are neither a natural property of the organism–environment interaction nor a result of experimental manipulation of that environment are considered to be "extraneous." Intraorganismic physiologic and developmental processes and various aspects of the experimental setting are commonly, though often unfortunately, accepted in this category. However, there is another source of variability that much of current practice fails to recognize. Its illusory source is the measurement process itself. If undetected, the resulting measurement system will confound its own sources of variability with those under analysis. No amount of statistical filtration will rid the data of measurement-produced variability. Dignifying the measurement product as "error variance," a common practice in social science methodology, reflects acceptance of intrinsic variability as a property of the organism rather than as a result of individual organism–environment interaction. Clearly antithetical to a natural science approach, the assumption of variability as an inherent or "given" organismic characteristic supports the statistical social science conception of behavior and obviates functional analysis of independent variables that might control intersubject differences (Johnston & Pennypacker, 1980; Sidman, 1960). A science in its developmental phase can ill afford acceptance of such obvious constraint on its potential knowledge base.

BEHAVIOR PROPERTIES

Quantifying behavior units requires assigning numbers and measurement units to selected dimensions representing fundamental properties of behavior. Accordingly, those properties that are both universal and relevant for adaptive functioning are of primary concern. Since behavior occurs in time, it is always time-referenced. The property of a single instance of behavior describing that point in time when the response occurs is its *temporal locus*. That a response takes time for completion is its universal property of *temporal extent*. Repetition through time, required for functional behavior–environment interaction, is the property of *repeatability*. These properties characterize every instance of behavior. Their dimensions are the subject of behavior measurement.

DIMENSIONAL QUANTITIES OF BEHAVIOR

Quantitative description of a behavior property requires selection of a suitable dimension to be measured. Length and weight are examples of dimensional quantities in the physical sciences; and their associated measurement units, centimeters and grams, are invariably related to each. Similarly, each of the fundamental behavior properties denotes quantifiable aspects. Thus the dimensional quantity representing temporal locus is *latency*, and the dimensional quantity of *duration* represents temporal extent or elapsed time. Repeatability through time, represented by *count*, completes the dimensional quantities of single behavior instances.

Additional dimensional quantities emerge for a series or class of responses, none of which is applicable to single occurrences. For multiple instances, temporal locus and repeatability find expression in the dimensional quantity of *frequency*—the number of responses per unit of time—considered by Skinner to be the fundamental datum of an analysis of freely emitted behavior (1938, 1953). That behavior frequencies change over time requires the dimensional quantity of *celeration* (Lindsley, 1969a)—the combination of temporal locus and repeatability which universally quantifies the dynamic property of behavior-environment interactions. The time interval between responses in a series, quantified as *interresponse time* (IRT), combines the properties of temporal locus and repeatability to represent their analogue in the case of more than one response.

Other dimensional quantities describing patterns of responses, such as quarter life and IRTs per opportunity, may describe serial occurrences of a class of responses. The generality and usefulness of these and other dimensional quantities fashioned for application in particular situations require extensive verification.

MEASUREMENT UNITS: THE REFERENTS FOR DIMENSIONS OF BEHAVIOR

Quantification of behavior events employs measurement units selected from appropriate scales representing the dimensions chosen for study. Just as standard units (yards, square feet, cubic yards, ounces, and gallons) describe physical dimensions, so standard units represent each of the fundamental dimensions of behavior. Standard *time units* (seconds, minutes, hours, days, weeks, months, years) are absolute and universal descriptors of both latency and duration. Because both are fundamental dimensions of any instance of behavior, time units are basic units of behavior measurement. No quantita-

tive description of behavior is complete without inclusion of at least one temporal referent.

Countability, denoting the recurring property of any unit of behavior analysis, has as its measurement unit the *cycle* which begins with the initiation and ends with the cessation of a single response. *Cycles per unit time* of the counting period describes the frequency dimension of a class of recurring responses. It combines the latency and countability dimensions applicable only to single instances. This compound unit is universally and invariantly descriptive of any repeatable motion. Simple counting determines the value of cycles.

Change in behavior frequency over time simply combines the second time dimension to become *cycles per unit time per unit time*. Such examples as cycles per minute per week and cycles per day per week describe the temporal changes in frequency seen both in skill proficiency growth and in behavior management endeavors. In the case of interresponse times, the cycle under study begins with the cessation of one response and ends with the onset of the next response. Its measurement unit is, then, the *time per cycle*. Table 6-1 summarizes the relations of behavior properties to dimensional quantities and their corresponding measurement units.

CONTROLLING THE CONTROLLABLE: VARIABILITY OF MEASUREMENT UNITS

Variation in behavioral phenomena, like variation in physical aspects of the environment, requires fixed measurement units for its description. Units of quantitative reference that vacillate from behaver to behaver, from scale to scale, can only confound the true variability of the phenomena being investigated. Use of such evanescent units of measurement not only ensures

TABLE 6-1. Properties, Dimensional Quantities, and Units of Behavioral Measurement

Property	Dimensional Quantity	Unit
Temporal locus	Latency	Time units
Temporal extent	Duration	Time units
Repeatability	Countability	Cycle
Temporal locus and repeatability	Frequency	Cycles/Unit time
Temporal locus and repeatability	Celeration	Cycles/Unit time/Unit time
Temporal locus and repeatability	IRT	Time/Cycle

See Johnston & Pennypacker, 1980, p. 128

limited generality of findings, but liberally bolsters popular assumptions of intrinsic variability beyond the reach of scientific study (Sidman, 1960). If accepted either wittingly or unwittingly, this assumption relegates behavior to prescientific explanatory concepts and precludes refinement of methodology that might reveal the lawful behavior-environment relations upon which a more exact prognostic-prescriptive taxonomy must be based.

Given the numerous sources of uncontrolled variability facing the assessor of human behavior, widespread acceptance of fluctuating measurement units bespeaks inadequate awareness of controllable sources of variability introduced in the measurement process itself. Judicious decision will remove this source of variation if the units included in the measurement system retain constant and universal values throughout the course of study. Such idemnotic units (Johnston & Pennypacker, 1980) reduce replication problems and enhance generality by ensuring direct comparability of measures.

Numerical values attached to time units are universally recognized and apply in any measurement of time. No confusion exists about the amount of time described by *t* minus 1 at Cape Canaveral or at the various tracking stations around the world. Were it not for the precise uniformity of temporal units, the operations of launching, tracking, and retrieving a space vehicle would be calamitous at best. Nor is there any question whether the value of a second will change from day to day. Time units are thus standard, absolute, and universal, applying to every instance of behavior, every series of repetitions, and every observation/counting/recording period. Similarly, counting conventions are universal. As long as the behavior units being counted retain descriptive stability so that the value or definition of each response remains constant throughout a series of repetitions, the response cycle then has a standard referent. Moreover, given accurate timing devices and an equally accurate description of response cycles, variations emerging from repeated measurement cannot be ascribed to measurement "errors."

UNITLESS VALUES

Contrast the descriptive precision and comparative power of the fixed unit and its standard referent with, for example, the rating scale's numerical values anchored only in verbal descriptions connoting subjectively determined amounts of an assumed dimension of behavior, for example, assertion or anxiety. If we accept *Webster's* (1961) definition of a unit as "a determinate quantity (as of length, time) adopted as a standard of measurement for other quantities of the same kind," and substitute for length one of the dimensions of behavior described above, it becomes immediately apparent that no true measurement units are attached to these values. The referents are indigenous to the scale itself, the respondent's personal interpretation of the accompanying descriptors, and the scale writer's interpretation of the often synthetic dimension represented by the scale. Personal histories and linguistic conventions substituted for standard, absolute units thus render applications of arithmetic inappropriate.

Nonetheless, the common practice of providing greater "sensitivity" by increasing the number of scale points reflects the assumption of intrinsic variability. Thus, presence–absence checklists, when expanded into three-, five-, or seven-point or greater scales, permit greater latitude of response individuality. However, expansion of the scale fails to provide a fixed unit from which to determine individual departure.

Further arithmetic digression occurs in practices such as summing unitless "values" across multiple scales to yield "scores," and in arbitrarily scaling such "scores" into profiles for purposes of comparison. Because a given numerically designated point of one scale is in no way equivalent to or representative of the same point on any other scale, such comparisons are clearly spurious. For the same reason, score or profile comparison across subjects is even less appropriate. Greater disparity of subjective interpretation from respondent to respondent increases the variation in referents for any given numerical "value."

Comparison of repeated measures through time constitutes the basis of scientific validation and, ultimately, prediction and prognosis. However, repeated measures composed of units resulting from unique histories or constraints of the examining situation confound *these* sources of variability with whatever change may appear in obtained scores. Any or all of the factors influencing the original choice of positions on a given scale may vary from one administration to another. A given respondent may attach different interpretations to the labeled points on a scale at different times. For example, a subject's definition of, say, "anxiety" may change from one "testing" to another, especially if treatment intervenes. The result is a shift of unknown quantity in the private "value" of scale points. Thus even the most gross change in scale value cannot reliably represent a true change. Without direct measurement in absolute units, verification is impossible.

INDETERMINATE UNITS

Another version of oscillating, indeterminate, or vaganotic units (Johnston & Pennypacker, 1980) is becoming more prevalent in the literature. Psychological attribute testing and large-scale probability testing of groups now threaten to replace the naturalistic analysis of behavior. Units such as MA, IQ, and scores on achievement tests derive from empirically assessed variability in the standardization group. As such, they have no standard referents and no absolute values. Puds, purps, and zuds are easily as indeterminate. Similarly, statistical significance testing further reifies indeterminate units as the "standard" against which to evaluate obtained differences. That such procedures, by virtue of their dependence on vaganotic units, are inappropriate for functional analysis of variability should be self-evident. "The data are asked to perform an impossible task, that of evaluating themselves against standards that are supplied by their own existence" (Johnston & Pennypacker, 1980, p. 371).

Alternatives that are consistent with the definition of behavior offered earlier, that retain the uniqueness of individual behavior values, and that make use of the standard and absolute idemnotic measures distinguishing the natural from the social sciences, are explained in detail by Johnston and Pennypacker (1980).

DIMENSIONLESS QUANTITIES

We have thus far discussed the instability of measures that are widely used to assess actual or synthetic dimensions of behavior. Their contribution of variability is indistinguishable from the other sources of variability in the behavior being studied. Perhaps even more prevalent are quantities bearing no relation to any real or putative dimensions of behavior. Recommended for application to opportunity-bound responses (Baer, 1982), the values associated with a measure such as percentage relate a count representing the subject's behavior, to a count representing the experimenter's or teacher's behavior in the form of opportunities offered. In calculation, the attached units of measurement cancel, and the remaining quantity describes no referent in any natural dimension of behavior. Similarly, relative durations and relative frequencies find no dimensional representation in the fundamental properties of an organism's behavior. Ratios of this nature dispense with the units accompanying the original measures; and this essential descriptive information is often unretrievable from published reports. Even worse, it is impossible to determine sources of variation in a series of such ratios, for instance, as commonly found in skill acquisition data. Percentage of correct responses may increase due to increased correct responding with the same number of opportunities, correct responding remaining constant while opportunities decrease, or a combination of these two. The problem remains identical whenever comparison of any two measures of the same dimensional quantity takes the form of a ratio that is divided (and multiplied by 100) to obtain a percentage.

The fact that percentage correct reveals nothing of the temporal charac-

teristics of individual performance permits some deluding and often damaging conclusions. Its use for comparison presents the risk of equating the skills of two individuals who may reach the same percentage score while performing at levels of proficiency disparate enough to place one in the gifted or normal range and the other in the retarded range. Additionally, this property of dimensionless ratios enables habilitators to obtain predetermined and often misleading quantification of progress simply by adjusting the number of response opportunities offered. Multiple examples of differing proficiency in individuals performing at 100% accuracy levels (Barrett, 1979) highlight but a vignette of the variability in basic behavior dimensions that find no expression in the percentage measure. Unfortunate conclusions of skill "mastery" continue to be drawn from a measure that imposes a ceiling that may be exceeded by any of the actual dimensional quantities of response classes. In failing to reveal sources and types of variability, ratios depicting relative measures of the same dimension are sufficiently insensitive to obscure potentially valuable information in the analysis of behavior.

A simple statement retaining the necessary dimensional descriptors would include, at the very least, correct response frequency and error frequency, indicating that the correct response frequency was x times greater than the error frequency. Treating the two frequencies separately acknowledges the reality that *each is free to vary* independently of the other (Barrett, 1980).

CHOICE OF DIMENSIONAL QUANTITIES

Any behavior can be described with each and all of the dimensional quantities mentioned earlier. Ideally, the use of all would yield the maximum potentially valuable information. However, most resources do not permit what could be construed as a measurement luxury. While choices are often necessary, sensitivity and precision are the primary criteria for scientifically suitable selection. Continuous temporal dimensions such as latency and duration are as sensitive as their timing devices, especially when operated automatically. And, although frequency may be limited by the discrete nature of integers used to count response cycles, the temporal parameter of frequency is continuously adjustable to achieve optimal sensitivity for the particular measurement application. The principal consideration is the smallest detectable change, represented by 1/time unit selected. Akin to increasing the power of a microscope or adding finer gradations on a ruler, increasing the length of measurement time increases measurement sensitivity. The longer one looks, the more one sees of the behavior frequency spectrum (Bourie, 1981).

Dimensional quantities may differ in their sensitivity to variations in responding (Springer, Brown, & Duncan, 1981). Appropriate selection is, therefore, a key to the resultant information yield. Counts of correct responses neglect temporal parameters. In addition, simple counting restricts observable variation to the range of response opportunities available to the behaver or to the observer/recorder or both. More restrictive yet are relative counts whose arbitrary ceilings obscure the orderly variation shown by frequency. Temporal dimensions likewise require choice considerations as, for example, if response duration stabilizes while frequency, though unmeasured, undergoes orderly changes. Automatic transduction with continuous direct recording avoids these constraints by retaining each of the dimensional quantities of behavior, thereby presenting more options for analysis. Its use yields the richest and most sensitive quantification of behavior variability.

CONVERTING BEHAVIOR UNITS INTO MEASURABLE NOTATION

The best test of decisions regarding response class definition, dimensional quantities, and units is the completeness and accuracy of information emanating from the measurement of recorded behavior. The stream of ongoing behavior must undergo filtration into defined response classes. Transducing the dimensional quantities of those classes into a recordable form is the process of observation.

Converting the output of observation into permanent notation is the function of recording. These two pivotal processes partly determine the quality of data from which conclusions will be drawn. Decisions affecting the accuracy of their products are therefore critical to the validity of the information obtained from their use. Regardless of what procedures may be chosen, the correspondence between occurrences of selected responses and their recorded representations is the accuracy criterion against which to evaluate their merit. It is essential to rectify sources of inaccuracy in these processes prior to application. Once in use, such contributors to extraneous variability will be inseparable from their recorded products and therefore unavailable for post hoc appraisal.

OBSERVATION

Detecting amounts of dimensional quantities of response classes and transforming them into permanently recordable events makes ongoing behavior amenable to measurement. Performed either by instruments or by people, the transducing function interfaces selected acts of the behaver with the recording procedure. In the case of automated environments, the interface, in addition to activating a recording mechanism, also may activate the environment to respond according to whatever contingencies have been programmed. For the purpose of detection, instrumentation affords sensors responsive to a myriad of behaviors, and current technology offers a wealth of such devices providing capabilities far beyond those of human observers. When calibrated for appropriate sensitivity and designed for stability of response definition, properly selected transducers eliminate fatigue and conditioning history (bias) factors that have long been acknowledged as sources of uncontrolled variability in human observation. More importantly, they assure stable correspondence among occurrences of selected responses and their transmission to recording devices.

Observer/transducer accuracy, whether human or mechanical, electromechanical, or electronic, depends on sensitivity through time to selected dimensions of the response class to be studied. Ideally, each and every emission of a designated response should activate the observing response. The goal of observation is, then, twofold: to create conditions that (1) sensitize the transducer only to those response dimensions selected for recording, and (2) ensure that the transducer reacts appropriately to every occurrence of the selected behaviors. To approximate the accuracy of automatic devices with human observers requires arranging conditions under which the class of behaviors being studied exerts uncontaminated control over the observing response. To the extent that responses are reliably identified and the continuity of ongoing responses directly recordable, standard units of measurement representing dimensional quantities of behavior are applicable. Absence of standard units precludes accuracy comparisons with any dependable estimate of true values.

Response Class Characteristics

A critical source of observational error resides in the clarity and functionality of response class definitions. Any ambiguity will be reflected in inaccurate observing responses. Due to the subjectivity of their definition, hypothetical states elude detection and ensure inexact measurement. As a result, questions of validity inevitably arise. Restricting inquiry to observable events not only provides for detection but also offers the advantage of idemnotic measurement, therefore a closer approximation to true values.

Given that behavior consists of movement in space through time, the observing response should be capable of reacting to each occurrence of specified movement cycles as well as to their temporal distribution. The latter necessitates one or more timing devices coupled with a response-specific sensor, human or otherwise. If the response class is functionally defined, target response emissions *in temporal association with* their related antecedent and subsequent conditions should occasion the observing response. In addition, observation of response class effects in the form of physical products may satisfy the functional definition requirement. Determining one-to-one correspondence between responses and products as well as the authorship of responses may pose problems, especially in group settings. However, response product observation may be useful when online observation is not practical. If topographic characteristics define the class of interest, mechanical properties (distance, velocity, inertial force) delineating the specified limits of the class form must be observed and transmitted to the recording mechanism.

Definitional Drift

Establishing one-to-one correspondence between response emissions and activation of the transducer is merely the first step toward ensuring accuracy of the resulting data. Faithful transmission of responding over time is an additional requirement. Drift in accuracy may occur if the observing response fails to track each response, either because it cannot keep pace with a high frequency response class or because it does not filter out other events. The latter is more likely to occur with human observers who, unlike devices, react to subtle changes in response class features. Frequent checks on definitional drift and recalibration of the transducer will help provide the desired definitional stability.

Sensitivity Drift

Both human and automatic sensors require frequent calibration checks to assure that the detection response continues to react selectively only to the events of interest throughout the course of study. Procedures for training and calibrating human observers, including self-observers, are discussed below. For the moment, it should be clear that assuring accuracy of observation is a prerequisite for application of standard measurement procedures.

There are some safeguards for reducing the variability in observer reactions. First, careful selection of observers to rule out obvious biases is a necessity. Second, rigorous training should incorporate the full range of stimulus control technology to achieve accurate reactivity of the observing response. Third, frequent calibration should check and adjust the correspondence between observer behavior and a set of true values. All such observer preparation should take place under carefully controlled conditions with subjects offering the full range of all pertinent response class dimensions. In addition, observer recalibration should take place at intervals throughout the course of study to ensure continued sensitivity to the changing range of values that may result from treatment effects.

Temporal Dimensions of Observation

Another usually undetected source of observational inaccuracy occurs in the temporal programming of the observing response. Just as behavior occurs continuously through time, its observation should track the entire course of its occurence. Continuous and precisely response-synchronized observation will yield the most complete account of naturally occurring fluctuations in behavior. Observation scheduling that departs from the continuous and complete ideal will necessarily produce an intermittent and therefore less than true account of behavior flow. The question becomes one of the extent of observation-imposed "noise" that reduces the accuracy and stability of the obtained data. The more complete the observation of all responses in the class, the more closely will it yield true values of the behavior dimensions being investigated.

Complete observation of all responses in a class is most often not feasible. Empirical analysis should determine the maximally sensitive frequency and duration of observation periods. Since most behavior dimensions vary in response to environmental contingencies, prior analysis of samples under different conditions should reveal the nature of temporal and setting variations in dimensional quantities. An appropriately designed observation schedule can then minimize loss of information while maximizing the representativeness of the sample with respect to periods of relative stability and variability during prestudy probes.

Obviously, frequency of responding will play a major role in observation scheduling, irrespective of which dimensional quantities are being assessed. Low frequency behaviors may demand longer observation periods than behaviors of higher frequency. Whatever sampling schedule is used, the objective is to obtain the best approximation of true parameters of the response class under study. The more continuous the observation, the less likely will there be misrepresentation of naturally occurring temporal variations.

Intermittent sampling within the observation period imposes experimenter-determined but unanalyzed discontinuity on the observation product. In the absence of continuous records as a comparison standard, the extent of error resulting from incomplete observation remains unknown. Yet it is confounded with the natural variability of the behavior being studied as well as with the variability imposed by treatment. The result is reduction in both the accuracy and the generality of conclusions.

Popular forms of incomplete observation include time-sampling with its momentary observations, alternating with relatively longer intervals between observations. Although inter-look intervals may be regularly scheduled, the cumulative observation time within a session may be only a minor fraction of the total session time. The resulting counts describe the number of momentary looks that happen to coincide with the observer's judgment regarding presence or absence of an ongoing behavior.

A variation of time sampling often referred to as interval recording divides the session into a series of temporally contiguous observation intervals. Although it frees the observer to look continuously, it grossly restricts the observation product to checking the presence or absence of behaviors occurring within each interval. Definitional requirements related to the temporal coincidence of the target behavior with observation intervals may complicate the procedure. For example, the behavior of interest may be scored only if it occurs throughout an interval, or it may qualify for checking if it occurs during only part of an interval.

In both interval recording and time-sampling, the passage of arbitrarily determined time intervals rather than the behaver's responses occasion the observing response. Both procedures yield counts of only the number of experimenter-imposed time intervals within observation sessions when the observer's looks and the behaver's actions coincide. Thus, in addition to the problem of definitional slippage, both the continuity and the amount of ongoing behavior remain obscure. Moreover, since both may vary in unexpected directions during treatment, what may appear to be an accurate representation of a subject's behavior during pretreatment assessment cannot be assumed to predict observational adequacy throughout the course of intervention.

Efforts to improve the accuracy and reliability of the human observing response are essentially those that increase its automaticity to make it more closely approximate the precision of nonhuman transducers. A definition of behavior that includes its repeated displacement in space (i.e., its countable cycles) as well as its continuity (i.e., its temporal properties) requires continuous transduction of its response dimensions. Whether human or nonhuman observers are used and regardless of how accurately calibrated they may be, discontinuous procedures preclude continuous transduction and are, therefore, incapable of transmitting the true amount and temporal distribution of the behavior.

None of the discontinuous observation procedures produces accurate information on the number of responses emitted, their duration, their frequency, or their distribution in real time (Binder & Jameson, 1982; Bourie, 1981; Powell, 1984; Powell, Martindale, & Kulp, 1975; Powell, Martindale, Kulp, Martindale, & Bauman, 1977; Repp, Roberts, Slack, Repp, & Berkler, 1976; Springer, Brown, & Duncan, 1981). Futhermore, *agreement among independent observers using the same schedule of intermittent observation cannot rectify these shortcomings*. The constraints of discontinuous observation thus render it inappropriate for assessing the dimensional quantities of response classes.

RECORDING

Just as behaver responses must occasion observing responses, so must they control recording responses. It is the output of the observation procedure that constitutes the record of what was observed. Due to the intimacy of this relation, sources of variability in the observing response will inevitably reflect themselves in the notation of response emissions and therefore in the accuracy of the data product. Thorough observer training, design excellence, and calibration checks for observer sensitivity will fail to compensate for a notation format and an observation/recording schedule that prevent complete and accurate recording. Conversely, the best recording procedure cannot purge the resulting data of input flawed by imprecise or discontinuously programmed observation.

The nature of the recording response either facilitates or diminishes the accuracy of the resulting data. Its automaticity, like that of the observing response, is a critical factor, for the two must function in fluent coordination to complete the transduction process. Compatibility of the observing and recording components is the objective.

Again, precisely calibrated instrumentation provides the model. When attached to the transducer, the only limit on the latency of the recording response is the operating time of the associated equipment: milliseconds for deflection of an electromechanical recording pen or microseconds for electrooptical, electronic, or magnetic recording media. Given the speed of microprocessing, the upper limits of recording frequency far exceed that of the most fluent human motor responses. In addition, the temporal distribution of all behaver responses can appear as permanent tracings for both on-line and post hoc analysis. If behavior assessors could only become as hooked by current technology as the nation's youth, even obsolescent PacMan could reveal a picture of human behavior assets and liabilities beyond the fantasies of a fatigued army of timer-controlled pencil pushers.

How can the human recording response approach this ideal? Given that it should not interrupt visual observing, the least intrusive recording response should require no visual-motor component. Push buttons or squeeze switches operated merely by fluent finger-thumb opposition with minimal force and excursion provide the simplest and least expensive method of transmitting observed responses to a recording device. Tallying on a blank page or moving a bead on a string offer an approximation for behaver responses below the frequency range of the recording response itself. However, tally marks and slashes in time-ruled squares during intermittent observation are incompatible with maximally accurate observing and recording.

While these tactics may afford a summary record of behavior events, there remains the temporal dimension of recurring responses that reveals the fluctuations of ongoing behavior. Only a continuously operating recording device with a constant time base can portray the continuity of behavior. Automatic recorders are plentiful. If the human recording response is an easily repeatable, relatively nonfatiguing operant, the output can operate such recorders. For example, a small console equipped with five push buttons for each hand provides recording of ten different movement cycles. Adding a toggle switch above each push button permits recording the durations of ten different events. Fluency with such a device requires little practice. Yielding 20 channels of behavior information on a permanent record of response distributions through time, such recording allows uninterrupted observing. With computer readout, frequency distributions and summations are available for whatever measures are desired—all in a choice of the universally recognized standard, absolute measurement units. Today's prices make this a cost-effective buy.

Thus, even without automatic transduction of behaver movements, the observing-recording cycle becomes less cumbersome with assistance from automatic recording devices. Most observation tasks require a dedicated transducer. Therefore, the more automatic the observing-recording chain, the more accurate and complete will be the quantitated portrayal of behavior. Concurrent recording of antecedent and subsequent events along with behaver movement cycles results in permanent records revealing the dynamic interactive property of behavior that defines its functionality—the common focus of both behavior assessment and behavior analysis.

We have stressed both the scientific utility and the application efficiency that result from functionally defined response classes. But functional definition itself requires *functional recording* of the *relations* between behaver movements and their controlling environmental variables, both antecedent and subsequent. For this purpose, transducers that both detect and define selected responses may be arranged so that different antecedents are associated with operation of each. One or more of these antecedent-movement relations may also produce specified subsequent events, thus permitting direct measurement of functional relations between antecedents, responses, and subsequents. Such a system requires a recording channel for each transducer *with respect to each antecedent*. Thus, two antecedents and two transducers require four recording channels. Each channel records from a given transducer only when one of the two antecedents is occurring. Separate recording of the four functionally defined classes facilitates analysis of each class with respect to whatever reinforcement contingencies may be thought useful to study. With such a functional definitional/recording system, each additional transducer necessitates as many recording channels as there are antecedents associated with its operation. Recording of subsequent events occurs in association with whatever antecedent is in effect at the time each is delivered. An automated system that transduces each response emission and subsequent event to a recording mechanism that operates only when specified environmental antecedents are occurring produces a series of permanent tracings of separately recorded antecedent-response-subsequent relational occurrences. This operational definition of functional response classes can be confirmed only by the subject's continuing interaction in the assessing/analyzing environment. One such system revealed lawful commonalities in functionally defined response classes with demonstrated generality throughout the range of psychometrically defined retardation (Barrett, 1965, 1969, 1973, 1974, 1977; Barrett & Lindsley, 1962). Any definitional vagary in a recording and measurement system of this nature will be due to drift in transducer sensitivity. Frequent calibration checks minimize the problem.

EVALUATING MEASUREMENT PROCEDURES

Appraisal of behavior measurement depends on the extent to which it represents a true portrayal of behavior. All subsidiary criteria derive from the conception of behavior as the fundamental datum of assessment and analysis.

BASIC QUESTIONS

Accordingly, each of the major considerations in designing behavior measurement procedures that follow from a natural science conception of behavior also form a set of criteria for judging measurement adequacy.

- Does it represent the fundamental properties of behavior: movement, the time it takes, its place in time, its continuity through time, and its interactive (measurable) effects on the environment?
- Are its response class definitions unambiguous enough to ensure repeatable and accurate detection, sensitive enough to reveal variability, and functional with respect to controlling antecedent and subsequent events?

- Are the dimensional quantities chosen to represent behavior properties sensitive to the relevant aspects of the behavior being assessed?
- Are the units standard, absolute, and appropriately sensitive to reveal the variability associated with intervention effects as distinct from the variability associated with pretreatment conditions?
- Are its observation procedures exclusively and continually reactive to the dimensional quantities selected for analysis?
- Are its recording mechanisms operating continuously, activated solely by occurrences of selected response classes, and capable of revealing their distribution through real time?

These are questions the behavior assessor might ask of the procedures currently prevalent in the behavior assessment literature. Their answers should reveal the major impediments to replication, generality, and expansion of our knowledge base in behavior assessment and applied behavior analysis.

ACCURACY OF BEHAVIOR PORTRAYAL

The extent to which obtained measures approximate the true values of a naturally occurring phenomenon defines accuracy of measurement. An accurate measurement procedure will provide a one-to-one correspondence between movement cycles and their recorded products. Perfect correspondence (i.e., perfect accuracy) produces true values. The closest approximation to true values results from instruments that are sensitively calibrated to yield the same reading on successive occurrences of the same defined movement cycle. To portray the continuity and temporal distribution of true values requires precise timing and recording devices whose operation is continuous and whose units are appropriate to the task at hand.

If humans perform the functions of detecting and recording, their accuracy may be determined by comparing *each* of their recorded responses with *each* of the same series of responses detected and recorded automatically. For example, training experimentally naive women to detect small lumps in a simulated breast can be brought to high levels of accuracy for lumps as small as 2 mm by comparing reported detection with a continuous automatic recording from pressure-sensitive transducers in each lump (Adams *et al.*, 1976). Simply comparing summations of the responses recorded by the human against those recorded automatically will not suffice. Although summations may be identical, the responses detected by automation may not be those detected by the human observer/recorder. The same logic applies to use of two humans in place of automation. Agreement of one with the other can in no way test the accuracy of the product. This is especially so when neither person has been trained to observe and record

accurately the range of response values that may result from treatment effects.

Issues of validity arise only when independently measured true values are not available for comparison with obtained values. If transduction is direct and the recording mechanism responds in one-to-one correspondence with each transduced response through time, the product will be a faithful reproduction of response occurrences, the time they took, their locus in time, and their distribution throughout the recording period.

STABILITY OF MEASUREMENT

In order to reveal relations between behavior and its controlling variables, transduction accuracy must remain stable over time. Of course, a relentlessly accurate measurement procedure meets this criterion by definition. Demonstrating measurement stability over time requires that a succession of constant behavior inputs produce the same values from the instrument or measuring procedure. The greater the number of invariant response repetitions yielding identical recorded values, the more stringent the demonstration and the more confidence one can have regarding the reliability of both the process and its products.

A constant response input will not yield identical values if variability lurks in the measurement units chosen to quantify them. Moreover, absence of standard units precludes comparison with true values. Given demonstrated stability of other components of the measurement process, the universally recognized units denoting countability, duration, latency, frequency, celeration, and IRT remove the unreliability associated with elastic, idiosyncratic units. Accurate timers, continuous automatic recorders, and response-specific transducers make it possible to track the accuracy of a measurement system over time, thereby determining its stability or reliability.

CALIBRATION

Obtaining and maintaining accurate and stable transduction requires calibration—checking and adjusting the measuring system to achieve systematically standardized relations between response class input and recorded output. Initially, calibration procedures vary the sensitivity of the transducer until it operates regularly in perfect correspondence to only the events selected for study. Frequent periodic rechecking and adjustment of sensitivity assures that definitional drift is not occurring and that recording devices are faithfully reproducing what is detected.

Use of automated transducers eliminates the influence of observer fatigue and bias, but it does not replace humans in the calibration process. Although devices may be preferred to detect repeated instances of the same response, frequent observation of their operation is essential for the desired outcome. Astute human observers engaged in periodic concurrent monitoring often supply added sensitivity to the many events that may influence variations in what a transducer is detecting. Human detection of such subtleties could contribute substantially to design refinement that sensitizes transducers to nuances deemed significant.

Just as the operating characteristics of automatic transducers determine their suitability for specific applications, so the personal histories of human observers may play a significant role in their effectiveness as detectors and recorders of specified response classes. Vested interests may create biases that remain unknown unless independently obtained accurate values are available for checking the observing-recording product. Screening and carefully selecting observers is an obvious first step toward achieving the best approximation of accurate recording.

Training observers to perform the observing and recording responses to the highest level of accuracy is akin to sensitizing an automatic transducer to only the response inputs selected for assessment. Stimulus control technology is just as applicable for this purpose as it is for teaching new skills to handicapped people. Shaping is often necessary to reach stable topography of the detecting-transmitting response. Amounts of selected dimensional quantities of the defined response classes must then become discriminative stimuli for the observing response. Once obtained, stimulus control by the appropriate dimensional quantities must retain stability throughout exposure to a wide range of values over repeated observation periods.

Measurement practice sessions with scripted performance by bogus subjects provides known input in the form of behavior episodes with predetermined values of dimensional quantities. Otherwise, videotaping or automatic transduction are useful to provide true values for training criteria. As a precaution against drift, it is important that scripted or otherwise arranged known inputs to observer trainees contain the full range of values that might occur during the course of real treatment.

Once the formal study is underway, frequent periodic checking of observer accuracy against automatically transduced recordings or against products of the subject's responding (calibration) will reveal whether retraining sessions are necessary to adjust selective reactivity to the dimensional quantities under study. Data reporting these calibration checks should replace determinations of interobserver agreement in routine descriptions of behavior assessment methods employing human observers.

Given the advances in detecting and transmitting technology, to say

nothing of the recording and analytic capabilities of microcomputers, "the use of unassisted human observers can only be seen as the Stone Age of behavioral observation" (Johnston & Pennypacker, 1980, p. 126).

IMPLICATIONS FOR BEHAVIOR ASSESSMENT

Today's behavioral assessment literature offers a smorgasbord of techniques sufficient to suit nearly anyone's taste. Interpretive orientation rather than methodology distinguishes "behavioral" from other approaches. The result is a verbal veneer superimposed on a set of practices derived from developmental field studies, psychodynamics, and psychometrics, tempered by expedient demands for practicality and pressure for social acceptance. The relation of "behavioral" assessment methodology to any distinct concept of behavior is tenuous at best, and often difficult if not impossible to detect. More specifically, the measurement aspects of "behavioral" assessment appear to reflect something other than behavior as defined in the natural science tradition.

TOWARD CONCEPTUAL-METHODOLOGIC CONSISTENCY

Accepting the label "behavioral" while overlooking both a natural science definition of behavior and its associated measurement procedures leaves behavioral assessment as a field with an identity that is far from obvious—perhaps even obscure. Creating a theoretical superstructure (Lang, 1977) will not heal the very apparent conceptual-methodologic schism. Nor will it create order in the absence of substantiating data. Such an undertaking requires generality of findings for a secure foundation, and generality has not yet emerged from behavioral assessment research. The state of "anarchy" (Cone & Hawkins, 1977) resulting from nonfunctional definitions cannot help but perpetuate itself in the absence of a unifying concept of behavior and methodology that addresses and explicates the properties of that naturally occurring phenomenon.

Reviewing behavior properties and their dimensions should generate instances and noninstances of each. With these categories as guides, a classification of currently measured dependent variables could serve as a starting point for delineating the behavioral aspects of contemporary "behavioral" assessment procedures. Those consistent with a natural science definition, its properties, and dimensions would be the objective of such a survey. Movement cycles and measurement of their effects on the environment through time are necessary elements. Procedures that do not meet the criteria of definitional consistency are not "of or relating to behavior." To retain them under the rubric "behavioral assessment" will only sustain the current level of conceptual-methodologic dissonance. Attendant inconsistencies will inevitably plague the search for generality and thus retard progress toward functional classification and prognostic effectiveness.

IMPROVED SPECIFICATION OF MOVEMENT CYCLES

Having sorted out those measurement practices that are compatible with a natural science concept of behavior, it should become obvious that measuring movements in terms of their effects on the environment invites the use of sensing devices other than human observers. Indeed, such devices now enable physiologic events to be included in the domain of observable behavioral events. Bioelectrical activity and the technology of transduction produce continuous recordings of muscle activity cycles [electromyogram (EMG) and electrocardiogram (EKG)], cycles of electrical activity recorded from the scalp [electroencephalogram (EEG)] and from eye movements [electrooculogram (EOG)]. Heart rate, blood and pulse volume, and respiration are other internal changes now transduced into environmental effects which make them accessible to measurement (Nietzel & Bernstein, 1981). Advances in telemetry now make it possible to obtain measurements of various internal behaviors in natural environments (Rugh & Schwitzgebel, 1977).

By contrast, transducers of observable movement cycles, though more easily obtainable and more straightforward in design, appear infrequently in the literature of behavioral assessment and analysis. A review and classification of some existing devices is available (Schwitzgebel, 1976). Examples include a slouch-transducing device (Arzin, Rubin, O'Brien, Ayllon, & Roll, 1968), a smoking controller (Azrin & Powell, 1968), a urinary accident transducer (Azrin & Foxx, 1971), and a detector of appropriate toileting responses (Azrin, Bugle, & O'Brien, 1971). Other transducers in the literature include those for eye movements (Doran & Holland, 1971), workshop tool usage (Schroeder, 1972), anal sphincter pressure (Kohlenberg, 1973), penile circumference (Laws & Rubin, 1969), blood alcohol concentration (Sobell & Sobell, 1975), multiple tics (Barrett, 1962), duration of speech (Lane, 1964), speech pauses (Ruder, Jensen, & Brandt, 1970), detection of lumps in simulated human breasts (Adams *et al.*, 1976), and private events (Hefferline & Bruno, 1971). The last work began as investigations of electromyographic changes that were demonstrated to be conditionable without "awareness" in the frontier era of human operant conditioning (Hefferline, Keenan, & Harford, 1959). As more sophisticated methods evolved, Hefferline (1962) foresaw the advent of biofeedback and the limitless possibilities

for productive interaction between psychophysiologic, clinical-psychological, and behavioristic endeavors:

> Technological advance has been greatly spurred by the crash program directed at telemetering a man in space. Improved transducers, amplifiers, and recorders, developed for use in satellites, are already available commercially at reasonable prices. Obviously such devices can be used efficiently with earth-bound . . . subjects. (p. 125)
>
> Modern instrumentation . . . provides automatic programming of experimental operations too fast or too slow for manual execution and processes voluminous data with superhuman speed and accuracy. With artificial eyes, ears and hands, recording impartially and without lapse of attention, the psychologist, as he [sic] deals with a complex situation, gets a fuller and more trustworthy answer to the question of what is going on. (p. 100)
>
> With such instrumentation the therapist interested in research can have a consulting room which is at the same time a laboratory. Or any recordings obtained as an adjuvant to therapy can supply data for someone else's scientific use. The feasibility of arrangements of this sort is what leads me to speak of symbiotic relations which could obtain between experimental and clinical psychology. (p. 101)

PARTICIPATION IN A UNIVERSAL MEASUREMENT LANGUAGE

Instrumentation also provides benefits beyond definitional specificity and accuracy. Its recorded output translates continuous behavior-environment interactions into universally recognized idemnotic units of measurement. In so doing, instrumentation affords behavior assessment and analysis a language common not only to its parent science but to other relevant disciplines as well. The need has long been expressed and the impediments recognized (Eiduson, Geller, Yuwiler, & Eiduson, 1964):

> The two central problems in relating biochemistry to behavior are those of quantification and correspondence. Biochemical data are nearly always parametric, consisting of real numbers, real intervals, and a real zero point. In addition, the variable can be expressed in explicit, operationally defined units. Many extremely important behavioral variables, however, are nonparametric, with none of these properties. Indeed some are essentially generic names for a whole class of loosely related behavioral complexes and, as commonly used, are both unitless and dimensionless. . . . Only relatively few . . . have true parametric properties. . . . Response rate is as parametric a measure as any biochemical one and is ideal for crossdisciplinary studies. (pp. 468–469)
>
> But what of the neurochemist whose discipline presently deals not with the formal composition of the final painting but with the discrete substances that comprise its pigments? . . . His [sic] ultimate goal, of course, is to achieve a point-for-point (or sequence-for-point) correspondence with behavior. . . . It

> will be necessary not only to quantify isolated segments snipped off a behavioral complex but also to dissect the entire structure into quantifiable units appropriate to biochemical studies. Not only must the behavioral complex be more rigorously analyzed, but the analysis must be carried out with an eye to the problem of correspondence. Essentially this becomes a search for comparable dimensions and units. (p. 471)

Preparatory to such felicitous cross-science interface, behavior assessment must engage in vigorous efforts at in-house communication. Reproducibility and generality of its findings, the test of its viability, will depend on its awareness and integration of the fundamental functions of measurement: description, comparison, and prediction.

To fulfill any communicative function, quantitative *description* must attach standard units to its obtained values. Idiosyncratic units derived from in-house scales of convenience (e.g., "levels of independence" or "levels of assertiveness") fail to provide standard distinguishing descriptors for their attached numerical values. For this reason, they are useless as quantitative descriptors with any generality beyond the special situations for which they are contrived.

Reproducibility and generality require *comparison* of obtained values both from study to study and longitudinally from condition to condition within intrasubject replications. For this purpose, numbers with vacillating values and vaganotically conceived units are of no use. To find the functional relations between behavior and the environment requires absolute units bearing an invariant relation to the dimension they describe. Without this basic requisite, laws of arithmetic and higher mathematics are inapplicable, generality will remain obscured, and interscience investigations will not be possible.

Prediction requires repeated measurements over time. Unstable units and quantification that neglects the temporal dimensions of individual behavior simply do not provide the information necessary to predict the outcome of measurement of an individual's behavior at a future time. It is well to remember that vaganotically defined phenomena, quantitated by their own empirically determined variability as measured in groups of people, are not the subject matter of behavior assessment or analysis. Similarly, vaganotically conceived units, demonstrating interpretive elasticity and emanating from assumed variability, are not suitable for quantitative description of behavior as a naturally occurring phenomenon (see Cone, 1981). Their origin derives from social science statistical assumptions; their product applies neither to the description or prediction of an individual's behavior over time. For this reason, their application is appropriate only to phenomena that are conceptually consistent with that approach.

Germination and cross-fertilization of experimental and clinical endeavors characterizes the model so successfully applied in the medical sciences for over a century (Bernard, 1865/1957). It cannot be approached without nurture from a medium of discourse that retains its communication function regardless of the purpose for which it is used or the philosophical leanings of its users. Terms of a common language must, therefore, be universal and invariant. Their denotations must be standard across applications. In the reciprocal nurturing process, the vernacular of scientific undertakings and those of clinical exploration and demonstration should be equally precise if for no other reason than to facilitate observation of whatever lawfulness may underlie the phenomenon being studied. Both are concerned with lawfulness or predictability. Neither is likely to advance either the basic science or its technological development if the ingredients of the communication system contain more variability than the behaviors they both address.

To demonstrate generality within its domain as well as productive communication and eventual collaborative investigation with other relevant sciences, behavior assessment and analysis must, at the very least, measure the behaviors it studies in a common language (see Churchman, 1959). Standardization at this level should precede and hopefully would presage standardization at other levels of assessment methodology (Goldfried, 1976, 1979; Kanfer, 1972).

ASSESSING THE AIMS, TECHNIQUES, AND RESULTS OF TREATMENT

It is becoming increasingly popular to engage in the trappings of "behavioral" treatment in order to please—or appease—third parties. Indeed, a recommended procedure for setting treatment goals, choosing behaviors to be altered, selecting procedures for altering them, and determining treatment effectiveness is to submit these decisions for approval by parents, friends, peers, and so on (Kazdin, 1977; Wolf, 1978). As a result, the rating scale is staging a dramatic comeback (Barrett, 1981). Consumers are asked to rate their satisfaction with or approval of all aspects of treatment. If its current rate of growth continues, the rating scale with its unitless values and synthetic dimensions promises to supplant other assessment methods—they will simply be unnecessary accoutrements. Therapeutic decisions and effects will depend on social value judgments. Units of behavior analysis will become inseparable from the linguistic convenience and conditioning histories of rating scale respondents.

If given the chance, standard measurement procedures can halt this regression to prebehavioristic methods. We do not have to revert to the vagaries of societal language to determine what behaviors to change or how much change they need in order to increase the competence of our clients. The *behavior* of skilled classmates, peers, and friends measured as functional response classes can provide unbiased answers to these questions. Quantified performance ranges obtained by measuring the behaviors of competent individuals provide standards for assessing a client's deviations and for determining the aims of intervention (Barrett, 1977; Haughton, 1972; Kunzelmann, 1970; Van Houten, 1979; Walker & Hops, 1976; Willis, 1974). Tailored for each individual's deviant or deficient functional response classes, procedures that produce increasing approximations to these performance standards are, then, the procedures of choice. Determining the response class functions to be altered and selecting the most efficient procedures for doing so require measured behavior for their most effective solution. The measured behavior of choice is the client's. The more standardized the language of measurement we provide, the more precise and enlightening will be the client's answers.

Experimental behavior analysis brought the most sensitive of all measurement systems to the task of quantifying what a behaver does under different environmental conditions. It is an integral component of the operant "preparation" that will facilitate productive collaboration with other disciplines seeking to understand human behavior (Skinner, 1986). From its fruits emerged a methodology that revised the way professionals and others regarded themselves and their troubled brethren. It also revised the way professionals write about what they do. But printed words do not suffice for the progress of either the science of human behavior or its application in improved technology of treatment. Assessment of human behavior must elucidate the written word with measurement—measurement that allowed small animals to produce their own precise recordings—measurement that will permit clients to reveal to investigators the most promising and effective treatments.

ACKNOWLEDGMENT

The authors are grateful to Wayne Robb and Carl Binder for their critical reading and many suggestions.

REFERENCES

Adams, C. K., Goldstein, M. K., Hench, L., Hall, D. C., Madden, M., Pennypacker, H. S., Stein, G. H., & Catania, A. C. (1976). Lump detection in a simulated human breast. *Perception and Psychophysics, 20*(3), 163–167.

Angle, H. V., Hay, L. R., Hay, W. M., & Ellinwood, E. H. (1977). Computer assisted behavioral assessment. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 369–380). New York: Brunner/Mazel.

Azrin, N. H., Bugle, C., & O'Brien, F. (1971). Behavioral engineering: Two apparatuses for

toilet training the institutionalized retarded. *Journal of Applied Behavior Analysis, 4*, 249–253.

Azrin, N. H., & Foxx, R. M. (1971). A rapid method of toilet training the institutionalized retarded. *Journal of Applied Behavior Analysis, 4*, 89–99.

Azrin, N. H., Holz, W., Ulrich, R., & Goldiamond, I. (1961). The control of the content of conversation through reinforcement. *Journal of the Experimental Analysis of Behavior, 4*, 25–30.

Azrin, N. H., & Lindsley, O. R. (1956). The reinforcement of cooperation between children. *Journal of Abnormal and Social Psychology, 52*, 100–102.

Azrin, N. H., & Powell, J. (1968). Behavioral engineering: The reduction of smoking behavior by a conditioning apparatus and procedure. *Journal of Applied Behavior Analysis, 1*, 193–200.

Azrin, N. H., Rubin, H., O'Brien, F., Ayllon, T., & Roll, D. (1968). Behavioral engineering: Postural control by a portable operant apparatus. *Journal of Applied Behavioral Analysis, 1*, 99–108.

Baer, D. M. (1982, May). Some recommendations for a modest reduction in the rate of current recommendations for an immodest increase in the rate of exclusive usages of rate as a dependent measure. Unpublished paper in the symposium, "Is rate of response a universal datum?" Presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing similarity to a model. *Journal of the Experimental Analysis of Behavior, 10*, 405–416.

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.

Barrett, B. H. (1958). The role of insight. In J. M. Hadley (Ed.), *Clinical and counseling psychology* (pp. 62–120). New York: Knopf.

Barrett, B. H. (1962). Reduction in rate of multiple tics by free operant conditioning methods. *Journal of Nervous and Mental Disease, 135*, 187–195.

Barrett, B. H. (1965). Acquisition of operant differentiation and discrimination in institutionalized retarded children. *American Journal of Orthopsychiatry, 35*, 862–885.

Barrett, B. H. (1969). Behavioral individuality in four cultural-familially retarded brothers. *Behaviour Research and Therapy, 7*, 79–81.

Barrett, B. H. (1973). *Annual Report*: 1 July, 1972–30 June 1973. Belmont, MA: W. E. Fernald State School of Behavior Prosthesis Laboratory.

Barrett, B. H. (1974). *Annual Report*: 1 July, 1973–30 June, 1974. Belmont, MA: W. E. Fernald State School Behavior Prosthesis Laboratory.

Barrett, B. H. (1975). Course of acquisition: Accuracy and frequency comparisons. In *Instructional Development and Behavior Analysis Program: Annual Report, 1 July 1974–30 June 1975* (pp. 69–74). Belmont, MA: Walter E. Fernald State School, Behavior Prosthesis Laboratory.

Barrett, B. H. (1977). Behavior analysis. In J. Wortis (Ed.), *Mental retardation and developmental disabilities* (Vol. 9, pp. 141–202). New York: Brunner/Mazel.

Barrett, B. H. (1979). Communitization and the measured message of normal behavior. In R. L. York & E. Edgar (Eds.), *Teaching the severely handicapped* (Vol. 4, pp. 301–318). Columbus, OH: Special Press.

Barrett, B. H. (1980, June). From accuracy to fluency with a standard measure. Unpublished paper presented at the meeting of the Association for Behavior Analysis, Dearborn, MI.

Barrett, B. H. (1981, May). Measurement issues. Unpublished paper in the symposium "Progress in Precision Teaching." Presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

Barrett, B. H., & Lindsley, O. R. (1962). Deficits in acquisition of operant discrimination and differentiation shown by institutionalized retarded children. *American Journal of Mental Deficiency, 67*, 424–436.

Bernard, C. (1957). *An introduction to the study of experimental medicine*. New York: Dover. (Original work published 1865).

Binder, C. V., & Jameson, D. (1982). An analysis of interval size in a momentary time-sampling procedure. *Journal of Precision Teaching, 3*, 9–15.

Birnbrauer, J. S. (1979). Applied behavior analysis, service and the acquisition of knowledge. *The Behavior Analyst, 2*(1), 15–21.

Birnbrauer, J. S. (1981). External validity and experimental investigation of individual behaviour *Analysis and Intervention in Developmental Disabilities, 1*, 117–132.

Birnbrauer, J. S., Burchard, J. D., & Burchard, S. N. (1970). Wanted: Behavior analysts. In R. H. Bradfield (Ed.), *Behavior modification: The human effort* (pp. 19–76). San Rafael, CA: Dimensions Publishing Company.

Bourie, C. J. (1981, May). Some dimensions of sampling behavior. Unpublished paper presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

Churchman, C. W. (1959). Why measure? In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 83–94). New York: John Wiley & Sons.

Cohen, D. J. (1962). Justin and his peers: An experimental analysis of a child's social world. *Child Development, 33*, 697–717.

Cohen, D. J., & Lindsley, O. R. (1964). Catalysis of controlled leadership in cooperation by human stimulation. *Journal of Child Psychology and Psychiatry, 5*, 119–137.

Cone, J. D. (1977). The relevance of reliability and validity for behavioral assessment. *Behavior Therapy, 8*, 411–426.

Cone, J. D. (1981). Psychometric considerations. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (2nd ed., pp. 38–68). New York: Pergamon.

Cone, J. D., & Hawkins, R. P. (Eds.), (1977). *Behavior assessment: New directions in clinical psychology*. New York: Brunner/Mazel.

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

Doran, J., & Holland, J. G. (1971). Eye movements as a function of response contingencies measured by blackout technique. *Journal of Applied Behavior Analysis, 4*, 11–17.

Eiduson, S., Geller, E., Yuwiler, A., & Eiduson, B. T. (1964). *Biochemistry and behavior*. New York: Van Nostrand.

Ferster, C. B., & DeMyer, M. K. (1961). The development of performances in autistic children in an automatically controlled environment. *Journal of Chronic Diseases, 13*, 312–345.

Ferster, C. B., & DeMyer, M. K. (1962). A method for the experimental analysis of the behavior of autistic children. *American Journal of Orthopsychiatry, 32*, 89–98.

Flanagan, B., Goldiamond, I., & Azrin, N. H. (1958). Operant stuttering: The control of stuttering behavior through response-contingent consequences. *Journal of the Experimental Analysis of Behavior, 1*, 173–178.

Fraley, L. E. (1981). On the spread of behavior analysis to the applied fields. *The Behavior Analyst, 4*, 33–41.

Goldfried, M. R. (1976). Behavioral assessment. In I. B. Weiner (Ed.), *Clinical methods in psychology* (pp. 281–330). New York: John Wiley & Sons.

Goldfried, M. R. (1977). Behavioral assessment in perspective. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 3–22). New York: Brunner/Mazel.

Goldfried, M. R. (1979). Behavioral assessment: Where do we go from here? *Behavioral Assessment, 1*, 19–22.

Goldfried, M. R., & Kent, R. N. (1972). Traditional versus behavioral personality assessment: A comparison of methodological and theoretical assumptions. *Psychological Bulletin, 77*, 409–420.

Greene, E. B. (1941). *Measurements of human behavior*. New York: Odyssey Press.

Hake, D., & Vukelich, R. (1973). Analysis of the control exerted by a complex cooperative procedure. *Journal of the Experimental Analysis of Behavior, 19*, 3–16.

Hake, D., Vukelich, R., & Kaplan, S. J. (1973). Audit responses: Responses maintained by access to existing self or coactor scores during non-social, parallel work, and cooperation procedures. *Journal of the Experimental Analysis of Behavior, 19*, 409–423.

Hake, D., Vukelich, R., & Olvera, D. (1975). The measurement of sharing and cooperation as equity effects and some relationships between them. *Journal of the Experimental Analysis of Behavior, 23*, 63–79.

Hall, D., Goldstein, M. K., & Stein, G. H. (1977). Progress in manual breast examination. *Cancer, 40*, 364–370.

Haughton, E. C. (1972). Aims—Growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing* (pp. 20–39). Seattle, WA: Special Child Publications.

Hayes, S. C., & Cavior, N. (1980). Multiple tracking and the reactivity of self-monitoring: II. Positive behaviors. *Behavioral Assessment, 2*, 283–296.

Haynes, S. N., & Jensen, B. J. (1979). The interview as a behavioral assessment instrument. *Behavioral Assessment, 1*, 97–106.

Hefferline, R. F. (1962). Learning theory and clinical psychology—An eventual symbiosis? In A. J. Bachrach (Ed.), *Experimental foundations of clincial psychology* (pp. 97–138). New York: Basic Books.

Hefferline, R. F., & Bruno, L. J. (1971). The psychophysiology of private events. In A. Jacobs & L. B. Sachs (Eds.), *The psychology of private events* (pp. 164–192). New York: Academic Press.

Hefferline, R. F., Keenan, B., & Harford, R. A. (1959). Escape and avoidance conditioning in human subjects without their observation of the response. *Science, 130*, 1338–1339.

Hersen, M., & Bellack, A. S. (Eds.). (1976). *Behavioral assessment: A practical handbook*. Elmsford, NY: Pergamon Press.

Johnston, J. M. (1982, May). Behavior technology in the year 2000: What do we want it to look like? Invited address presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum.

Kanfer, F. H. (1972). Assessment for behavior modification. *Journal of Personality Assessment, 36*, 418–423.

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

Kohlenberg, R. J. (1973). Operant conditioning of human anal sphincter pressure. *Journal of Applied Behavior Analysis, 6*, 201–208.

Kunzelmann, H. P. (Ed.). (1970). *Precision teaching*. Seattle, WA: Special Child Publications.

Lane, H. (1964). Differential reinforcement of vocal duration. *Journal of the Experimental Analysis of Behavior, 7*, 107–115.

Lang, P. J. (1977). Physiological assessment of anxiety and fear. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 178–195). New York: Brunner/Mazel.

Laws, D. R., & Rubin, H. B. (1969). Instructional control of an autonomic sexual response. *Journal of Applied Behavior Analysis, 2*, 93–99.

Lindsley, O. R. (1956). Operant conditioning methods applied to research in chronic schizophrenia. *Psychiatric Research Reports, 5*, 118–139.

Lindsley, O. R. (1960). Characteristics of the behavior of chronic psychotics as revealed by free-operant conditioning methods. [Monograph]. *Diseases of the Nervous System, 21*, 66–78.

Lindsley, O. R. (1962a). Operant conditioning techniques in the measurement of psychopharmacologic response. In J. H. Nodine & J. H. Moyer (Eds.), *Psychosomatic medicine: The first Hahnemann symposium* (pp. 373–383). Philadelphia: Lea & Febiger.

Lindsley, O. R. (1962b). A behavioral measure of television viewing. *Journal of Advertising Research, 2*(3), 2–12.

Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education, 147*, 62–81.

Lindsley, O. R. (1966). Experimental analysis of cooperation and competition. In T. Verhave (Ed.), *The experimental analysis of behavior* (pp. 470–501). New York: Appleton-Century-Crofts.

Lindsley, O. R. (1969a). Personal communication.

Lindsley, O. R. (1969b). Direct behavioral analysis of psychotherapy sessions by conjugately

programmed closed circuit television. *Psychotherapy: Theory, Research and Practice, 6*, 71–81.
Lindsley, O. R., Hobika, J. H., & Etsten, B. E. (1961). Operant behavior during anesthesia recovery: A continuous and objective method. *Anesthesiology, 22*, 937–946.
Linehan, M. M. (1977). Issues in behavioral interviewing. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 30–51). New York: Brunner/Mazel.
Linehan, M. M. (1980). Content validity: Its relevance to behavioral assessment. *Behavioral Assessment, 2*, 147–159.
Lockhart, K. A. (1979). Behavioral assessment of human preference. *The Behavior Analyst, 2*(2), 20–29.
MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*, 95–107.
Mahoney, M. J. (1977). Some applied issues in self-monitoring. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 241–254). New York: Brunner/Mazel.
Mash, E. J. (1979). What is behavioral assessment? *Behavioral Assessment, 1*, 23–29.
Mash, E. J., & Terdal, L. G. (1981a). Behavioral assessment of childhood disturbance. In E. J. Mash & L. G. Terdal (Eds.), *Behavioral assessment of childhood disorders* (pp. 3–76). New York: Guilford Press.
Mash, E. J., & Terdal, L. G. (Eds.). (1981b). *Behavioral assessment of childhood disorders*. New York: Guilford Press.
Michael, J. (1980). Flight from behavior analysis. *The Behavior Analyst, 3*(2), 1–22.
Michael, J. (1985). Fundamental research and behaviour modification. In C. F. Lowe, M. Richelle, D. E. Blackman, & C. M. Bradshaw (Eds.), *Behaviour analysis and contemporary psychology* (pp. 159–164). London: Lawrence Erlbaum.
Morgan, B. J., & Lindsley, O. R. (1966). Operant preferences for stereophonic over monophonic music. *Journal of Music Therapy, 3*, 135–143.
Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace & World.
Nathan, P. E., Schneller, P., & Lindsley, O. R. (1964). Direct measurement of communication during psychiatric admission interviews. *Behaviour Research and Therapy, 2*, 49–57.
Nelson, R. O. (1977a). Methodological issues in assessment via self-monitoring. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 217–240). New York: Brunner/Mazel.
Nelson, R. O. (1977b). Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 5, pp. 263–308). New York: Academic Press.
Nelson, R. O. (1980). The use of intelligence tests within behavioral assessment. *Behavioral Assessment, 2*, 417–423.
Nelson, R. O., & Hayes, S. C. (1979). Some current dimensions of behavioral assessment. *Behavioral Assessment, 1*, 1–16.
Nietzel, N. T., & Bernstein, D. A. (1981). Assessment of fear and anxiety. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (2nd ed., pp. 215–245). New York: Pergamon Press.
O'Leary, K. D. (1979). Behavioral assessment. *Behavioral Assessment, 1*, 31–36.
Pennypacker, H. S. (1981). On behavior analysis. *The Behavior Analyst, 4*, 159–161.
Pierce, W. D., & Epling, W. F. (1980). What happened to analysis in applied behavior analysis? *The Behavior Analyst, 3*(1), 1–9.
Powell, J. (1984). On the misrepresentation of behavioral realities by a widely practiced direct observation procedure: Partial interval (one-zero) sampling. *Behavioral Assessment, 6*, 209–219.
Powell, J., Martindale, A., & Kulp, S. (1975). An evaluation of time-sample measures of behavior. *Journal of Applied Behavior Analysis, 8*, 463–469.
Powell, J., Martindale, B., Kulp, S., Martindale, A., & Bauman, R. (1977). Taking a closer look: Time sampling and measurement error. *Journal of Applied Behavior Analysis, 10*, 325–332.

Repp, A. C., Roberts, D. M., Slack, D. J., Repp, C. F., & Berkler, M. S. (1976). A comparison of frequency, interval, and time-sampling methods of data collection. *Journal of Applied Behavior Analysis, 9*, 501–508.
Ruder, K. F., Jensen, P. J., & Brandt, J. F. (1970). An apparatus and procedure for the perceptual study of speech pauses. *Journal of the Experimental Analysis of Behavior, 14*, 287–290.
Rugh, J. D., & Schwitzgebel, R. L. (1977). Instrumentation for behavior assessment. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 79–113). New York: John Wiley & Sons.
Schroeder, S. R., & Holland, J. G. (1969). Reinforcement of eye movements with concurrent *Applied Behavior Analysis, 5*, 523–525.
Schroeder, S. R., & Holland J. G. (1969). Reinforcement of eye movements with concurrent schedules. *Journal of the Experimental Analysis of Behavior, 12*, 897–903.
Schwitzgebel, R. L. (1976). Behavioral technology. In H. Leitenberg (Ed.), *Handbook of behavior modification and behavior therapy* (pp. 604–626). Englewood Cliffs, NJ: Prentice-Hall.
Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
Sidman, M., & Stoddard, L. (1966). Programming perception and learning for retarded children. In N. R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 2, pp. 151–208). New York: Academic Press.
Sidman, M., & Stoddard, L. T. (1967). The effectiveness of fading in programming a simultaneous form discrimination for retarded children. *Journal of the Experimental Analysis of Behavior, 10*, 3–15.
Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *Journal of General Psychology, 12*, 40–65.
Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
Skinner, B. F. (1986). Some thoughts about the future. *Journal of the Experimental Analysis of Behavior, 45*, 229–235.
Sobell, M. B., & Sobell, L. C. (1975). A brief technical report on the MOBAT: An inexpensive portable test for determining blood alcohol concentration. *Journal of Applied Behavior Analysis, 8*, 117–120.
Springer, B., Brown, T., & Duncan, P. K. (1981). Current measurement in applied behavior analysis. *The Behavior Analyst, 4*, 19–31.
Swan, G. E., & MacDonald, M. L. (1978). Behavior therapy in practice: A national survey of behavior therapists. *Behavior Therapy, 9*, 799–807.
Van Houten, R. (1979). Social validation: The evolution of standards of competency for target behaviors. *Journal of Applied Behavior Analysis, 12*, 581–591.
Wahler, R. G. (1975). Some structural aspects of deviant child behavior. *Journal of Applied Behavior Analysis, 8*, 27–42.
Walker, H. M., & Hops, H. (1976). Use of normative peer data as a standard for evaluating classroom treatment effects. *Journal of Applied Behavior Analysis, 9*, 159–168.
Walls, R. T., Werner, T. J., Bacon, A., & Zane, T. (1977). Behavior checklists. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 77–146). New York: Brunner/Mazel.
Watson, J. B. (1924). *Behaviorism*. Chicago: University of Chicago Press.
Webster's third new international dictionary. Springfield, MA: Merriam, 1961.
Weiner, H. (1964). Conditioning history and human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 7*, 383–385.
Willis, B. (1974, July). *Project report IV of State of Washington's child service demonstration programs for precise educational remediation for managers of specific learning disabilities programs*. Tacoma, WA: Intermediate School District No. 111.
Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203–214.
Woods, T. S. (1980). On the alleged incompatibility of analysis and application: A response to Pierce and Epling. *The Behavior Analyst, 3*(2), 67–69.