CURRENT MEASUREMENT IN APPLIED BEHAVIOR ANALYSIS

Bonnie Springer, Tom Brown, and Philip K. Duncan
Drake University

ABSTRACT

The analysis of behavior began with a form of data, rate of responding, which allowed for efficient study and for the description of the basic principles of behavior. Especially important were the facts that rate of responding was a direct reflection of fundamental properties of behavior, and that rate of responding was measured continuously within an experimental session. As behavior analysts moved from purely experimental to applied settings, discontinuous, time-based methods of measurement evolved, which neither directly reflect fundamental properties of behavior nor continuously record behavior within an experimental session. This paper offers a critical discussion of current measurement practices, and discusses factors possibly related to the use of discontinuous, time-based observing/recording procedures. A theoretical basis for observing/recording procedures is presented which emphasizes continuous measurement of response dimensions directly related to fundamental properties of behavior.

The datum on which a large portion of behavior analysis is based is rate of responding (Skinner, 1938). To be sure, this measure was selected in part for its convenience, but the more important consideration was that rate of responding provided for the description of functional relations between the behavior of the organism and its environment. As Skinner pointed out: "With a systematic formulation of behavior, it is usually possible to know in advance what aspect of behavior is going to vary during a given process and what must, therefore, be measured." (1938, p. 58). Thus, both the aspects of behavior which are chosen for study, and the way those aspects are measured, determine the value of a particular measurement system.

Johnston and Pennypacker (1980) have described fundamental properties that compose behavior. The following discussion of behavioral properties, response dimensions, and units of measurement is based heavily on their description. Behavior is a naturally-occurring phenomenon, made up of fundamental properties, just as other natural phenomenon are. These properties, tem-

poral locus, temporal extent, and repeatability, can be used as the basis for measuring aspects of a particular response. Temporal locus refers to the point in time where a response occurs, and often involves a measure of latency. Temporal extent refers to the elapsed time of a response, and involved a measure of duration. Repeatability refers to the recurring nature of a response, and often involves a response count. These three, latency, duration, and countability, are among several response dimensions or aspects which derive from the fundamental properties of behavior.

In order to measure a specific response dimension, units of measurement sensitive to that dimension must be identified. Since both latency and duration are temporal aspects of a response, time units (sec, min, etc.) are appropriate units of measurement. Since countability involves the repetitive nature of a response, the appropriate unit of measurement is the cycle.

Response dimensions can be combined to form compound units. One of the most common is the one Skinner (1938) began with, rate, or cycles/unit time. Clearly, Skinner and the field of behavior analysis began with a measurement strategy which was based on measures sensitive to the fundamental properties of behavior. It

Requests for reprints should be addressed to Phillip K. Duncan, Department of Psychology, Drake University, Des Moines, IA, 50311.

will be the thesis of this paper that it is in the best interest of applied behavior analysis to continue the strategy of measurement begun by Skinner by implementing measurement tactics based on the analysis offered by Johnston and Penypacker (1980).

In developing this theme, this paper will examine measurement systems currently used in applied behavior analysis. Emphasis will be on the accuracy of data produced by current measurement systems, and the relation of those data and systems to the seminal strategy used by Skinner and advocated by Johnston and Penypacker. A conceptual framework for determining what aspects of behavior should be measured and proposals for alternatives to current measurement practices are presented.

CURRENT PRACTICES

Discontinuous, Time-based Measurement

In a recent survey, Kelly (1977) reported that in studies published by the *Journal of Applied Behavior Analysis*, 41% used some form of discontinuous time-based observation/recording procedure.¹ These procedures included time sampling, partial interval, and whole interval methods.

The time sampling method requires an observer to note whether or not a specific behavior is occurring at the end of a fixed interval of time (Hall, 1971; Kazdin, 1975). For example, a 60-minute observing/recording session might be divided into 10-second intervals. At the end of each interval, an observer looks at the subject and records whether or not the behavior is occurring. Repeated occurrences of the behavior within each interval are not recorded by the observer. Time

sampling was used in 21% of the articles reviewed by Kelly (1977).

The interval recording procedure allows the human observer to note whether or not a specific behavior occurred within an interval (Hall, 1971; Kazdin, 1975). Partial and whole interval procedures are commonly used in the applied literature. The partial interval observing/recording procedure requires an observer to record a behavior as occurring regardless of the response duration or frequency. For example, an observer watches the subject as in the previous example, and at the end of each interval records whether the behavior did or did not occur at any time during the interval.

The whole interval observation and recording procedure requires that a response be scored only if it continues throughout the interval. An observer watches the subject and at the end of each interval records whether the behavior has occurred for the entire length of the interval, has not occurred at all, or has occurred only partially through the interval. Variations on the interval method have added time intervals for recording. For example, an observer might watch a subject for 10 seconds, as in previous examples. At the end of the interval, the observer has 5 seconds to record the data. During the 5 second recording period, the behavior of the subject is not measured. The interval observation and recording procedures were used by 20% of the applied researchers reported by Kelly (1977).

These procedures are discontinuous in that within an experimental session the actual recording of responses does not necessarily occur upon the occasion of each response, even though observation of the response may or may not occur and simply be ignored as far as permanent records are concerned—a discontinuity in recording. These procedures are time-based in that records are not occasioned solely by the occurrence of a response, but by the passage of time. After time has elapsed, observation and recording oc-

¹Kelley's review showed the following measurement procedures and their relative use: time sampling recording 21%, interval recording 20%, event recording 29%, trial scoring 35%, response duration 9%, other 6%, and undetermined 2%. The percentages add up to more than 100% because more than one method could have been used in a study.

curs, depending on the particular procedure employed. Thus, the occurrence of a response not only fails to be the exclusive occasion for recording, it may, in fact, not be the occasion at all for observing and recording.

Data produced by time sampling, partial interval, and whole interval observation and recording procedures have recently been analyzed in applied literature. These analyses have led to concerns about the relation between behavioral events and the data which represent them.

Powell, Martindale and Kulp (1975) assessed measurement error due to the length of the observation interval of discontinuous time-based observing/recording procedures. Eight time sampling interval lengths and five interval recording lengths were compared to a continuous measure of "in-seat" behavior. The results indicate that data from the time sampling method more closely approximated the known values when interval lengths shorter than 40 seconds were used. Longer interval lengths overestimated the actual percentage of responding per session. The actual percentage of responding per session was determined by dividing the total response duration per session by the number of minutes per session and multiplying by 100. The partial interval recording procedure overestimated responding regardless of the interval length while the whole interval procedure underestimated the actual percentage of responding. Powell et al. (1975) kept response characteristics such as frequency and duration constant across conditions.

The study by Powell et al. (1975) suggests that measurement error produced by the different observation and recording procedures might be influenced by characteristics of the response. Repp, Roberts, Slack, Repp, and Berkler (1976) compared data obtained by time sampling and interval recording to response frequency. The data was generated across

two conditions: (1) frequency (i.e., high, medium, and low response rates); (2) response patterning (i.e., constant vs. bursting). Repp et al. (1976) generated data using electro-mechanical equipment and thus did not incorporate "live" data from an applied setting as had Powell et al. (1975). The results of Repp's study demonstrated that the time sampling procedure underestimated by the "10/0"² partial interval procedure during high and moderate rate conditions but overestimated low rates of responding. Comparisons made by Repp et al. (1976) to "10/0" procedures are difficult to interpret, since the "10/0" procedure itself may fail to accurately represent responding.

Powell, Martindale, Kulp, Martindale, and Bauman (1977) investigated the effect of response duration on accuracy by varying the percentage of occurrence of "in-seat behavior" per session. They varied the duration of responding so that an adult male was "in-seat" 20%, 50% and 80% of the session. The data from 9 interval lengths of time sampling, partial interval and whole interval observing/recording procedures were compared to the actual percentage of behavior. The time sampling procedures both over and underestimated the actual percentage of behavior. The partial interval observing/recording procedure overestimated the actual percentage of behavior while the whole interval observing/recording procedure underestimated the actual percentage of responding. Powell et al. (1977) did not vary the frequency of the response nor did they show how the combination of response frequency and duration can effect the degree of measurement error pro-

²The 10/0 notation indicates an observing and recording interval in which the initial period is used exclusively for observing the subject and the second period exclusively for recording the data. For 10/0, ten seconds is devoted to observing with recording occurring immediately after and 0 seconds allotted for recording.

duced by the different observing/recording procedures.

Brown and Duncan (Note 1) investigated the ability of the time sampling and partial interval procedures to reflect controlled dimensional changes of frequency and duration. Using electromechanical equipment, they simulated three rates of responding (high, medium, low) and three duration values (long, medium, short). In addition, they obtained data that were produced by the recording of a free operant response of a resident in an institutional setting. Both procedures produced a permanent and continuous record of behavior on paper tape from an event recorder. Ten-second time sampling and ten-second partial interval procedures were used to form data records from the continuous event tapes. Measurement error was computed by subtracting the percent of intervals scored by time-sampling and interval observing/recording procedures from the

actual percent duration of the session time which the "response" occupied. (Compare Powell *et al.*, 1977).

The measurement error found by Brown and Duncan (Note 1) replicated the findings of Repp *et al.* (1976) and Powell *et al.* (1977), in that both time sampling and partial interval observing/recording methods failed to accurately represent actual events. Figure 1 shows that partial interval observing/recording overestimated each known value while time sampling both under and overestimated each known value regardless of frequency or duration.

That discontinuous, time-based measurement systems are common is undeniable; that applied behavior analysis has flourished is clear. But the data presented here just as clearly illustrate that information collected through discontinuous, time-based procedures contains inaccuracies. The inaccuracies occur not just because humans are fallible, but because

Fig. 1. Deviation of time sampling and partial interval data from known values of response duration.

the measurement system itself is not sufficiently sensitive to response dimensions discussed earlier. Science grows by utilizing new discoveries built on past learning. Errors in measurement have been discovered and the goal would seem to be to correct them. Yet powerful arguments remain which discount the importance of the measurement error, or despair of alternative practical solutions.

JUSTIFICATIONS AND CRITICISMS

A variety of reasons are posed for the use of discontinuous, time-based observing and recording systems. What follows is a critical examination of some points offered in support of the ongoing use of these procedures.

Reasonable Accuracy

An influential factor in maintaining the use of discontinuous, time-based measurement practices is the suggestion, by the same research which raises the issue of inaccuracy, that under certain limited circumstances they provide close approximations to continuous measures of selected response dimensions. For example, Powell et al. (1975, 1977) suggest that when time sampling occurs sufficiently often (every 5 seconds) or when time sample values are summed and deviated from data obtained through continuous measurement, close approximations to continuous measurement of response duration data are obtained. Similarly, Alevizos, De Risi, Liberman, Eckman, and Callahan (1978) report procedures designed to illustrate the most "efficient and representative" (p. 249) time sampling procedure. Research reported by Repp et al. (1976) suggests that 10-second interval recording provides reasonably accurate, although conservative, data when response rates vary, whereas time sampling data are not accurate.

In spite of the fact that time or interval sampling may occasionally approximate continuous measures of a response dimension, there is no guarantee that these approximations will be consistent. Research

already discussed demonstrates empirically that these procedures guarantee inaccuracy regardless of response conditions; and even under the best of conditions, predictions about direction of measurement error are impossible. Thus, the researcher knows neither how much nor in what direction the error will occur. Powell et al. (1977) states that "An analysis of interval sampling demonstrates that the results are not directly related to either the frequency or duration of observed behavior" (p. 331). Powell et al. (1977) also comments that "... momentary time samples must be conducted surprisingly frequently for the sample results to mirror consistently the true state of nature, i.e., a continuous measure of behavior" (p. 332).

"Appropriate" Observation and Recording Intervals

Secondly, the point of any applied study is to change a response in some significant manner, often its frequency. But what may be an "appropriate" discontinuous, time-based observation and recording procedure at the beginning of the study may be totally "inappropriate" during final treatment conditions, assuming that the behavior does change. For example, the chosen observing/recording system may appear fairly accurate in measuring stable baseline responding but is unable to detect and accurately represent the variability which occurs with the introduction of the new contingencies. A system which is designed to measure low frequency responding is particularly unsuited to measure behavior occurring at a high frequency. What then is truly the best observation and recording procedure, one which "fits" baseline conditions, or one which "fits" the expected outcome? If the data are biased in order to decrease the chances of obtaining insignificant positive treatment effects, true behavior changes may be masked; if bias favors the intervention, of course, false

positives will occur. In each case, existing relationships between independent and dependent variables remain undiscovered.

Observer Ease

An aspect of several forms of discontinuous time-based observation and recording procedures which makes them particularly alluring is that they do not require constant observation of the subject. Except for the "whole interval" procedure, observers need only observe the subject until the response of interest occurs—as in the "partial interval" procedure, or only at highly restricted points in time, as in the "time sampling" procedure. In the interim the observer may watch the subject or anything else, thereby reducing the tedium which can accompany observing/recording procedures.

Even though they may be less tedious, it should be noted that the discontinuous nature of time and interval sampling techniques serves to obscure relations between environmental events and the response of interest. The study of discriminative control, for example, requires precise records of the temporal relation between S^D and response.

Definitional Ease

Another factor encouraging their use appears to be the lack of precision required in formulating response definitions. Whereas data reflecting the number or frequency of a response would require an explicit designation of response beginning and ending, data reflecting the number or percent of intervals during which a response occurred would not require such precision. For example, when considering children's mutual play, a count or frequency measure would require that each instance of play be clearly demarcated. Discontinuous procedures, on the other hand, would only require that the observer be able to discern whether or not mutual play was occurring at the time of observation. Constructing a response definition from which rate or

duration data can be gathered may be tedious, but the risk in lost measurement accuracy for not adopting this approach to defining the dependent variable should not be ignored.

An appealing aspect of discontinuous observing/recording procedures is the ease with which the observation of multiple responses and/or multiple subjects within the same observation session results in the collection of a great deal of data. Using the mutual play example and a "partial interval" or "time sampling" procedure, the observer could track such diverse responses as play, verbalizations, aggressive responses, and so forth. As soon as one of these responses is noted in the interval, the observer can concentrate on the others. Similarly, an observer can record the play responses of several children by rotating attention at each observation interval.

Again, inaccuracy may be the price paid, in this case, for larger volumes of data. The fact that multiple responses and subjects are being observed and recorded says nothing about the accuracy of these data. Observing and recording multiple responses simultaneously with discontinuous time-time-based intervals can only multiply the error found when one response is so observed and recorded.

A final difficulty with attempts to obtain simultaneous, multiple-response data is that so little data may be gathered on a particular response that the data are not representative of the overall response pattern. When observations rotate from subject to subject, as the number of subjects increases, the number of observations per subject within a session decreases. Data are gathered on more subjects, but the price paid is decreased representativeness. It is difficult to support the contention that either ease of acquisition or quantity of data gathered justifies the loss of measurement accuracy.

Legal and Ethical Issues

It is true that any measurement system

will have some degree of error, but it is a highly questionable practice for professionals to employ measurement procedures which their own literature has demonstrated to be inaccurate. It is important to note here that the inaccuracy objected to is **not** that which results from human error—presumably this will always be with us. The objection is that the very design of discontinuous, time-based procedures creates inaccuracy. It is unclear to what extent one is liable for making decisions about persons based on a measurement system which guarantees errors, however, some precedents have already been set. Policies enacted by the Equal Employment Opportunity Commission require the validation of measurement devices which have disparate effect minorities. In this case, there are clear legal sanctions for failure to provide accurate data for decision making. Similar sorts of sanctions are to be or perhaps have been identified for behavior therapies delivered in the public or private sector.

From a different vantage, it is unclear how benefits accrue to an individual client, the discipline of behavior analysis, or the general culture when the use of a particular measurement system excludes or retards the development of alternative, refined systems and the increased understanding of behavior which might then be obtained.

We've Come A Long Way

A very persuasive and correct argument is that a great deal has and continues to be achieved in applied behavior analysis utilizing discontinuous, time-based observing and recording procedures. This is not to be gainsaid. The critical point is that we still have a long way to go, and several investigators have described some problems in one of the most critical factors which will help us advance, namely our measurement systems. One has only to consider advances in biology which occurred in relation to the development of

the microscope, or advances in astrophysics related to the development of radio telescopes to see how influential changes in measurement procedures can be. Clearly biology and physics had progressed prior to the development of these measurement procedures, but the rate of progress changed dramatically afterward. To the extent that measurement procedures in applied behavior analysis fail to accurately reflect aspects of responding, and that failure can be considerable, our progress will be delayed.

The objections to discontinuous time-based observation and recording procedures raised here require at least suggestions about viable alternative measurement strategies. Such suggestions follow, with their common theme being a return of measurement strategies which provided the basis for the identification of behavioral principles.

DIRECT, CONTINUOUS MEASUREMENT

The measurement practices offered as an alternative to discontinuous, time-based systems are based on the identification of units of measurement which reflect response dimensions derived from fundamental properties of behavior. If a particular dimension of a response is experimentally or clinically important, a unit of measurement sensitive to that dimension can be used. For example, if response duration is of interest, a form of time units, possibly minutes, could be the unit of measurement, and the duration data reported as "elapsed minutes" or "total duration in minutes."

At this point, the objection might be raised that too much is being asked of measurement practices, and that some sort of sampling procedure is more reasonable, namely time or interval sampling. Whether or not to sample is not the issue. Unless behavior can be monitored 24 hours a day, samples must be taken. The frequency and length of the samples are important but separate

research issues. The principle point is that, once observation and recording begin, the data acquired represent the universe of response dimensions of interest, i.e., the total number of instances, the total duration, etc., which occur during the observation session. Sampling of this sort provides data about response classes which directly reflect relevant dimensions. Since continuous measurement does not necessarily require any more observation time than discontinuous measurement, it offers the possibility of more information, in terms of accurate descriptions of response dimensions, at no increase in cost.

Procedures which allow for the direct, continuous measurement of response dimensions may have to be more technologically sophisticated than the commonly used time and interval-sampling procedures. This sophistication, to a large extent, already exists. Where it does not, it must be provided by research. This research can take at least two forms: one directed at developing better equipment and the other focused on variables which contribute to the effective use of human observers as transducers of behavioral events.

Skinner and subsequent researchers in the experimental analysis of behavior designed measurement procedures which obviated many of the problems raised by current measurement practices—they arranged for the response to directly activate an automated recording device. In the applied analysis of behavior, relatively little use is made of equipment to observe and record behavior. Some examples do exist, however. Schulman, Stevens, Swan, Kupst, and Naughton (1978) describe a device (a biomotometer) which utilizes mercury switches to reflect activity levels. O'Brien and Azrin (1970) described a portable apparatus which both measured postural changes and provided feedback to a subject. Research on sexual problems utilizes data from strain gauges (e.g., Abel, Blanchard, Barlow &

Mavissakalian, 1975) which are directly activated by the subject's sexual arousal.

These examples provide, perhaps, the best form of direct, continuous measurement since accuracy is guaranteed within the limits of the devices themselves. Human observation error does not play a role. The response of the subject is both defined and measured directly by the apparatus.

When automatic response—activated measurement is not possible, devices may still be used by human observers to record behavior. A device with great promise is a small (3 x 4 x 1.5"), 3-channel encoder/decoder developed by Magyar and Fitzsimmons (1979), which allows a human observer to record laboratory and field behavior. Records taped on this device can then be played through standard programming equipment, allowing for as sophisticated an analysis of the behavior as desired. These examples are cited to emphasize the availability and utility of equipment which will itself make direct, continuous records or which permits a human observer to make such records. Finally, it may be that human observers are used with no more sophisticated equipment than a stopwatch, paper, and pencil. Time units and response counts can still be made with such "equipment," but human error can also occur. It is likely that many factors influence the accuracy of data collected by humans using direct, continuous procedures, and research designed to identify and reduce errors is indicated. At the very least, however, the accuracy of data so collected will not be reduced by inherent difficulties in the measurement procedure.

MULTIPLE RESPONSE DIMENSIONS

The continuous measurement of response dimensions in the applied setting is not a new idea. Kelly (1977) found in his sample that 29% of the studies measured frequency. While 9% measured duration, none of the studies that measured fre-

quency or duration as reported by Kelly (1977) recorded more than one dimension of behavior. However, more than one dimensional quantity can have clinical or experimental importance in an applied setting. Measuring only one response dimension, at the exclusion of others, results in a loss of information that might have had important clinical or experimental implications. If the information is not recorded as it exists in the natural environment than it can never be retrieved. Also, as environmental events are delivered contingent upon the performance of a specific response dimension, the probability of other dimensions changing is high. That is, as one response dimension is designated and programmed for change the other response dimensions also change. Changes in other response dimensions could be a variable which affects the future probability of the response dimension of initial interest. Failure to monitor other response dimensions precludes knowledge of multiple dimensional changes. The result is a lack of clarity in describing the effects of the independent variable.

On a conceptual basis there is strong support for the premise that dimensions not identified for change will often do so. When an event is made contingent upon a specific response, all dimensions of the response may be affected. For example, if the dimensional quantity of interest is frequency, as in number of times "out-of-seat" in a classroom setting, then the introduction of the independent variable may also affect duration, time spent out-of-seat, and IRT (time between instances of out-of-seat behavior). The implications of such a conceptual premise on the behavior of an investigator in the applied setting are multi-faceted. The investigator would be wise to target both response dimensions that are of current clinical significance, and those that might change as a result of the changes in contingencies and therefore become significant.

Data from the experimental literature

show that the various response or response-sequence dimensions can vary independently. A few studies have examined changes in response duration as an indirect effect of altering the schedule of reinforcement. Schaefer and Steinhorst (1959) measured duration of responding in the transition from continuous reinforcement (CRF) to fixed ratio 5 (FR5), FR10, and CRF to FR10. Their data indicate that duration temporarily increases following schedule changes and then returns to its previous level. Millenson, Hurwitz, and Nixon (1961) describe three experiments which examine the relationship between duration and various fixed interval (FI) schedules. Their primary conclusion was that when rats are exposed to fixed interval contingencies following CRF, the central tendency of the durations of these responses remain two to three times higher than under CRF.

A number of other investigators have reported changes in response duration when the organism was placed on extinction (Marguiles, 1961; Millenson & Hurwitz, 1961; Notterman, 1959). Their general finding was that extinction disrupts duration stability, i.e., results in greater numbers of long responses. Thompson, Heistad, and Palermo (1963) studied the effects of amount of training prior to extinction on frequency and duration of responding during extinction. Using rats and children as subjects they concluded that the length of training differentially affects the tendency of organisms to react to extinction onset by responding faster and for longer periods per response.

In most instances, research monitoring changes in more than one response dimension has been confined to responding under fixed ratio and fixed interval contingencies or extinction. These studies have also relied primarily on the use of laboratory animals as subjects.

Springer, Springer, Brown and Duncan (Note 2), using human subjects and a lever press response, manipulated inter-

response-times (IRT's) and duration in an attempt to determine the effects these changes would have on other response dimensions. Two subjects were exposed to alternating IRT requirements of ≤ 1 second and ≥ 4 seconds. A third subject was exposed to alternating duration requirements of ≤ 1 second and ≥ 4 seconds. The differential reinforcement of duration (DRD) schedule required the subject to depress a lever for a minimum time before reinforcement was delivered. Reinforcement was delivered when the microswitch was released; there were no exteroceptive stimuli to indicate when the required time had elapsed. In all conditions of the study, both rate and duration were recorded through the use of electro-mechanical equipment. The results of the study can be seen in Figures 2, 3, and 4.

Data from the three subjects illustrate the independence of the response dimensions under investigation. Figures 2 and 3 show that results from the two IRT sub-

jects are similar with respect to rate and IRT's, but that concurrent changes in duration were unique for each individual. The first subject showed only slight fluctuations in duration following the onset of new schedule requirements. This finding is comparable to that obtained with individuals in transition from CRF to FI (Millenson, *et al.*, 1961). The second subject, however, showed considerable changes in response duration.

Figure 4 suggests that Subject 3 had great difficulty in conforming to the requirements of the duration schedules. Large numbers of short duration presses occurred throughout the first two phases and the subject was very slow in reducing press durations when the schedule was changed from $\text{DRD} \geq 4$ seconds to $\text{DRD} \geq 1$ second. Although the control exerted over Subject 3's behavior was much weaker than it was for Subjects 1 and 2, these data served as a good example of the inadequacy of rate as the only measure of

Fig. 2. Data for subject 1, showing the interaction of IRT, duration, and rate.

Fig. 3. Data for subject 2, showing the interaction of IRT, duration, and rate.

Fig. 4. Data for subject 3, showing the interaction of IRT, duration, and rate.

behavior change. As duration varied both within and between experimental phases, rate remained comparatively constant. At points where there was a significant increase or decrease in rate, duration changed in the opposite direction. Obviously then, the effect of these contingencies on responding cannot be described in terms of a single dimensional quantity.

The data obtained by Springer *et al.* (Note 2) indicate that duration is a dimensional quantity which can vary independently of other response dimensions. Subject 1 exhibited only slight, temporary changes in duration as rate increased and decreased according to the schedule. The second subject's response duration increased as rate decreased and Subject 3 showed great session to session fluctuations in duration while rate remained generally stable.

In similar research, Stephenson and Johnston (Note 3) report that the duration and frequency of rumination episodes/day by an institutionalized individual showed that across days with similar frequencies the total duration of rumination fluctuated substantially. Again, failure to measure multiple response dimensions would have prevented this analysis from occurring.

The experimental analysis of behavior began and grew using response rate as its fundamental datum. Among the reasons for this selection was the fact that other response dimensions, especially duration, were trivial, the duration of a lever press or key peck not being of particular importance in describing functional relations. Applied behavior analysts, however, work daily with responses in which several dimensions are of importance. Tantrums may be important because of their frequency, duration, or other dimensional characteristics. The ruminating child studied by Stephenson and Johnston (Note 4) also illustrates multi-dimensional importance. Which dimensions show functional relations? They all might. To obtain the clearest picture of existing

functional relations, then, those dimensions of interest or importance should be measured.

SUMMARY AND CONCLUSIONS

The preceding discussion has examined issues related to measurement in applied behavior analysis. This discussion has suggested, on both conceptual and empirical grounds, that the use of discontinuous, time-based observation and recording procedures are inherently inaccurate, and that their continued use can but retard the discovery of clinically or experimentally important functional relations.

Offered as an alternative are observation and recording practices which allow for direct, continuous measurement of response dimensions. Oftentimes, for either clinical or theoretical reasons, multiple response dimensions must be measured to determine the impact of a particular contingency. Whereas direct, continuous measurement is ideally conducted with automatic devices, human-transduced measurement is also possible.

The improvement of measurement practices in applied behavior analysis is an issue integrally related to the rapid advancement of the field. This concern with current measurement practices is a specific example of a more general concern expressed by Dietz (1978) that the field of applied behavior analysis was losing its contact with the theory and methods of the experimental analysis of behavior. Dietz calls for a return to an analytic, investigative approach characteristic of traditional behavior analysis. The same logic suggests that the strategies which have formed the bases for measurement in the experimental analysis of behavior, and which have led to the current levels of theoretical/conceptual understanding of behavior, should serve as the foundation for measurement in the field of applied analysis of behavior as well.

REFERENCE NOTES

1. Brown, T. & Duncan, P.K. *A comparison of the accuracy of continuous and discontinuous observation and recording procedures*. Paper presented at the Association for Behavior Analysis, 1979.
2. Springer, B., Springer, M., Brown, T. & Duncan P. *An empirical analysis of the relations between response frequency and duration using continuous measurement*. Paper presented at the Association for Behavior Analysis, 1979.
3. Stephenson, H. & Johnston, J.M. *Measurement of variability in dimensional quantities of behavior*. Paper presented at the Association for Behavior Analysis, 1979.

REFERENCES

- Abel, G., Blanchard, E., Barlow, D., & Mavissakalian, M. Identifying specific erotic cues in sexual deviations by audiotaped descriptions. *Journal of Applied Behavior Analysis*, 1975, 8, 247-260.
- Alevizos, P., DeRisi, W., Liberman, R., Eckman, T. & Callahan, E. The behavior observation instrument: A method of direct observation for program evaluation. *Journal of Applied Behavior Analysis*, 1978, 11, 243-257.
- Dietz, S. Current status of applied behavior analysis. *American Psychologist*, 1978, 33, 805-814.
- Hall, R.V. *Behavior modification*, Vol. I: The measurement of behavior. Lawrence, Kansas: H & H Enterprises, 1971.
- Johnston, J.M. & Pennypacker, H.S. *Strategies and tactics of human behavioral research*: Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.
- Kazdin, A.E. *Behavior modification in applied settings*. Homewood, Illinois: The Dorsey Press, 1975.
- Kelly, M.B. A review of the observational data-collection and reliability procedures reported in the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 1977, 10, 97-101.
- Magyar, R.L. & Fitzsimmons, J.R. A multi-channel, portable, "real-time" event encoder-decoder for laboratory and field experiments. *Behavior Research Methods & Instrumentation*, 1979, 11, 47-50.
- Marguiles, S. Response duration in operant level, regular reinforcement, and extinction. *Journal of the Experimental Analysis of Behavior*, 1961, 4, 317-321.
- Millenson, J.R. & Hurwitz, H.M.B. Some temporal and sequential properties of behavior during conditioning and extinction. *Journal of the Experimental Analysis of Behavior*. 1961, 4, 97-106.
- Millenson, J.R., Hurwitz, H.M.B. & Nixon, W.L.B. Influence of reinforcement schedules on response duration. *Journal of the Experimental Analysis of Behavior*, 1961, 4, 243-250.
- Notterman, J.M. Force emission during bar pressing. *Journal of Experimental Psychology*, 1959, 58, 341-347.
- O'Brien, F. & Azrin, N. Behavioral engineering: control of posture by informational feedback. *Journal of Applied Behavior Analysis*, 1970, 3, 235-240.
- Powell, J., Martindale, A. & Kulp, S. An evaluation of time-sample measures of behavior. *Journal of Applied Behavior Analysis*, 1975, 8, 463-464.
- Powell, J., Martindale, B., Kulp, S., Martindale, A. & Bauman, R. Taking a closer look: Time sampling and measurement error. *Journal of Applied Behavior Analysis*, 1977, 10, 325-332.
- Repp, A.C., Roberts, D.M., Slack, D.J., Repp, C.F. & Berkler, M.S. A comparison of frequency, interval and time-sampling methods of data collection. *Journal of Applied Behavior Analysis*, 1976, 9, 501-508.
- Schaefer, H.H. & Steinhurst, R.A. The effect of changing the schedule of reinforcement upon duration of responding. *Journal of the Experimental Analysis of Behavior*, 1959, 2, 335-341.
- Schulman, J., Stevens, T., Swan, B., Kupst, M. & Naughton, M. Modification of activity level through biofeedback and operant conditioning. *Journal of Applied Behavior Analysis*, 1978, 11, 145-153.
- Skinner, B.F. *Behavior of Organisms*. Englewood Cliffs, N.J.: Prentice-Hall, 1938.
- Thompson, T., Heistad, G.T. & Plaermo, D.S. Effect of amount of training on rate and duration of responding during extinction. *Journal of the Experimental Analysis of Behavior*, 1963, 6, 155-161.