JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1976, 9, 501-508 NUMBER 4 (WINTER) 1976

A COMPARISON OF FREQUENCY, INTERVAL, AND TIME-SAMPLING METHODS OF DATA COLLECTION

ALAN C. REPP,[1] DAVID M. ROBERTS, DANIEL J. SLACK, CHRISTINA F. REPP, AND MARGO S. BERKLER

GEORGIA RETARDATION CENTER

Data representing high, medium, and low response rates in constant and nonconstant patterns were generated by electromechanical equipment to determine whether the same data collected by time-sampling, interval recording, and frequency recording would be represented similarly by each method. Results indicated: (1) that time-sampling provided an extremely inaccurate estimate of responding, and (2) that interval recording accurately represented responding of low and medium rates, but grossly underestimated high-rate responding.

DESCRIPTORS: behavioral recording, time sample, frequency recording, interval recording, response rates

---

In the experimental analysis of behavior, responses are usually defined as the electrical or electromechanical operation of equipment, which results in discrete records. The data produced by such records usually are summarized and presented as rate of responding (Ferster and Skinner, 1957; Skinner, 1938, 1959, 1966). In field experiments, observers often serve as transducers between the subject's behavior and the record of that behavior. For various reasons, usually associated either with the complexity of the observation system or with the duration of the response, rate is not so ubiquitously used as the basic datum in field studies (Pisor, Note 1). It is often replaced by per cent, which is frequently used to describe the fraction of intervals of observation in which behavior is recorded as having occurred.

To produce data for applied studies, various collection methods are used, the most common of which are *frequency recording, interval recording,* and *time sampling.* In frequency recording[2] (Bijou, Peterson, and Ault, 1968), data may be collected over several relatively large segments of the session or over the entire session (with the latter by far the most prevalent of the two), and the data are presented as responses per unit time or as the number of responses per session.

In interval recording[2] (Bijou *et al.,* 1968), the session is either: (a) divided into a number of equal-interval observation periods and the same number of equal-interval recording periods, or (b) divided only into a number of

---

[1]Reprints may be obtained from Alan C. Repp, Department of Special Education, Northern Illinois University, DeKalb, Illinois 60115.

[2]While these terms are common, other authors have used equally appropriate terms to describe these methods of data collection. For example, Thomson, Holmberg, and Baer (1974) differentiated between recording procedures and labelled them either *continuous time-sampling* or *intermittent time-sampling.* In those terms, the present study employed: (1) continuous time sampling with a 10-sec interval, and (2) intermittent time sampling with observations: (a) 10 of every 15 sec, (b) 5 sec of every 10 min, and (c) 10 sec of every 10 min. Powell, Martindale, and Kulp (1975) used two terms related to the procedures described here: (1) *partial interval time sampling* (our interval recording), in which an interval is scored as one of occurrence if a response occurs in any portion of it, and (2) momentary time sampling (our time sampling). Jackson, Della-Piana, and Sloane (1975) discussed several procedures, three of which are related to this study: (1) *interval method* (the same as our interval method), (2) *instantaneous time sampling* (our time sampling), and (3) *tally method* (our frequency method). Both Miller (1975) and Alevizos, Campbell, and Callahan (1975) used the terms interval recording and time sampling, but used *event recording* to describe the procedure labelled in this paper as frequency recording.

equal-interval observation-recording periods. In the former procedure, the observer watches the subject for a brief period (*e.g.,* 10 sec) and, during what is usually a shorter period (*e.g.,* 5 sec), records whether or not responding occurred in the preceding interval; in the latter procedure, observation and recording occur in the same interval. Data then are presented as the per cent of the observation intervals in which responding occurred.

In time sampling[2] (Baer and Wolf, 1967), the observer ignores the behavior for a relatively long period (*e.g.,* 10 min), and then, after a signal, quickly scans the situation and records the behaviors observed. Data usually are presented as the per cent of observations in which responding occurred, although they can also be presented as responses per unit time if the observer counts all the responses and divides by the total observation time.

The purpose of the present experiment was to compare data obtained by time sampling, interval recording, and frequency recording to determine whether the same data would be similarly represented by each method. The data were generated in two conditions: (1) as high, medium, or low rates of responding, and (2) as responding of a constant rate across the session or as responding that occurred in bursts and was not constant across the entire session.

METHOD

Data

Data were generated by electromechanical equipment, with records of the data preserved by an event recorder. To produce records of pseudobehavior at different rates and of different patterns, a pulse of 35-msec duration was sent through a probability gate (BRS/LVE PB-902/235-11) to a pen on the event recorder. To produce a permanent record of time, other pens were activated every 5 sec, 10 sec, and 10 min.

Rates of responding. Conditions of constant rate were generated in three ways. With the probability gate set at $p = 0.3$ and the electrical pulses spaced at 1.8 sec, the mean rate generated was 10 per minute. With the probability gate set at $p = 0.3$ and the pulses spaced at 18 sec, the mean rate was 1 per minute; with the probability gate set at $p = 0.3$ and the pulses spaced at 180 sec, the mean rate was 0.1 per minute.

Patterns of responding. Two patterns of responding were generated through the probability gate and were either: (1) constant across the entire session, or (2) burst, where for each 60 min, 75% of the pen deflections occurred in the first 30 min, and 25% occurred across the next 30 min. The constant condition was produced by establishing a constant pulse train across the session, while the burst condition was produced by changing the pulse train after the first half hour (For example, in a 180-min session, high-rate responding [$\bar{x} = 10$ per minute] was simulated by generating two rates of responding. In the first 30 min, pulses spaced at 2 sec were generated at a probability of 0.5 and were scheduled to produce 450 pen deflections. In the next 30 min, pulses spaced at 6 sec were generated at a probability of 0.5 and were scheduled to produce 150 pen deflections. This series then was alternated throughout the 3-hr session. Similar programs produced records to simulate both the medium- and the low-rate conditions).

Sessions. Sessions were arbitrarily defined as 180 min in length. Eighteen sessions defined the length of the experiment and provided three sessions for each of the six conditions [(high, medium, or low rates) x (constant or burst response patterns)].

Methods of Data Collection

Time sampling. Two time-sampling conditions were used. First, all pen deflections for the first 9 min and 55 sec of the session were ignored. Pen deflections during the next 5 sec were then scored as responding. This 9-min 55-sec/5-sec series was repeated throughout the session, and data were reported either as (a) the per cent of the 18, 5-sec intervals per session in which at least one pen deflection occurred, or (b) as the rate of responding, expressed in responses per minute, for the 90 sec during which the total number of responses in the 18, 5-sec time samples could be counted. This 9-min 55-sec/5-sec series was selected to simulate field experiments in which the observation-recording period is typically quite short.

In the second time-sampling procedure, all pen deflections were ignored during the first 9 min and 50 sec of the session and scored during the next 10 sec. This series was then repeated throughout the 3-hr session, providing 18, 10-sec observation-recording intervals. As in the previous example, data were reported either as the per cent of intervals in which responding occurred or as the rate of responding.

Interval recording. To simulate interval recording, two procedures were again used. In the first, the initial 10-sec period of the session was categorized as an observation interval, while the next 5-sec period was categorized as a recording interval, with pen deflections during this period being ignored. This 10-sec/5-sec series (labelled 10/5) was continued and presented 720 observation periods per session. For the second (labelled 10/0), all 1080, 10-sec intervals were scored. An observation interval was scored as one of responding if at least one pen deflection was present, and the data were reported as the per cent of the observation intervals in which deflections occurred.

Frequency recording. In this procedure, the responses occurring in the session were counted and divided by 180 to provide a summary of responding expressed as responses per minute.

Interobserver Agreement

Each tape was scored independently by two observers, and observer-agreement scores were calculated in each session for each of the three methods. In the time-sampling method, the Exact Agreement—Response Intervals Only (Repp, Deitz, Boles, Deitz, and Repp, 1976) procedure was used as a stringent means for assessing interobserver agreement. In this method, each of the 18 observation-recording intervals was classified as "nonresponding" if neither scorer reported a pen deflection in that interval. If, however, at least one observer scored the interval as containing at least one pen deflection, that period was classified as a "responding" interval. The number of "responding" intervals that both observers scored as containing the same number of pen deflections was divided by the total number of "responding" intervals and multiplied by 100 to provide a measure of the per cent of interobserver agreement. In all 18 sessions, the agreement was 100%. In the interval recording method, interobserver agreement was assessed in similar fashion, and agreement was 100% in all 18 sessions.

Interobserver agreement for the frequency recording method was assessed quite differently. The scorers independently counted the pen deflections over the entire session, divided the larger count into the smaller and multiplied by 100 to provide an assessment of interobserver agreement. In each session, the agreement reported was 100%.

RESULTS AND DISCUSSION

Comparisons of Per Cent Measures

In the methods reporting the per cent of intervals in which responding occurred, the observe-and-record (10/0) method was the only one to provide a record of all intervals in which responding occurred; all others were, in effect, samples of these data. As such, the observe-and-then-record (10/5) method, the 5-sec time-sampling, and the 10-sec time-sampling methods were each analyzed as the deviation in raw score from the 10/0 method. Figure 1 describes this difference for one representative session from each of the possible rate (high, medium, or low) *x* response pattern (constant or burst) conditions. The data from the other two sessions of each condition were so similar to the single session reported that they were omitted to provide graphs that were easier to read and in order to eliminate redundant data.

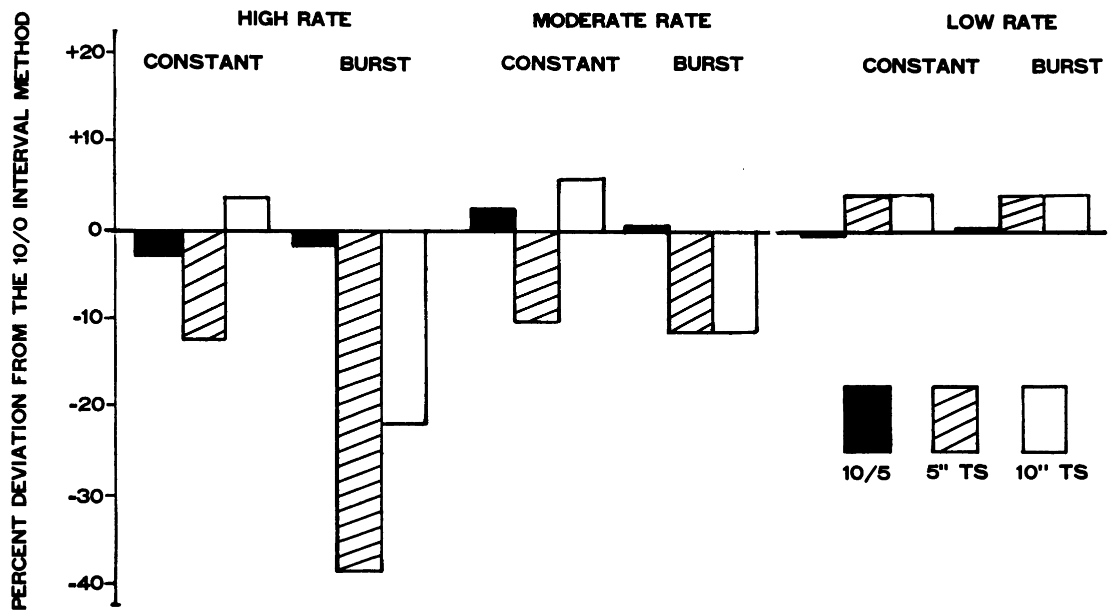

Fig. 1. The difference in representative sessions between each of the 10/5 interval, 5-sec time-sampling, and 10-sec time-sampling methods and the 10/0 interval method. The per cent of intervals in which responding was reported for the 10/0 method for the six conditions were: (1) high-rate/constant = 84.7%, (2) high-rate/burst = 82.8%, (3) medium-rate/constant = 16.2%, (4) medium-rate/burst = 16.9%, (5) low-rate/constant = 1.8%, and (6) low-rate/burst = 1.7%.

The comparisons in this figure indicate that the 10/0 and 10/5 interval methods differed very little (<3%) in each condition. In effect, there was no appreciable difference in the representativeness of the data between the interval method in which a person both observes and records in the same interval and the method in which a person observes in one interval and records in the succeeding interval. One must be careful, however, to restrict this conclusion to the representativeness of the data and not to extend it to the accuracy of the data. Because all pen deflections in this experiment were generated by programming equipment, and because the deflections themselves were the environmental events of interest, rather than reflections of other response events in the environment, the data were accurate in and of themselves. However, in studies in which the data are not simulated, response records are reflections of other environmental events and recording errors certainly occur. Whether the 10/0 or the 10/5 method is more accurate is a separate question that should, perhaps, be answered by each experimenter for each study in which interval recording is used. Because in applied studies, a human typically acts as a transducer between the subject's behavior and a record of that behavior, such variables as the number of response classes being recorded, the rates of responding, and the size of the observation interval could result in differences in the accuracy of the data produced by the 10/0 and 10/5 methods.

Figure 1 shows that the time-sampling method does not produce representative record samples of data, particularly when responding is not of a constant rate across the entire session and when it is of a medium (1 rpm) or high (10 rpm) rate. In the low-rate (0.1 rpm) conditions, the time-sampling method overestimated responding by less than 4% (5.5% to 1.8%), a figure that might lead to the conclusion that data from the time-sampling method are very similar to data from the 10/0 method. However, when the data estimates from time sampling are divided by data from the 10/0 interval method, the result indicates an overestimation of approximately 300% (5.5% ÷ 1.8). In these terms, the differences are particularly important both for those who are interested in reducing low-rate—high-intensity behaviors and for those who report their results as ratios of responding between baseline and treatment phases.

These results clearly indicate that: (1) time sampling did not provide data that were accurately reflective of environmental events, (2) the degree of difference was perhaps greater than expected from its considerable use (Hutt and Hutt, 1970), (3) the absolute difference between the 10-sec time-sampling and interval methods decreased as the rate of responding decreased, (4) the ratio of the scores generated by these two methods increased as the rate of responding decreased, and (5) the data from the burst condition tended to be less representative of the 10/0 method than data from the constant condition.

Comparisons of Rate Measures

In the methods reporting the rate of responding, the frequency method provided a record of all occurrences of responding, while the 5-sec and 10-sec time-sampling provided, of course, only samples of the session. To compare data from these three sources, rates of responding were calculated from a representative session for: (1) the 180 min provided by the frequency method, (2) the 90 sec provided by the 5-sec time-sampling method, and (3) the 180 sec provided by the 10-sec time-sampling method. These data are presented in Figure 2 as the difference in responses per minute between the frequency method and the two time-sampling methods. For all but one condition (medium-rate/burst), the rates provided by the 10-sec time samples were as accurate or more accurate than the rates provided by the 5-sec time samples. The absolute differences decreased as the rate of responding decreased from 10 rpm to 1 rpm, but the ratio of differences (*e.g.,* rpm from time-sampling ÷ rpm from frequency) increased as the rate of responding decreased, with the latter approximately 500% for the 5-sec time sample/frequency comparison. Again, although the absolute difference expressed in rpm may be small when the response rate is low, the difference when expressed as a ratio of the two reported rates can be large, and can be very important for those who express behavioral change as a ratio of responding between treatment and baseline conditions. As there is no reason that the time-sampling methods should consistently overestimate responding at these rates, the difference is not a constant across both baseline and treatment conditions and cannot be ignored.

Comparison of Interval and Frequency Methods[3]

The interval and frequency methods represent differences in both: (a) recording procedure [*i.e.,* marking the absence/occurrence of responding during many short intervals *versus* marking every occurrence of responding during a much longer interval (the session)] and (b) calculation (per cent *versus* rate) and, as such, cannot be compared to each other as directly as each can be compared to time sampling.

One method of comparing these methods is to measure the amount of information lost when one collects data by the interval method, which indicates only that at least one response has occurred instead of indicating how many responses occurred. To aid in this comparison, all 1080, 10-sec blocks in each session were categorized as containing 0, 1, 2, 3, 4, or 5 responses. The number of intervals in which *at least* one response occurred was counted and is represented by x. The number of intervals in which *only* one response occurred was then counted and is represented by y. An "index of information lost", represented by I, was then calculated to represent the per cent of responding about which the interval method, by virtue of indicating only that at least one response occurred, provided no information.

[3]Another description, in which the relationship between interval and frequency recording is expressed mathematically, is available on request.

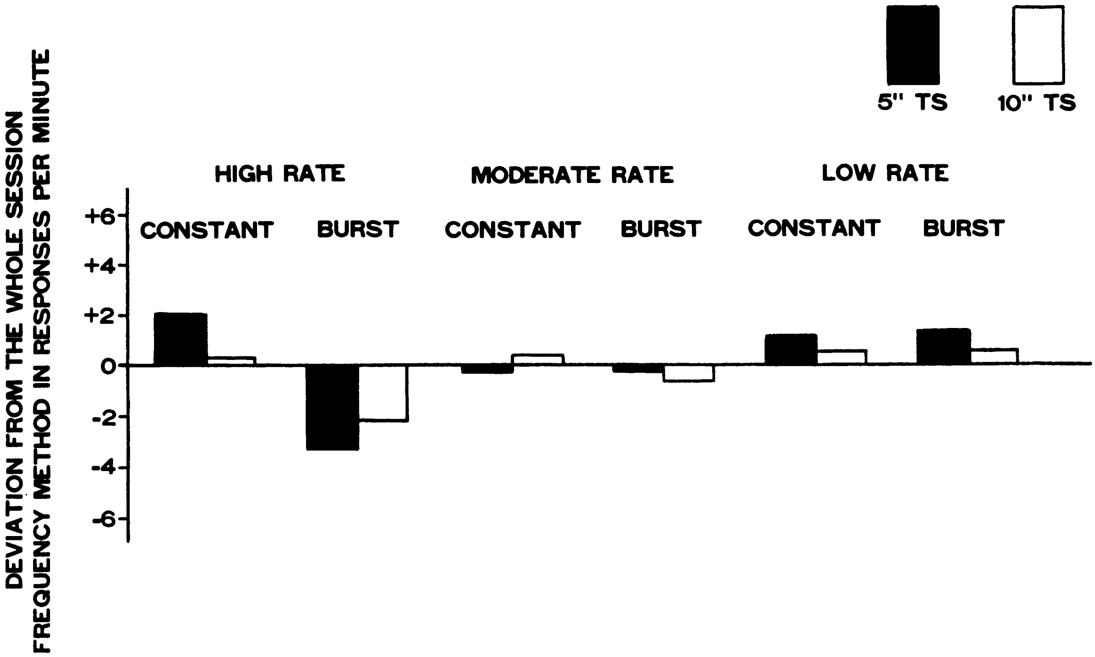

Fig. 2. The difference in representative sessions between each of the two time-sampling methods and the frequency method. The rates of responding reported for the frequency method for the six conditions were: (1) high-rate/constant = 9.9 rpm, (2) high-rate/burst = 10.1 rpm, (3) medium-rate/constant = 1.0 rpm, (4) medium-rate/burst = 1.0 rpm, (5) low-rate/constant = 0.1 rpm, and (6) low-rate/burst = 0.1 rpm.

For the low-rate and medium-rate conditions, there were no intervals in which more than one response occurred; for the two high-rate conditions, there were many such intervals. Where $I = (x - y)/x$, the index of lost information for the (1) high-rate/constant and the (2) high-rate/burst conditions were:

$$(1)\quad I_1 = (920 - 350)/920$$
$$I_1 = 0.62$$
$$(2)\quad I_2 = (895 - 365)/895$$
$$I_2 = 0.59$$

These results indicate that for the two high-rate conditions, the interval recording method produced no information on 62% and on 59% of the events. Thus, with high rates of responding (10 rpm), experimenters interested in maximum information about the environment should either use the frequency method of recording or the interval method with extremely small intervals. One means of obtaining an estimate of appropriate interval size for responding of very short duration would, of course, be to record all responding for several sessions, convert the data to rpm, and then divide this number into one. The result would estimate the average interresponse time and would provide the experimenter with an interval size that should not, for stable responding, contain too many intervals with more than one response. Another, more exacting method, would be to determine the smallest interresponse time during a prebaseline period and use this value as the interval size during the ensuing sessions (Repp, Note 2).

CONCLUSION

With the present findings, three of the conclusions relevant to findings of other experiments are:

(1) that time-sampling of even moderate interobservation periods does not provide data that properly represent events in the environment, and some conclusions from studies using this data collection method are clearly in question.

(2) that interval recording may be inaccurate for high-rate responding but, instead of being weakened, some conclusions from studies employing this data collection procedure could be strengthened. Because interval recording with small intervals reports most responding that occurs at medium or low rates, baseline data in experiments that intend to increase responding are reasonably reflective of environmental events. However, because interval recording employing 10-sec intervals does not report a considerable amount of responding that occurs at high rates, successful treatment phases do not report all appropriate responding, and many experiments may actually have greater treatment-baseline differences than reported.

Similarly, experiments concerned with the reduction of high-rate responding may have been more successful than authors initially reported. Baseline reports of inappropriate responding probably underestimate high-rate responding, while successful treatment phases employing interval recording probably report almost all the medium- or low-rate responding that actually occurs. Again, baseline-treatment differences may be greater than those reported.

Choice of the frequency method may have produced data much more indicative of changes in responding in these experiments. The degree of difference can be shown quite simply with either the stable or the burst data from the present experiment. Presume that a procedure successfully reduced responding from high rate to medium rate. If this change were expressed in terms of rate, the change would be a factor of 10 (10 rpm ÷ 1 rpm); if the change were expressed in terms of per cent, as provided by interval recording of the same event recorder tapes, the change would be a factor of 5 (84.7% ÷ 16.2% for stable responding and 82.8 ÷ 16.9% for burst responding). Similarly, if responding were reduced from high rate to low rate, the reduction would be by a factor of 100 (100 ÷ 1) if the data were recorded by the frequency method and by a factor of approximately 48 (84.7 ÷ 1.8 and 82.8 ÷ 1.7) if recorded by the interval method.

(3) that interval recording may be responsible for some experimenters discarding results as not dramatic enough to pursue with further experiments or to publish intact. If an independent variable reduced responding from high to medium rates, the procedure might have been discounted as not being powerful enough for further investigation. However, because the data collection procedure may have been responsible for underestimating the occurrence of high rate events, the results from the procedure may have been more significant than the baseline-treatment differences suggested. Unfortunately, the effect of this error produced merely by the data collection procedure cannot be rectified.

The present experiment indicated that there is an interaction among the rate and pattern of responding, the method of data collection, and the data reported. As such, it adds: (a) to the work by Thomson, Holmberg, and Baer (1974), which showed that the manner in which the observation periods are arranged within the session affects the data produced, and (b) to the work by Powell, Martindale, and Kulp (1975), which showed that the length of the interobservation period affects the representativeness of time-sampling data and that the length of the observation interval affects the representativeness of data collected through interval recording.

REFERENCE NOTES

1. Pisor, K. Data display techniques. Unpublished manuscript, Department of Psychology, Georgia State University, 1976.
2. Repp, A. C. (Chmn.) Strategies for observation systems in applied studies of the behavior of retarded persons: A symposium. American Association on Mental Deficiency, 1976.

REFERENCES

Alevizos, P. N., Campbell, M. D., and Callahan, E. J. Communication. *Journal of Applied Behavior Analysis,* 1975, **8**, 660.

Baer, D. M. and Wolf, M. M. The reinforcement contingency in remedial and preschool education.

In R. D. Hess and D. M. Baer (Eds.), *Early education: current theory, research, and action.* Chicago: Aldine, 1968. Pp. 119-129.

Bijou, S. W., Peterson, R. F., and Ault, M. H. A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis,* 1968, **1,** 175-191.

Ferster, C. B. and Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

Hutt, S. J. and Hutt, C. *Direct observation and measurement of behavior.* Springfield, Illinois. Charles C Thomas, 1970.

Jackson, D. A., Della-Piana, G. M., and Sloane, H. N. Jr. *How to establish a behavior observation system.* Englewood Cliffs, New Jersey: Educational Technology, 1975.

Miller, L. K. *Principles of everyday behavior analysis.* Belmont, California: Wadsworth Publishing Company, 1975.

Powell, J., Martindale, A., and Kulp, S. An evaluation of time-sample measures of behavior. *Journal of Applied Behavior Analysis,* 1975, **8,** 463-464.

Repp, A. C., Deitz, D. E. D., Boles, S. M., Deitz, S. M., and Repp, C. F. Differences among common methods for calculating interobserver agreement. *Journal of Applied Behavior Analysis,* 1976, **9,** 109-113.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

Skinner, B. F. *Cumulative record.* New York: Appleton-Century-Crofts, 1959.

Skinner, B. F. Operant behavior. In W. K. Honig (Ed.), *Operant behavior: areas of research and application.* New York: Appleton-Century-Crofts, 1966. Pp. 12-32.

Thomson, C., Holmberg, M., and Baer, D. M. A brief report on a comparison of time-sampling procedures. *Journal of Applied Behavior Analysis,* 1974, **7,** 623-626.

Received 1 December 1975.
(Final acceptance 22 April 1976.)