*SIMPLIFYING CONTINUOUS MONITORING OF
MULTIPLE-RESPONSE/MULTIPLE-SUBJECT
CLASSROOM INTERACTIONS*

THOMAS M. SKRTIC AND HARVEY J. SEPLER

THE UNIVERSITY OF KANSAS AND VANDERBILT UNIVERSITY

In order to facilitate the field monitoring of three subjects interacting according to one or more of 18 response categories, a modified version of several available, but oftentimes mechanically incompatible, observational procedures was designed. Its continuous recording strategy, sectioned into one-minute observational units, enabled researchers to derive highly representative behavior samples, and when accompanied by the specially tailored coding form and recording apparatus, observers achieved over 90% agreement across all reliability sessions. This procedure provides applied researchers with a simple, highly reliable, and adaptable observation tool for continuously and simultaneously monitoring the behaviors of one or more subjects.

DESCRIPTORS: measurement, observer agreement, observational technology, classroom behavior

Efforts to overcome methodological and reliability problems associated with naturalistic classroom observations continue to spur the development of novel observation systems and procedures. Faced with the task of monitoring 18 interactional responses between one teacher and each of two target students across seven instructional formats, a special coding sheet with accompanying observational procedures and equipment was devised. The frequency-within-interval system described below was developed as a response to three concerns raised in the literature: (a) the system was designed to enable rapid, reliable multiple-response and multiple-subject recording of naturally occur-

ring classroom behaviors, (b) its continuous format avoided many of the controversies associated with interval, time-sampling procedures for recording low-frequency behaviors (e.g., Powell, Martindale, Kulp, Martindale, & Bauman, 1977; Repp, Roberts, Slack, Repp, & Berkler, 1976), and (c) the system demonstrated that reliable, continuous multi-element event recording need not be expensive or complex.

The observational apparatus consisted of 11 recording sheets, a holder, and one tinted acetate cover. A 55-min class period was divided into 1-min intervals. (This interval length seemed most preferable to us for assessing low-frequency behaviors; see, for example, Sanson-Fisher, Poole, & Dunn, 1980.) A coding sheet was designed to allow 5 min of recording time on each sheet, as shown in Figure 1. The abscissa of each sheet was divided into two sections of five 1-min intervals—one section for each targeted student. Along the ordinate of each scoring sheet were spaced the interactional categories of concern, their qualitative subcategories, and categories for indicating the onset and duration of instructional activities occurring during each observational interval. This

Preparation of this article was supported by contract number 300-77-0494 from the Office of Special Education and Rehabilitative Services, United States Department of Education to The University of Kansas Institute for Research in Learning Disabilities. Portions of this paper were presented at the meeting of the Association for Behavior Analysis, Dearborn, 1980. The authors are particularly grateful for the assistance and comments of Sharie L. Myers, R. Don Horner, and P. Douglas Guess. Copies may be obtained from Thomas M. Skrtic, Department of Special Education, The University of Kansas, Lawrence, Kansas 66045.

LD										NLD									
			1	2	3	4	5				1	2	3	4	5				
APRVL	S	V						S	V										
		N																	
	A	V						A	V										
		N																	
DISAPRVL	S	V						S	V										
		N																	
	A	V						A	V										
		N																	
CALLDON	+							+											
	-							-											
	0							0											
VLNTRANSWR	+							+											
	-							-											
	0							0											
REQSTHELP	+							+											
	-							-											
	0							0											
Teach / Assist			1	2	3	4	5	T/A			1	2	3	4	5				
Manage		M						M											
Lecture		L						L											
Qus/Ans		Q/A						Q/A											
Discus		D						D											
Sml/Grp		S/G						S/G											
ldv/Sldy		I/S						I/S											
ldv/Rpt		I/R						I/R											
Other		O						O											

Fig. 1. Each recording sheet represents five 1-min observational units for learning disabled and nonlearning disabled students. Along the ordinate are spaced the 18 interactional categories that pertain directly to teacher responses to student behaviors. Teacher approval (APRVL) and disapproval (DISAPRVL) refer to verbal (V) and nonverbal (N) responses to student social (S) and academic (A) behavior. Positive (+), negative (-), and neutral (0) teacher responses to teacher-initiated (CALLDON) and student-initiated (VLNTR ANSWER) answers as well as student-initiated requests for assistance (REQST HELP) are coded in the next three sections. The frequency of teacher-initiated assistance (T/A) is recorded in the last section. Finally, a minute-by-minute record of classroom events allows for the assessment of interaction rates relative to behavioral opportunities. For a more complete description of the categories and the classroom events to which they pertain, readers are referred to Skrtic (Note 1).

latter information enabled researchers to assess appropriate levels for behavior and reinforcement frequencies relative to classroom opportunities. The 55-min class period was recorded in the 55 intervals contained in 11 scoring sheets.

To increase recording ease and accuracy, an acetate overlay was built to slide atop the recording sheets. First, a clear plexiglass holder, built to hold $8\frac{1}{2} \times 11$ -sized paper was purchased. (Transparency holder (THP4) is available from Behavior Research Company, Box 3351, Kansas City, KS 66103.) Because one side of the holder was unrimmed, papers and

acetates could be slid in and out of the holder easily. Next, as shown in Figure 2, a lightly tinted acetate was cut horizontally so that only two columns of intervals (one for each targeted student) were exposed for each category. By marking only in the open spaces on the recording sheet, observers could be more certain to avoid inaccurate tally marks. When the interval changed, cued by a stopwatch, the acetate was simply slid to the right to expose the next column of intervals for each target student. When the 5-min sheet was completed, it was removed from the holder exposing the next 5-min recording sheet. Finally, a hole was cut on the left

LD						NLD						
		1	2	3	4	5		1	2	3	4	5
APRVL	S	/					S	/				
	V						V					
	A						A					
DIS	S	/					S	/				
	V						V					
	A						A					
CALLD	+	/					+	/				
	-						-	/				
	0						0	/				
VLTNR	+	/					+	/				
	-						-	/				
	0						0	/				
REQST	+	/					+	/				
	-						-	/				
	0						0	/				
T/Asst						T/A						
Manager	M						M					
Lecture	L	/					L	/				
Q/A	Q/A						Q/A					
Dispt	D						D					
Sml/Grp	S/G						S/G					
Idr/Std	I/S						I/S					
Idr/Rpt	I/R						I/R					
Other	O						O					

Date _____

Sheet _____

Observer _____

Teacher Code: _____

LD Code: _____

NLD Code: _____

Total in class: _____

Present today: _____

Time Start: _____

Time End: _____

Total Minutes Observed: _____

Reliability: _____

Other Observer: _____

Fig. 2. Clear plexiglass holder with movable, tinted acetate exposing one column of intervals for each targeted student.

side of the acetate so that by depressing the thumb, observers could slide the acetate back and forth without disturbing the placement of the remaining recording sheets. Using this simple device, observers were able to record 18 separate interactional responses simultaneously for two students and their teacher reliably in 55 1-min intervals.

Circumventing the controversy surrounding reliability indices and calculations for low-rate behaviors (see, for example, Hartmann, 1977; Hawkins & Dotson, 1975; Hopkins & Hermann, 1977; and the first 11 articles appearing in the special issue dedicated to "Behavioral Assessment" in the *Journal of Applied Behavior Analysis*, 1979, **12**, 491-571), this study employed an interval-by-interval formula in which both scored and unscored intervals were counted in the calculation. Agreement was scored when the primary and reliability observers reported the same number of behaviors occurring in identical intervals while a disagreement was scored when either a different number of occurrences for each behavior was recorded per interval or when one observer recorded at least one occurrence and the other recorded none. The percentage of interobserver agreement was calculated using $\text{agreements} / (\text{agreements} + \text{disagreements}) \times 100$ for each of the 55 intervals across each of the 18 observational categories. This calculation procedure resulted in an agreement score for each category. These scores were then averaged across all reliability sessions resulting in an overall average agreement score for each category. Finally, these scores were averaged across all categories to yield an overall percent agreement score for the system.

Using this reliability procedure, in conjunction with a recording format that creates the opportunity for scoring multiple responses within each interval, establishes a highly stringent criterion for deriving interobserver agreement. Whereas in traditional interval recording, only the first occurrence of the behavior per interval is recorded (i.e., either a + or 0

is scored leaving a chance correct figure of 50% for each interval), a frequency-within-interval recording procedure requires every new behavioral occurrence to be scored. Accordingly, agreements are "more difficult" to obtain and the opportunity for chance-agreement (Hartmann, 1977) is made less probable.

The result is an increased demand for tally-by-tally accuracy. Although observers initially experienced difficulties maneuvering among the numerous response categories, once the acetate device was used, an average agreement score of 91% was obtained. Averages and ranges of the reliability calculations for each of the 18 interactional responses are presented in Table 1.

Table 1
Percent Agreement*

<i>Behavioral Category</i>	<i>Mean</i>	<i>Range</i>
Approval	84	
Social Verbal	87	(83-89)
Social Nonverbal	75	(70-88)
Academic Verbal	92	(90-95)
Academic Nonverbal	80	(76-82)
Disapproval	82	
Social Verbal	79	(75-83)
Social Nonverbal	70	(58-81)
Academic Verbal	92	(87-95)
Academic Nonverbal	88	(86-90)
Called On	93	
Positive	96	(94-98)
Negative	93	(92-93)
Neutral	89	(85-92)
Volunteers Answer	91	
Positive	93	(87-96)
Negative	91	(85-97)
Neutral	89	(88-91)
Requests Help	96	
Positive	97	(95-99)
Negative	95	(92-98)
Neutral	95	(92-97)
Teacher Assistance	99	(98-100)
TOTAL MEAN	91	(7)

*Calculated on four observers who each were checked against the other three for a total of 12 reliability sessions or 20% of a total of 60 observational sessions. All percent scores are rounded to the nearest whole number.

This agreement score is particularly significant in view of the number and diversity of dependent measures monitored and that, according to Kelly (1977), fewer than half of the observational studies appearing in the *Journal of Applied Behavior Analysis* between 1968 and 1975 reported reliability coefficients above .90. Although the acetate device may not be necessary for all observational situations, its contribution to increased interobserver agreement appeared beneficial to us because of the large number of behaviors recorded. Setting requirements prevented formal experimental comparisons between acetate and non-acetate conditions from being made. Nonetheless, informal observer-trainee reports confirmed increased recording capabilities and accuracy once the apparatus was implemented.

In conclusion, these results suggest that by combining the frequency-within-interval observational design with a simplified recording procedure (such as the one presented above), a practical and reliable observational format is made available for recording absolute behavioral frequencies across a large number of responses occurring in the natural environment. This is particularly useful since, according to Deitz (Note 2), the logistic difficulties associated with multiple-response and/or multiple-subject observation have been major justifications for using time sampling procedures. In view of the controversies surrounding discontinuous measurement, the procedures described here offer applied researchers a practical alternative for incorporating desirable features of several observational formats, thereby increasing their capacity to describe and analyze behavioral relations accurately.

REFERENCE NOTES

1. Skrtic, T. M. *The regular classroom interactions of learning disabled adolescents and their teachers*. (Research Report No. 8) Lawrence: The University of Kansas Institute for Research in Learning Disabilities, 1980. (Available from the University of Kansas Institute for Research in Learning Disabilities, 313 Carruth-O'Leary, The University of Kansas, Lawrence, Kansas 66045).
2. Deitz, S. M. *Some issues on the rationale and practice of measurement in applied behavior analysis*. Paper presented at the meeting of the Association for Behavior Analysis, Dearborn, 1980.

REFERENCES

- Hartmann, D. P. Considerations in the choice of interobserver reliability coefficients. *Journal of Applied Behavior Analysis*, 1977, **10**, 103-116.
- Hawkins, R. P., & Dotson, V. A. Reliability scores that delude: An Alice in Wonderland trip through the misleading characteristics of inter-observer agreement scores in interval recording. In E. Ramp & G. Semb (Eds.), *Behavior analysis: Areas of research and application*. Englewood Cliffs, N.J.: Prentice-Hall, 1975.
- Hopkins, B. L., & Hermann, J. A. Evaluating interobserver reliability of interval data. *Journal of Applied Behavior Analysis*, 1977, **10**, 121-126.
- 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.
- 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.
- Sanson-Fisher, R. W., Poole, A. D., & Dunn, J. An empirical method for determining an appropriate interval length for recording behavior. *Journal of Applied Behavior Analysis*, 1980, **13**, 493-500.

Received September 11, 1980

Final acceptance July 3, 1981