63 Best Practices in the Systematic Direct Observation of Student Behavior

John M. Hintze
University of Massachusetts at Amherst
Robert J. Volpe and Edward S. Shapiro
Lehigh University

OVERVIEW

Direct observation is one of the most widely used assessment procedures by school psychologists. In a survey of more than 1,000 school psychology practitioners, Wilson and Reschly (1996) found that of the 26 different types of assessment instruments listed across seven different assessment categories (e.g., ability/intelligence, social/emotional, visual/motor), structured observational methods were ranked highest in terms of frequency of use. Overall, practitioners report that they conduct more than 15 behavioral observations of student behavior during the course of a typical month.

Systematic direct observation refers to observation of behavior other than behavior that has been explicitly elicited by a predetermined and standardized set of stimuli (i.e., test behavior (Salvia & Ysseldyke, 2001)). School psychology practitioners commonly use both naturalistic and systematic direct approaches to observing student behavior. Briefly, naturalistic observation approaches refer to those observational procedures where the observer enters specific situations (e.g., a classroom) and observes all that is going on with no predetermined behaviors in mind. Here, the most common way of recording observations is to keep an anecdotal record of the behaviors that seem important to the observer. In summarizing the information, the observer provides a complete description of the many behaviors and the context in which they occurred.

In contrast, systematic direct approaches to behavioral observation are distinguished by five characteristics (Salvia & Ysseldyke, 2001). First, the goal of observation is to measure specific behaviors. Second, the behaviors being observed have been operationally defined a priori in a precise manner. Third, observations are conducted under standardized procedures and are highly objective in nature. Fourth, the times and places for observation are carefully selected and specified. Fifth, scoring and summarizing of data are standardized and do not vary from one observer to another.

Both naturalistic and systematic direct approaches to observing behavior have proven useful in developing theory and practice related to the assessment and intervention of student behavior. This chapter provides an overview of (a) naturalistic observational procedures, (b) systematic direct observational procedures, and (c) observational instruments as they related to the observation of student behavior in the classroom.

BASIC CONSIDERATIONS

Rationale for the Systematic Direct Observation of Behavior

With the reauthorization of the Individuals with Disabilities Education Act (IDEA, 1997), now more than ever it is absolutely incumbent that school psychologists adopt assessment and evaluation practices that

allow for gathering relevant functional information on student behavior and performance patterns. Specifically, sections 300.532(b) and (c) under the *Procedures for Evaluation and Determination of Eligibility* of IDEA indicate that a variety of assessment tools and strategies be used to gather relevant *functional* and developmental information about a child and that they be *validated* for the specific purpose for which they are used. Moreover, section 300.532(d) requires the use of tests and other evaluation materials that include those tailored to assess specific areas of educational need and not merely those that are designed to provide a single general intelligence quotient. Similarly, sections 300.533(a)(1)(ii) and (iii) indicate that as part of the initial evaluation for special education, evaluation data include current classroom-based assessments and *observations* conducted by teachers and related services providers. Once deemed eligible for special education, section 300.347(2)(ii) requires the specification of *measurable* annual goals in the development of the individualized education program that meet each of the child's needs as they pertain to the disability classification. The spirit of IDEA is also seen in the *Principles for Professional Ethics* and *Standards for the Provision of School Psychological Services* of the *Professional Conduct Manual* (National Association of School Psychologists, 2000). Here, school psychologists are required to use assessment techniques and instruments that have established validity and reliability and to adopt assessment practices that increase the likelihood of developing effective educational interventions.

The call for the increased use of systematic direct observational procedures and a movement toward more ecologically sensitive functional assessment practices contrast directly with those of a more traditional approach that have long been the standard in school psychological assessment practices. Table 1 provides a comparison between traditional and behavioral assessment across a number of key fundamentals. Central to the difference between the two is how they view the role and causes of behavior and the type of inferences that can be made from observing behavior. From a behavioral perspective, all behavior is a result of the dynamic interplay between an individual and the environment. Thus, observing a person's behavior is useful only to the extent that it can be used to summarize a person under a specific set of circumstances. Behavior is thought to be specific to a situation. For this reason, assessment needs to be ongoing and direct until a certain level of stability or generalizability is observed. This is in contrast to a more traditional approach whereby observed behavior is assumed to be enduring and stable across all situations and intrapsychic or residing within the individual. For this reason, assessment can be global and static; that is, occurring just once or enough to measure the internal construct.

Table 1. Comparison between behavioral and traditional assessment

	Behavioral	*Traditional*
Causes of behavior	Interaction between individual and environment	Enduring underlying states or traits
Role of behavior	Important as a sample of an individual's behavior in a specific situation	Behavior is important only to the extent that it represents the underlying trait.
Consistency of behavior	Behavior thought to be specific to a specific situation	Behavior assumed to be consistent across time and settings.
Use of data	To describe and quantify behavior under specific conditions. To select appropriate interventions. To monitor effectiveness of interventions.	To describe personality Functioning and etiology. To diagnose and classify. To make prognosis or prediction.
Level of inference	Low.	High.
Comparisons made	Both within and between individual.	Between individual.
Timing of assessment	Ongoing. Before, during, and after intervention.	Before and occasionally after intervention.

Note. Adapted from Goldfried and Kent (1972) and Hartmann, Roper, and Bradford (1979).

BEST PRACTICES IN SYSTEMATIC DIRECT OBSERVATION

Methods of Systematic Direct Observation

Naturalistic Observation

Naturalistic observation refers to the recording of behavioral events in their natural settings at the time they occur, using trained or impartial observers, where descriptions of behaviors require little if any inference beyond that which is observed and recorded (Jones, Reid, & Patterson, 1979). Similar to descriptive, narrative, or anecdotal observations, naturalistic observation employs the recording of salient behaviors and discriminative stimuli as they are observed chronologically in time. For example, a school psychologist might observe a student in the classroom and note the sequence of behaviors or activities that are hypothesized to be important or serve a reinforcing function in maintaining patterns of behavior. If a student is observed to refuse the request of a teacher, then this datum would be noted and recorded as a behavioral observation or event.

According to Wilson and Reschly (1996), naturalistic observation is the most frequently used type of direct observation, used nearly twice as often as other more systematic approaches to observation. The reason for this practice most likely lies in its ease in use and minimal training requirements. In practice, naturalistic observation usually takes the form or makes use of one of two recording methods. The easiest form involves simply observing and noting behaviors and events descriptively or anecdotally as they occur in the natural setting. Typically, such behaviors and events would be recorded in a written fashion, listed chronologically as they appeared in real time. Interpretation is limited to a descriptive account of the types of behaviors and events observed and their temporal ordering in time. Because interpretation is limited, such anecdotal and descriptive accounts cannot be used for making high-stakes decisions. In fact, one of the limitations posed by naturalistic observation is the inclination to "over interpret" the data, or make inferences regarding student behavioral patterns from a limited and unstandardized sample of behavior. Likewise, observers are prone to use confirmatory search strategies whereby increased attention is directed toward those behaviors that confirm the observers' original hypotheses. In both situations, bias in the selection and interpretation of the observation is present.

A second and equally popular method of naturalistic observation involves the use of A-B-C (Antecedent-Behavior-Consequence) observation and recording. Similar to anecdotal or descriptive observational techniques, here the practitioner makes careful note of those environmental arrangements, behaviors, or events occurring just before the behavior of concern is observed (i.e., the antecedent) and what behaviors or events are observed as a result of the behavior (i.e., the consequence). By using the example of refusing a teacher request above, the antecedent would most likely be some type of teacher request or directive, the behavior would be the act of refusing, and the consequence would most likely be a brief description of how the teacher responded to the student refusal. Figure 1 presents an example of the type of recording schedule that is typically used during A-B-C observations. This type of recording schedule can be easily constructed by dividing a sheet of paper into three columns, each of which corresponds to one of the three conditions (i.e., antecedents, behaviors, and consequences). Once constructed, the observer provides a brief narrative of each condition as they are observed in the natural setting. Although listed as A-B-C, the behavior column is generally completed first and then followed by the antecedents and consequences. The reasoning behind this rests on the understanding that without the presence of some salient recordable behavior, there is little use in recording antecedents and consequences. While numerous behaviors will obviously occur, only those of clinical importance are noted. Given this, it is often helpful to specify beforehand those behaviors that will be recorded during the observation period.

Naturalistic observation techniques have become increasingly popular owing in part to their utility as part of an overall functional assessment strategy. Here, descriptive or anecdotal and A-B-C analyses are used as a preliminary step in data collection and serve the purpose of developing testable hypotheses regarding the motivation and maintenance of student behavior. For example, O'Neill et al. (1997) use naturalistic observation strategies as the first step in an overall functional assessment procedure that allows the clinician to operationally define target behaviors and formulate preliminary hypotheses regarding the function of behavior. Once salient behaviors and environmental events are observed naturalistically,

Figure 1. Example of naturalistic observation using A-B-C observation and recording

A-B-C Observation and Recording Sheet

Antecedent	Behavior	Consequence
Teacher asks students to take out paper and pencil.	Target student does not take out paper and pencil. Plays with toy car on desk instead.	Teacher reprimands target student.
Teacher takes paper and pencil out of target student's desk.	Target student pushes paper and pencil off desk and onto floor. Target student puts head on desk with arms folded around head.	Teacher removes request. Teacher picks up paper and pencil and places on desk. Tells target student he can begin work when he calms down.
Teacher continues lesson with rest of class.	Target student begins to play with car on desk again.	Teacher stops lesson and takes car away from target student.
Teacher directs student to put name on top of paper.	Target student kicks desk away. Sits in chair with arms folded across chest.	Teacher ignores target student.

they are observed systematically by using time-sampling procedures.

The advantages of using naturalistic observation procedures to identify target behaviors in this manner are that (a) their importance or social validity of behaviors can be assessed by noting the frequency of their occurrence in the natural setting, (b) their relationship to important environmental antecedents and consequences can be examined systematically, (c) data can be used to develop testable hypotheses regarding the function of behavior, and (d) the data gathered serve an important step in allowing the clinician to make decisions regarding the function of behavior rather than focusing on topographical and descriptive accounts of what is observed. As was previously noted, one of the restrictions of natural descriptive accounts of behavior is their limited utility in decision making. Because the data gathered are purely descriptive in nature, decisions are restricted to summary statements of what was observed and little else. For the most part, such data prove limited in a more problem-solving assessment orientation. Nonetheless, naturalistic descriptive accounts can be extremely useful when used as a preliminary step in a problem-solving functional assessment paradigm. Here the data gathered form the basis for developing initial hypotheses that can be subsequently observed in a more systematic fashion. The next two sections describe such systematic observational methods and the way in which practitioners can use such methods to quantify behavior from an ecological perspective.

Observational Procedures

Observational procedures refer to a set of techniques that school psychologists can use to quantify behavior along one or multiple dimensions (Kratochwill, Alper, & Cancelli, 1980). For example, a school psychologist might be interested in assessing the frequency in which a referred student is out of his seat. After operationally defining the behavior of interest, the target child would be directly observed for a specified length of time with the number of times he got out of his seat noted. Additionally, the length of time spent out of his seat for each occurrence might be noted. If out-of-seat behavior does not appear to be the main issue of concern, then another behavior can be specified and observed in a similar manner. Likewise, multiple behaviors can be identified and observed concurrently. Although the example is oversimplified, here the advantages of using observational

procedures are that they are flexible and can be tailored to suit the specific needs of the assessment situation.

Measuring and Recording Behavior Systematically. There are various types of data that can be collected during systematic direct observation. A workable definition of a target behavior is one that provides an accurate description of the behavior that clearly defines the parameters of its existence and nonexistence (Heward, 1987). As such, constructs and reifications do not lend themselves well to direct observation. For example, raising one's hand to be called on is an observable and measurable behavior. Behaving "off the wall" is not something that can be directly observed. In developing explicit behavioral definitions Hawkins and Dobes (1977) offer the following suggestions:

1. The definition should be objective, referring only to observable characteristics of the behavior and environment.

2. A workable behavioral definition is readable and unambiguous such that an experienced observer could read it and readily paraphrase it accurately.

3. The definition should be complete, delineating the boundaries of what is to be included as an instance of the behavior and what is to be considered a noninstance of the behavior.

As such, operational definitions must be objective, ensuring that specific instances of the defined target behavior can be readily observed and recorded. In addition, an operational definition is a technological definition that enables others to use and replicate it (Baer, Wolf, & Risely, 1968). A complete operational definition identifies what is not the target behavior and aids observers in discriminating the target behavior from similar responses.

Once behavior is defined, the calibration of the operational definition is determined by the nature of the data; that is, the frequency of its occurrence and the particular interests of the observer (Hintze & Shapiro, 1995). In addition, practical considerations such as the availability of observers, the amount of time the student is accessible, or any combination of these factors, all dictate the type of data collected. Because each of these data may yield different results, the method of data collection must be clearly understood.

Frequency or Event Recording: The type of data known as frequency or event recording involves counting the number of occurrences of behavior observed during a specified time period. When the time periods in which the behavior is counted vary, frequencies are converted to rates of behavior per unit of time. For example, an observer may report that a target child raised a hand at an average rate of one time per minute during three separate observations conducted over the course of three days, even though the actual duration of each observation period varied. By using rate of behavior allows the practitioner to compare the occurrence of behavior across observational periods (Shapiro, 1987).

Frequency or event recording is most useful when observing behaviors that have a discrete beginning and ending. Throwing paper airplanes, hitting, and the raising of a hand are all examples of such behaviors. Behaviors that are continuous or persist for longer durations sometimes prove difficult to observe by using event recording. For example, pencil tapping, talking, or on-task behavior would be difficult to observe by using such an observational system. As with all recording schedules, a very clear operational definition of the target behavior helps ensure that accurate frequency data are being collected. In instances such as pencil tapping and talking, episodes of the behavior may be defined by using a time dimension as well as a topographical description of the behavior. For example, pencil tapping may be defined as the continuous tapping of the pencil against a physical surface that produces audible noise for at least five consecutive seconds. In this way, the frequency count for pencil tapping is easily recorded rather than the actual number of times the pencil is tapped against a physical surface.

A second consideration when using event recording has to do with the actual length of time each episode of the behavior occurs for. Generally speaking, each episode of the behavior should take approximately the same amount of time for each instance of the behavior (Barton & Ascione, 1984). For example, if an observer is observing occurrences of "noncompliance," the length of time might vary widely from a simple refusal to follow a direction to a knock-down

drag-out tantrum that lasts 20 minutes. Because each episode would be coded as the occurrence of one act of noncompliance, each would be weighted equally in terms of the way in which it was quantified. Obviously, the variability in the duration of each episode is lost. For this reason, when the length of time a behavior occurs is important, event recording might not be the most appropriate recording schedule. It may, however, be combined with other recording options (e.g., duration) to capture both frequency and time dimensions of a particular behavior.

Another instance in which frequency or event recording is particularly useful is when behavior occurs at a relatively low rate. Such behaviors often occur infrequently but are of interest because of their intensity or seriousness. For example, running out of the classroom may occur once or twice a day, but may represent significant difficulties for the student. The advantage here is that with low frequency behaviors, observational periods can be continuous and designed in a fashion to be relatively unobtrusive and at low time and cost to the observer. The disadvantage, however, is that if any instance of the behavior is not observed and recorded, the reliability of the observed data is sacrificed.

As can be seen, the methods for frequency or event recording are quite varied. Commonly, the frequency in which behavior occurs is recorded in a written format (e.g., tallies on a piece of paper) with the beginning and ending time of the observational session noted. In addition to simple paper-and-pencil recording, hand-held mechanical recorders such as those used to keep track of attendance at a social function, golf wrist counters, or a wrist abacus can be used. In the end, any device capable of keeping a cumulative frequency count can be used to perform event recording. Figure 2 presents an example of a paper-and-pencil frequency recording schedule.

Duration Recording: Another type of behavioral response that can be recorded is the duration of the behavior. Duration measures may be very helpful with certain types of school-related behaviors. Studying, temper tantrums, social isolation, and aggressive outbursts are good examples of behaviors in which duration is generally important. Duration is also appropriate in cases where changing the duration of the behavior is an important target for intervention. As in the case of event recording, behaviors that have discrete beginnings and endings may be assessed in the length of time the behavior lasts.

The duration of a behavior is usually standardized in two ways (Salvia & Ysseldyke, 2001). First, the average duration of each occurrence may be computed. For example, Gary got up and out of his seat five times during a 20-minute observation. The duration of each episode was 2, 3, 4, 5, and 6 minutes, respectively. In this case the average duration was 4 minutes (i.e., [2 + 3 + 4 + 5 + 6]/5). Second, the total duration may be computed. In this same example, Gary was out of his seat a total of 20 minutes.

The most precise nonautomated instrument for collecting duration data is a stopwatch. The procedure for recording total duration with a stopwatch is to start

Figure 2. Example of frequency recording. Target behaviors include getting out of seat (OS), calling out (CAL), and teacher redirections (TR)

Frequency Observation and Recording Sheet

Date: May 13, 2000

Observer: A. VanDelay

Time	OS	CAL	TR
9:00 a.m.–10:00 a.m.	X X X	XXXXX	XXXXXXX
10:00 a.m.–11:00 a.m.	X	XXX	XXXX
11:00 a.m.–12.00 p.m.	XX	X	XXXXXX

the stopwatch as the behavior begins and stop the timing at the end of the episode. Without resetting the stopwatch, the observer starts the stopwatch again at the beginning of the second occurrence of the behavior and stops timing at the end of the second episode. The observer continues to accumulate the durations of time in this fashion until the end of the observation period and then transfers the total duration of time showing on the stopwatch to a record sheet. Figure 3 presents an example of duration recording for thumb sucking.

When observation sessions are consistent in length (e.g., 20 minutes), total duration can be compared across sessions. However, when observation sessions vary in length, a percent ratio of total duration to observation length must be computed before comparisons across observational sessions can be made. For example, if the total duration of Gary's out-of-seat was 10 minutes in each of three observational sessions but the observational sessions varied from 20-, to 30-, and 40-minute time frames, then the total duration percent that he was out of his seat would be noted as 50, 33, and 25%, respectively.

The procedure for recording duration per occurrence with a stopwatch is to start the stopwatch as the behavior begins and stop the timing at the end of the episode. The observer transfers the duration of time showing on the stopwatch to a data sheet and resets the watch. The stopwatch is started again at the beginning of the second occurrence of the behavior and is stopped at the end. The duration of time for each episode is then summed and divided by the number of occurrences yielding an average duration per episode.

Latency Recording: Latency recording is the measurement of elapsed time between the onset of a stimulus or signal (e.g., a verbal directive) and the initiation of a specified behavior (Cooper, 1987). Latency recording should be used when the major concern is the length of time between an opportunity to elicit a behavior (e.g., after the presentation of the verbal directive) and the actual time it takes to begin performing the behavior. In this case, the response latency would be the length of time between the end of the teacher's directive and initiation of the student's compliance with the directive. As with event and duration recording, when latency is assessed, both the signal and the behavior of interest must have discrete beginnings. The procedure for latency recording is similar to that for duration recording. The observer uses a stopwatch and begins timing immediately after the signal or stimulus is delivered and stops timing at the instant the target behavior is initiated. For each

Figure 3. Example of duration recording for thumb sucking

Duration Observation and Recording Sheet

Student: Alex Observer: B. Matthews
Behavior: Thumb Sucking
Date: August 5, 2000

Time start: 10:10 Time stop: 10:30

Thumb sucking (separate incidents)	Elapsed time per episode (in minutes and seconds)
1	1' 17"
2	6' 42"
3	2' 11"
4	7' 26"
5	52"
	Total: 18' 28"

Average duration per episode, 3' 42"

episode, the summary datum is the time lapse between the signal and the behavior. The actual time that it takes to complete the target behavior *does not* figure into latency recording. This is because behaviors can vary widely in the time it takes to complete them. For example, the directive of "pick up your pencil and put your name on the top of the paper" and "pick up your pencil and complete the *New York Times* Sunday crossword puzzle" would each be expected to have similar latencies. Obviously, the duration of these two behaviors would vary significantly. If completion times of these two behaviors were of primary interest, then duration would be an appropriate recording schedule. Moreover, like duration, both average and total latency can be used to summarize observed behaviors.

Time-Sampling Interval Recording: Whereas frequency, duration, and latency recording are able to accurately capture the dimensions of behavior each represents, oftentimes it is difficult to use any one of the recording schedules because of practical or measurement concerns. Issues dealing with availability of observers, lack of time, or operational issues (e.g., complications in determining exact beginning and ending of behavior) all contribute to the difficulty in observing behavior continuously. The essential characteristics of time sampling interval recording involve selecting a time period for observation, dividing the observational period into a number of equal intervals, and recording whether or not a specified target behavior has occurred during each interval (Merrell, 1999). For example, a 30-minute observation period might be broken down into 120 15-second intervals. Within each interval, the presence or absence of one or multiple behaviors might be assessed. Presence or absence of the behavior will be determined by one of three, or any combination, of three recording schedules discussed below (i.e., whole, partial, or momentary time sampling recording).

Also, unlike event, duration, and latency recording that provide exact frequency or time dimensions of observed behavior, time sampling interval recording provides only approximates for the behavior as it occurs. That being so, for situations where the exact number of occurrences or time spent engaged in the behavior, or latency of the behavior is of concern, time sampling interval recording might not be the best option. Nonetheless, time sampling interval recording provides an excellent alternative when conditions warrant observing a number of behaviors simultaneously, or for behaviors that occur at a moderate to high rate or steady state.

Whole-Interval Recording: In whole interval recording, the target behavior is scored as having occurred only when it is present throughout the entire interval (intervals are scored usually with a plus sign or other mark to indicate the presence of the behavior, and empty intervals generally denote the absence of the behavior). Since the behavior must be present for the entire interval, whole-interval recording lends itself quite well to behaviors that are continuous or intervals that are of a short duration (Shapiro & Skinner, 1990). One of the drawbacks of whole-interval recording, however, is that has a tendency to underestimate the presence of the behavior in real time. Consider for example, a whole-interval recording schedule where each interval is 15 seconds long. If, for example, off-task behavior were our target behavior and it was observed for 13 of the 15 seconds during the interval, then the interval would not be scored for the presence of off-task behavior since it did not occur for the entire 15-second interval. In essence, it would appear as if the target student was on-task for the entire 15 seconds. Because of this, whole-interval recording is well suited for behaviors targeted for increase through intervention efforts (Sulzer-Azaroff & Mayer, 1991).

Partial-Interval Recording: In contrast to whole-interval recording, with partial-interval recording an occurrence of the behavior is scored if it occurs during any part of the interval. Thus, if a behavior begins before the interval begins and ends within the interval, then an occurrence is scored. Similarly, if the behavior starts after the beginning of the interval, then an occurrence is scored. Finally, if multiple occurrences of the behavior are observed within the same interval, then the interval is simply scored as if the behavior occurred once. Again, in comparison to whole-interval recording, partial-interval recording is a good choice for behaviors that occur at a relatively low rate or behaviors of somewhat inconsistent duration. Also, partial-interval recording tends to overestimate the actual occurrence of the behavior. By using the example above, if the target student were observed to be off-task for only 2 seconds of the 15-second

interval, then the interval would be scored for the presence of the behavior as if it occurred for the entire 15-second interval. Because of this, partial-interval recording is well suited for behaviors targeted for decrease through intervention efforts (Sulzer-Azaroff & Mayer, 1991).

Momentary Time-Sampling: Finally, with momentary time-sampling, a behavior is scored as present or absent only during the moment that a timed interval begins. With this technique, the observer notes either the presence or absence of the behavior at a brief instant during the interval. By using the above example, the target student would be considered off-task if at the moment of observation (e.g., very beginning of a 15-second interval) he was observed to be off-task, irrespective of any behavior observed during the rest of the interval. Salvia and Hughes (1990) have summarized a number of studies investigating the accuracy of these time-sampling procedures. As was previously noted, both whole-interval and partial-interval sampling procedures provide inaccurate estimates of the behavior in real time. Momentary time-sampling, although based on the smallest sample of behavior, provides the least biased estimate of behavior as it actually occurs (Suen & Ary, 1989).

Observational Instruments

In contrast to observational procedures, observational instruments have been developed to assess a specific range of behaviors. For example, a school psychologist might choose to use an observational instrument designed specifically to quantify the percentage of time a student is academically engaged or off-task or the frequency with which a teacher provides directives, provides opportunities to respond, or positively reinforces student efforts. Unlike more generic observational procedures, however, the flexibility of observational instruments is typically limited. With standardized administration and scoring procedures, practitioners cannot alter the operational definitions to suit their individual needs. However, because they have been developed with a specific purpose in mind, observational instruments tend to provide a more detailed account of a student's behavioral pattern across a variety of behaviors of common interest to the observer. What is lost in flexibility is gained in breadth of behaviors observed. While school psychologists could certainly develop their own observational instruments individually tailored to particular behavioral constructs, the time spent in development may be cost and labor prohibitive.

Nonetheless, the use of observational instruments continues to gain in popularity. The increased interest in such techniques is due in part to the optimization of laptop computers and hand-held data recording devices. What follows are two examples of published systematic observation codes. The Behavior Observation of Students in Schools (BOSS) (Shapiro, 1996) can be a useful part of an academic assessment in classroom settings. It may also be used in assessments of disruptive behavior, if aggressive behavior is not of principal concern. The Attention Deficit Hyperactivity Disorder School Observation Code (ADHDSOC) (Gadow, Sprafkin, & Nolan, 1996) was specifically designed to assess disruptive child behavior across a number of school settings and includes various measures of aggressive child behavior. These specific codes were selected for review owing to their ease of use, and their differential purposes.

Behavior Observation of Students in Schools. The BOSS is a useful observation code for assessing child academic behavior in the classroom environment. It is a relatively easy code to learn and use. In simple terms, the BOSS assesses levels of "on-task" and "off-task" behavior. The amount of time children are engaged in academic tasks appears to be an important instructional variable (see Gettinger, 1986, for a review). Although several existing observation codes include behaviors that represent academic engagement, the BOSS is unique in that it divides engagement into two categories: (a) active engagement (e.g., writing, raising hand, answering a question) and (b) passive engagement (e.g., looking at a worksheet, listening to teacher directions). Furthermore, off-task behaviors are assorted into three categories: (a) off-task motor (e.g., out-of-seat, fidgeting, playing with pencil), (b) off-task verbal (e.g., calling out, talking to a peer when prohibited), and (c) off-task passive (e.g., looking around, looking out the window). Finally, the BOSS also includes a measure of teacher directed instruction (TDI), which provides an estimate of the amount of time the teacher is engaged in direct instruction. For example, the TDI category would be scored as present if the teacher were lecturing to the class and absent if the teacher were sitting at the desk correcting papers.

The BOSS is administered in 15-second intervals for a period of at least 15 minutes. The on-task behaviors (i.e., active and passive engagement) are scored at the beginning of each interval by using momentary time-sampling. For the remainder of each interval, the off-task behaviors (i.e., motor, verbal, and passive) are noted by using partial interval scoring. In addition, during every fifth interval, rather than observing the target student, the behavior of a randomly selected peer is observed and noted. In doing so, comparisons can be made between the target student and a peer composite that represents "typical" behavior during the observational period. Once the observation is completed, scoring summaries are computed for on- and off-task behaviors of both the target student and the peer composite, as well as an overall estimate of how much time the teacher was engaged in direct instruction.

Attention Deficit Hyperactivity Disorder School Observation Code. According to its authors, the ADHDSOC was developed as both a screening measure and as a tool for evaluating the effects of interventions for children diagnosed with attention-deficit/hyperactivity and related disorders (e.g., oppositional defiant disorder). The ADHDSOC can be used across a number of school settings (e.g., classroom, lunchroom, playground). For example, the following seven behavior categories are scored in classroom situations: (a) interference (e.g., target student calls out when it is inappropriate to do so), (b) motor movement (e.g., target student gets out of seat without permission, (c) noncompliance (e.g., target student ignores verbal direction from teacher or aide), (d) verbal aggression (e.g., target student curses), (e) symbolic aggression (e.g., target student takes another student's pencil), (f) object aggression (e.g., target student kicks chair or desk), and (g) off-task (e.g., target student stops working on assignment and stares out the window). Though the various aggression scores (i.e., verbal, symbolic, object) may be coded individually, the authors suggest collapsing them into a single category termed "nonphysical aggression." For observations conducted in the lunchroom or on the playground the following categories are coded: (a) appropriate social behavior (e.g., target student is observed talking to another child appropriately), (b) noncompliance, (c) nonphysical aggression (this includes both object and symbolic aggression), (d) verbal aggression, and (e) physical aggression (e.g., target student trips another child). Across all settings, target behaviors are scored on a partial interval every 15 seconds.

For diagnostic purposes, the authors recommend selecting three or four average peers to observe for comparison. Selected peers and the target student are then observed in alternating 1-minute intervals. The observer rotates through each peer until all have been observed, and then returns to the initial peer. Specific guidelines for collecting observation data as well as statistical guidelines for comparing the scores of the target student to the peer composite are provided. In addition, guidelines for incorporating the ADHDSOC into treatment evaluation procedures are also presented. Validation of the ADHDSOC is presented in several studies of school-based medication evaluation studies (see Gadow, 1993; Gadow, Nolan, Poalicelli, & Sprafkin, 1991).

Both the BOSS and the ADHDSOC are relatively simple school-based observation codes to learn and use. They offer the opportunity to assess both within-student features of behavior (e.g., across various settings that vary with respect to task demands), and between-student features of behavior (e.g., target student as compared to peer composite). These are only two of the many coding systems available to school psychologists. Other available instruments are listed in Table 2.

General Issues in Systematic Direct Observation

Considerations in Selection of Target Behaviors

Considering the Social Significance of the Behavior. In daily practice, it is not uncommon for teachers and other school personnel to describe a litany of target behaviors that present as possible candidates for change. As most school psychologists are all too well aware, to target each for observation and intervention is likely to prove time and cost inefficient in the long run. As such, assessment information gathered before systematically observing behavior (e.g., interviews) is absolutely crucial to designing a sound observation strategy. In doing so, the school psychologist must determine which elements of a student's behavioral repertoire might serve as socially significant and ecologically valid target behaviors (Hintze & Shapiro, 1995).

Table 2. Other examples of school-based observation protocols

Code	*Published*	*Age group*	*Setting*	*Recording schedule*	*Length of intervals, seconds*
Systematic Screening of Behavior Disorders: Peer Social Behavior Code (Walker & Severson, 1990)	Sopris West (pre-school adaptation available[a])	Elementary	Playground	• Partial interval	10
The Preschool Observation Code (Bramlett, 1993; Bramlett & Barnett, 1993)	See references	Preschool	Classroom	• States (momentary time-sample) • Events (frequency of response)	30
State-Event Classroom Observation System (Saudargas, 1997)	Available from author	Elementary	Classroom	• States (momentary time-sample) • Events (frequency of response)	15

[a]Sinclair, Del'Homme, and Gonzalez (1993).

Procedures that affect behavior of one or more persons in ways that increase the probability or magnitude of benefits for any one or all of the persons involved are considered to be socially valid targets for behavior change (Hawkins, 1986). Teaching children to read, pay attention, or make friends with peers are likely to be socially valid, because these efforts generally increase benefits and decrease costs for both the targeted individual and the rest of society (Hawkins, 1991). As such, a goal, outcome, or procedure is valid only to the extent that choosing it as a target for change improves the benefit-to-cost ratio for the individual, for others, or both. A consumer's or professional's opinion about a targeted behavior for change is only valid to the extent that it is consistent with such improved benefit-to-cost ratio (Hawkins, 1991).

Although conceptually the notion of social validity makes sense, in practice it may be difficult to operationalize. At least one objective validation strategy is to use normative data from same age peers as the target student to identify which behaviors are likely to be adaptive, and the extent to which such behaviors are expected to be mastered. For example, observing the rates of academic engaged time of peers of the target child can help establish whether or not a discrepancy is present, in addition to the magnitude of the discrepancy and goals for change (Hawkins, 1991). Similarly, observing those behaviors considered to be associated with targeted behaviors (e.g., escape or avoidance) from a normative reference group may assist in determining which behaviors are of critical importance for success (Hawkins, 1991). Finally, the best validation of which behavior is most adaptive is to test experimentally the outcomes produced by different behaviors and different levels of their performance (Hintze & Eckert, 2000). By definition, those strategies that yield the greatest benefit at the least cost are the most adaptive (Hawkins, 1986).

Prioritizing Possible Target Behaviors. In many assessment situations, decisions must be made regarding the relative priority of possible target behaviors (Hintze & Shapiro, 1995). In a review of the research, Nelson and Hayes (1979) offer four suggestions that can help guide the practitioner faced with a multitude of potential target behaviors:

1. Alter the behavior that is most irritating to the person who has identified the problem (Tharp & Wetzel, 1969).

2. Alter a behavior that may be relatively easy to change (O'Leary, 1972).

3. Alter behaviors that will produce beneficial response generalization (Stokes & Baer, 1977).

4. When behaviors exist as part of a longer response chain, alter the behaviors at the beginning of the chain (Angle, Hay, Hay, & Ellinwood, 1977).

In addition, Hawkins (1986) suggests that:

1. Targeted behaviors should be those that represent "keystone" or pivotal behaviors within a behavioral response hierarchy.

2. Behaviors that have a "general utility" should be considered prior to those with highly specific functions.

3. The construction or acquisition of behavioral response repertoires should take precedence over the pure elimination of specific behaviors.

4. Behaviors that gain a student access to reinforcement in the natural environment should be given high priority.

5. Student choice should be considered when selecting possible target behaviors.

Antecedents and Consequences of Behavior. Both naturalistic observation and systematic direct observation strategies allow the observer to examine interdependencies among functional antecedents, behaviors, and consequences (Alessi, 1988). Careful examination of such functional response chains allow practitioners to develop hypotheses such as:

1. Are there avoidance, escape, or termination behaviors evident, contingent upon requests or demands made by the teacher?

2. Does the magnitude of the behavior change as a function of varying task demands?

3. Does the target behavior lead to accessing social attention or preferred tangibles or activities?

4. Are there particular setting or temporal characteristics associated with the target behavior?

Once developed, such hypotheses can be tested experimentally in a more controlled experimental analysis of the behavior (Iwata, Dorsey, Slifer, Bauman, & Richman, 1982). By using a brief functional analysis (Wacker et al., 1990), hypothesized behavioral functions can be assessed using predetermined analogue assessment strategies. Most assessments include alone, escape, and attention conditions. During the first phase of the analysis, reinforcement (e.g., social attention, withdrawal of a demand) is provided contingent on the occurrence of the target behavior. During the second phase of the brief functional analysis each condition is replicated; however, the reinforcement contingency (i.e., escape, attention) is provided for appropriate behavior rather than inappropriate behavior. By using such a methodology allows the practitioner to validate specific A-B-C chains and target specific environmental contingencies for ongoing assessment and intervention.

SUMMARY

Significant effort has been made over the past decade to shift the role of the school psychologist from one of problem identifier to one of problem solver (Deno, 1995). With a focus on hypothesis testing and scientific accountability, the school psychologist as problem solver seeks to use assessment instruments that provide clear links between assessment and intervention. Systematic direct observation provides one of the most useful strategies for accomplishing this goal. With a focus on socially significant and meaningful behavior change, systematic direct observation changes the approach of the observer from passive to active and reflects technological advances in both the study of human behavior and how behavior is recorded and summarized. Moreover, systematic direct observation procedures are in line with the reauthorization of IDEA and current standards for test use among school psychologists.

As with any new skill, it takes time to become fluent in the use of systematic observations. School psychologists should not be fooled by its apparent simplicity and should expect to devote as much time in training to use such procedures as they typically would in learning any standardized assessment instrument. With continued use, however, school psychologists will find systematic direct observation to be a crucial component of just about any of the services they offer.

REFERENCES

Alessi, G. (1988). Direct observation methods for emotional/behavior problems. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavior assessment in schools: Conceptual foundations and practical applications* (pp. 14–75). New York: Guilford.

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.

Baer, D. M., Wolf, M. M., & Risely, T. (1968). Current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Barton, E. J., & Ascione, F. R. (1984). Direct observation. In T. H. Ollendick & M. Herson (Eds.), *Child behavior assessment: Principles and procedures* (pp. 166–194). New York: Pergamon.

Bramlett, R. K. (1993). *The Preschool Observation Code*. Unpublished manuscript, University of Central Kansas.

Bramlett, R. K., & Barnett, D. W. (1993). The development of a direct observation code for use in preschool settings. *School Psychology Review, 22,* 49–62.

Cooper, J. O. (1987). Measuring and recording behavior. In J. O. Cooper, T. E. Heron, & W. L. Heward (Eds.), *Applied behavior analysis* (pp. 59–80). Columbus, OH: Merrill.

Deno, S. L. (1995). The school psychologist as problem solver. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 471–484). Washington, DC: National Association of School Psychologists.

Gadow, K. D. (1993). A school-based medication evaluation program. In J. L. Matson (Ed.), *Handbook of hyperactivity in children* (pp. 186–219). Boston: Allyn & Bacon.

Gadow, K. D., Nolan, E. E., Poalicelli, L. M., & Sprafkin, J. (1991). A procedure for assessing the effects of methylphenidate on hyperactive children in public school settings. *Journal of Clinical Child Psychology, 20,* 268–276.

Gadow, K. D., Sprafkin, J., & Nolan, E. E. (1996). *ADHD School Observation Code*. Stony Brook, NY: Checkmate Plus.

Gettinger, M. (1986). Issues and trends in academic engaged time of students. *Special Services in the Schools, 2,* 1–17.

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

Hartmann, D. P., Roper, B. L., & Bradford, D. C. (1979). Some relationships between behavioral and traditional assessment. *Journal of Behavioral Assessment, 1,* 3–21.

Hawkins, R. P. (1986). Selection of target behaviors. In R. O. Nelson & S. C. Hayes (Eds.), *Conceptual foundations of behavioral assessment* (pp. 331–385). New York: Guilford.

Hawkins, R. P. (1991). Is social validity what we are interested in? Argument for a functional approach. *Journal of Applied Behavior Analysis, 24,* 205–213.

Hawkins, R. P., & Dobes, R. W. (1977). Behavioral definitions in applied behavior analysis: Explicit or implicit? In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, method, and application* (pp. 167–188). Hillsdale, NJ: Erlbaum.

Heward, W. L. (1987). Selecting and defining target behavior. In J. O. Cooper, T. E. Heron, & W. L. Heward (Eds.), *Applied behavior analysis* (pp. 36–58). Columbus: Merrill.

Hintze, J. M., & Eckert, T. L. (2000). The use of functional assessment and analysis strategies to reduce the noncompliant behavior of a child with autism. *Proven Practice: Prevention & Remediation Solutions for Schools, 3,* 9–15.

Hintze, J. M., & Shapiro, E. S. (1995). Systematic observation of classroom behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp.

651–660). Washington, DC: National Association of School Psychologists.

Individuals with Disabilities Education Act: Amendments of 1997 (PL 105-17). 20 USC Chapter 33, Sections 1400 et seq. (Statute).

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities, 2,* 3–20.

Jones, R. R., Reid, J. B., & Patterson, G. R. (1979). Naturalistic observation in clinical assessment. In P. McReynolds (Ed.), *Advances in psychological assessment*, *3*, 42–95. San Francisco: Jossey-Bass.

Kratochwill, T. R., Alper, S., & Cancelli, A. A. (1980). Nondiscriminatory assessment in psychology and education. In L. Mann & D. A. Sabatino (Eds.), *Fourth review of special education* (pp. 229–286). New York: Grune & Stratton.

Merrell, K. W. (1999). *Behavioral, social, and emotional assessment of children and adolescents.* Mahwah, NJ: Erlbaum.

National Association of School Psychologists. (2000). *Professional conduct manual* (4th ed.). Bethesda, MD: Author.

Nelson, R. O., & Hayes, S. C. (1979). Some current dimensions of behavioral assessment. *Behavioral Assessment, 1,* 1–16.

O'Leary, K. D. (1972). The assessment of psychopathology in children. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (pp. 234–272). New York: John Wiley.

O'Neill, R. E., Horner, R. H., Albin, R. W., Sprague, J. R., Storey, K., & Newton, J. S. (1997). *Functional assessment and program development for problem behavior: A practical handbook* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Salvia, J., & Hughes, C. (1990). *Curriculum-based assessment: Testing what is taught.* New York: Macmillan.

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Princeton, NJ: Houghton Mifflin.

Saudargas, R. A. (1997). *State-Event Classroom Observation System (SECOS).* Unpublished manuscript, University of Tennessee-Knoxville.

Shapiro, E. S. (1987). *Behavioral assessment in school psychology.* Hillsdale, NJ: Erlbaum.

Shapiro, E. S. (1996). *Academic skills problems workbook.* New York: Guilford.

Shapiro, E. S., & Skinner, C. H. (1990). Best practices in observation and ecological assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology II* (pp. 507–518). Washington, DC: National Association of School Psychologists.

Sinclair, E., Del'Homme, M., & Gonzalez, M. (1993). Systematic screening for preschool behavioral disorders. *Behavioral Disorders, 18,* 177–188.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 19,* 349–367.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data.* Hillsdale, NJ: Erlbaum.

Sulzer-Azaroff, B., & Mayer, G. R. (1991). *Behavior analysis for lasting change* (2nd ed.). New York: Harcourt Brace.

Tharp, R. G., & Wetzel, R. J. (1969). *Behavior modification in the natural environment.* New York: Academic.

Wacker, D., Steege, M., Northup, J., Reimers, T., Berg, W., & Sasso, G. (1990). Use of functional analysis and acceptability measures to assess and treat severe behavior problems: An outpatient model. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 349–359). Pacific Grove, CA: Brooks/ Cole.

Walker, H. M., & Severson, H. H. (1990). *Systematic screening for behavior disorders: Observer training manual.* Longmont, CO: Sopris West.

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25,* 9–23.