Chapter 8

Behavioral Observations

Chapter Questions

- Why have direct observations often been considered the standard by which other assessment techniques are judged?
- What are some of the characteristics of behavioral observations that limit their usefulness in many clinical situations?
- What are the basic components of observational systems?
- What are some examples of observational systems that might be used as part of a clinical assessment of children and adolescents?

Direct observation of a child's or adolescent's overt behavior has held a revered status in the clinical assessment of youth. Frequently, the validity of other methods of assessment is judged by their correspondence with direct observations of behavior. In fact, behavioral observation is often viewed as synonymous with the practice of behavioral assessment (Shapiro & Skinner, 1990). There are two primary reasons for this importance provided to direct observations. First, as the term *direct* implies, observations of behavior are not filtered through the perceptions of some informant. Instead, the behaviors of the child are observed directly. As we have discussed in the chapters on behavior rating scales, information provided by others in the child's environment or by the child himself or herself can be influenced by a host of variables and biases. This increases the complexity of interpreting these types of assessment by requiring assessors to account for these influences in their interpretations. Therefore, direct observations of behavior eliminate a great deal of the complexity in the interpretive process. Second, direct observations of behaviors frequently allow

P.J. Frick et al., *Clinical Assessment of Child and Adolescent Personality and Behavior*,
DOI 10.1007/978-1-4419-0641-0_8,

for the assessment of environmental contingencies that are operating to produce, maintain, or exacerbate a child's behavior. For example, direct observations can assess how others respond to a child's behavior, or they can detect environmental stimuli that seem to elicit certain behaviors. By placing the behavior in a contextual framework, behavioral observations often lead to very effective environmental interventions.

To illustrate this potential of behavioral observations, Carroll, Houghton, Taylor, West, and List-Kerz (2006) conducted a study of 58 students (ages 8 to 11) in which two students, one with Attention Deficit Hyperactivity Disorder (ADHD) and one with no disorder, were observed for 40 min. As would be expected, the children with ADHD showed more off-task behavior. However, the observational system documented "triggers" to this off-task behavior. Nearly, half of the off-task behaviors of students with ADHD could be attributed to environmental distractions, and over a quarter were preceded by specific teacher behaviors.

While these characteristics of behavioral observations make their use an important component of many clinical assessments, we feel that the importance of direct observation is sometimes overstated. Like any assessment technique, direct observations have several limitations. One of their major limitations is that direct observations are often expensive and time consuming, if one is to obtain high-quality information. Because of their cost, many assessors simply eliminate this source of information from their assessment battery. Alternatively, assessors may attempt less rigorous observations than are appropriate. For example, an assessor may observe a child interacting on a playground for a 20-min period and record the child's behavior in a narrative form, without clearly specifying what behaviors will be observed or how they will be recorded. These informal observations are dangerous if the assessor is unaware of the severe limitations and potential biases in the data that are collected and, instead, interprets the data as if they were objective (see Harris & Lahey, 1982a).

Another result of the costliness of direct observations is that the development of many observational systems has ignored basic psychometric considerations (Hartmann, Roper, & Bradford, 1979). In the previous chapters on rating scales, we focused a great deal of attention on the psychometric properties of scales such as the different types of reliability that have been established, the information on the validity of the scales, and the normative base with which to compare scores. Because of cost factors, few observational systems have established their reliability or validity in multiple samples. An even more widespread problem for observational systems is the lack of a representative normative sample that would allow for a comparison of a child's scores with those from the general population. As we have discussed in earlier chapters, having norm-referenced scores is crucial in the clinical assessments of children and adolescents, given the rapid developmental changes they are experiencing.

Even if one were to use an observational system in the most sophisticated manner, direct observations are still limited by (1) the reactivity of the observational setting, (2) difficulties in obtaining an adequate sample of behaviors, and (3) an inability to detect internal events such as cognitions and emotions. *Reactivity* refers to a well-documented phenomenon that a person will change his or her behavior when it is being observed (Kazdin, 1981; Mash & Terdal, 1988). As a result, the sample of behavior may not be as objective as one would hope. There is a significant amount of research on factors that influence the degree of reactivity that results from direct observations (Harris & Lahey, 1982b; Kazdin, 1981). For example, the age of the child can affect the degree of reactivity, with preschool children showing less reactivity to observation than older children

(Keller, 1986). Also, steps can be taken to reduce reactivity during observation such as allowing the child time to get used to (habituate to) the observational setting and reducing the conspicuousness of the observational system (Keller, 1986). But, even under optimum conditions, reactivity is still likely to affect the results of the assessment to some degree.

Another liability of direct observations is the difficulty of obtaining an adequate sample of behaviors. There are several facets to this issue. The first issue involves ensuring that the sample of behavior is obtained under the most ecologically valid conditions; that is, under conditions that will generalize to other times and situations. Although the issue of ecological validity is most important for observational systems that use contrived (analog) conditions (e.g., observing the child in a clinic playroom), it is also important in selecting the natural setting most appropriate for conducting the observation. The second issue is that, even if one selects the best setting, one must ensure that a large enough time frame is used, so that behaviors will be representative and generalizable to other times and settings. In the previous example of a child being observed in a playground setting for a 20-min period, it cannot be determined how typical a child's behavior was during this observational period. He or she may have had an especially good or especially problematic day on the playground. A third issue, which encompasses both the selection of settings and adequacy of the observational period, is the difficulty in assessing many behaviors that are very infrequent (e.g., cruelty to animals, hallucinations, panic attacks) or by nature covert (e.g., stealing, lying). In most cases, one would not ethically want to contrive a situation that would prompt such behaviors, and the behaviors are often too infrequent to be observed naturally occurring in the child's environment.

A final issue in the use of behavior observations is the fact that observations are limited to the assessment of overt behaviors. They do not provide a means for assessing the cognitive, affective, and motivational components of a child's functioning (Mash & Terdal, 1988). This does not negate the importance of having a good assessment of a child's overt behavior in making diagnostic and treatment decisions. However, it has become increasingly clear that overt behavior is only one piece of a complex puzzle. Research in several areas of child psychopathology has supported the importance of intrapsychic variables for both assessing (e.g., Frick, 2006) and treating (e.g., David-Ferdon & Kaslow, 2008) children and adolescents.

In summary, direct observations are affected by some factors that often preclude their use in many clinical settings and limit the usefulness of the data obtained. We spent a great deal of time reviewing the factors that affect behavioral observations, not because of a bias against this form of assessment, but because we have found that assessors sometimes ignore these issues. We feel that a clinical assessor should be aware of these issues in deciding whether or not direct observations should be included in an assessment battery and should consider these issues when interpreting observational data. However, these limitations should not be considered any greater than those associated with other assessment techniques, and the limitations must be weighed against some very important advantages of direct observation (e.g., elimination of reporter bias and ready translation into environmental interventions). Direct observations can be an integral part of many assessment batteries but, as is the case for all assessment techniques, they also have limitations in the information they provide in isolation.

In the following section, we discuss basic issues in the development and use of observational systems. As was mentioned earlier in this chapter, many clinical assessors use informal observational techniques

in their assessment battery without establishing a well-defined system. Unfortunately, the information obtained from such systems is difficult to interpret. Unlike rating scales, there are few standardized observational systems that are readily available for clinical use that have well-established psychometric properties. Therefore, the next section focuses on basic considerations in designing an observational system for one's own clinical use. Following this discussion, some examples of observational systems that are commercially available or that have been used in research are reviewed.

Basics of Observational Systems

Defining Target Behaviors

The basic components of observational systems can be broken down into the what, where, how, and by whom of the system. The first part of developing a system of direct observation involves defining *what* behaviors one wishes to observe. Defining the behaviors of interest first involves deciding on the level of analysis one wishes to use (Barrios, 1993). Specifically, the level of analysis can be at the level of isolated behaviors, at the level of constellations of behaviors (syndromes), or at the level of interactions within a social unit. As an example of the social unit level of analysis, many observational systems allow for the recording of how a child behaves in response to parental behavior and how a parent responds to a child's behavior (Gelfand & Hartmann, 1984). Also, the example given previously demonstrates an observational system that focuses on behaviors by fellow students and teachers that can influence the on-task behavior of a student with ADHD (Carroll et al., 2006). Because these systems allow one to document events (stimuli) that elicit a behavior and responses to the behavior that may help to maintain or increase it, this level of analysis provides important information on potential targets of intervention. An example of a simple antecedent-behavior-consequence (A-B-C) type of observation is provided in Box 8.1. From this example, it is clear that recording antecedents and consequences allows one to determine the sequence of events within which a behavior is embedded.

After the level of analysis is chosen, one must operationally define what behaviors, what constellation of behaviors, or what antecedents/consequences will be observed within this window. These definitions are made prior to beginning a direct observation and must be specified in objective and understandable terms in order to reduce the potential for bias and increase the reliability of the observation. Some examples of target behaviors used in observational systems are described in Table 8.1.

The target behaviors in Table 8.1 are simply lists of behaviors from several domains that can be assessed by observational systems. In order to be reliable, coding systems must have very explicit definitions of each behavior. This is necessary to reduce the possibility that the observer will use subjective and idiosyncratic definitions of the behaviors, thereby making interpretations from the observations difficult. Without such definitions the primary advantage of direct observations, objectivity, is severely compromised. One would think that behaviors such as those in Table 8.1 are easy to define and that simple definitions would lead to different observers being able to code the same behavior in the same way. Decades of research have found that this is not true. To reliably code behaviors, one must develop very detailed definitions. Box 8.2 provides an example of a very detailed definition of behavior from a frequently used coding system.

Box 8.1

A Hypothetical Example of Simple A-B-C Observational System of an 8-Year-Old Boy (B)

Time/Setting	Antecedent	Behavior	Consequence
8:30/Math class-copying from board		B takes pencil from another child	Child ignores him
	Child ignores him	B tears paper on child's desk	Child tells teacher and teacher reprimands B
	Teacher reprimands B	B sulks	Teacher allows B to erase board
8:35/Math class-doing seatwork		B leaves seat to sharpen pencil	Teacher asks B to raise hand to leave seat
		B raises hand	Teacher continues to work with other student
	Teacher ignores B	B gets out of seat and pulls on teacher's shirt to get attention	Teacher scolds B for leaving seat and places name on board
	Teacher puts B's name on board	B starts to cry	Child teases B
	Child teases B	B tries to hit other child	B sent to office
8:55/Math class-completing seat work	B returns to class	B sullen and refuses to work	Teacher allows B to collect assignments

TABLE 8.1 Examples of Target Behaviors from Several Behavioral Domains

ADHD (Carroll et al., 2006)	**Conduct Problems (Patterson, 1982)**	**Social Competence (Dodge, 1983)**	**Depression (Kazdin, 1988)**	**Autism (Lord, Rutter, DiLavore, & Risi, 1999)**
Off task	Noncompliance	Solitary play	Talking	Asking for help
Fidgeting	Destructiveness	Cooperative play	Playing alone	Symbolic play
Inappropriate talking	Aggressive play	Smiling	Negativism	Taking turns
Gazing around	Insults/threats	Compliments	Frowning	Reciprocal play
Out of seat	Aggression	Rule making	Complaining	Telling a sequential story
Loud talking	Arguing	Turn taking	Whining	
	Teasing			

Box 8.2

Criteria for "Whine" from the Dyadic Parent–Child Coding System

Definition
A whine consists of words uttered by the child in a slurring, nasal, high-pitched, falsetto voice.

Examples
When can we go home?
Mommy, I hurt my finger.
I have to go to the bathroom. This is too hard.
I don't want to play this anymore.

Guidelines
1. The voice quality of the word or phrase is the primary distinguishing element for coding whine.
2. Each whined sentence constitutes a separate whine. Whined phrases separated from one another by a pause of 2 s or longer are coded as separate whines.

Examples:
Child: I have a headache. I want to go home. (2 whine)
Child: I don't like the red blocks… 2-s pause… and I don't like the Legos. (2 whine)
Child: Please let me take it home… 2-s pause… Please. (2 whine)

3. The content of the word or phrase may be anything except smart talk.

Examples:

I don't like this anymore.	(whine)
I hate you.	(smart talk)
I feel sick.	(whine)
You make me sick.	(smart talk)
You hurt my feelings.	(whine)
You're a jerk.	(smart talk)

4. Whining is a verbal behavior and can occur simultaneously with a nonverbal deviant child behavior (destructive or physical negative child).

Decision Rules
1. When uncertain as to whether the child's voice quality is actually a whine or normal voice quality, do not code whine.
2. When uncertain as to whether a child's verbalization is a whine, smart talk, or a cry, code whine.
3. When uncertain as to whether the deviant behavior is a whine or a yell, code yell.

Source: Summarized from the manual for the Dyadic Parent–Child Interaction Coding System (Eyberg & Robinson, 1983) with the authors' permission.

Setting

Once the target behaviors are defined, the next decision is to determine *where* to observe these target behaviors. Naturalistic observations involve observing the child in his or her natural setting (e.g., in the classroom, at home). The kind of behaviors of interest (e.g., social interactions during free play) often determines what natural setting is best to conduct the observation (e.g., on the playground). In its purest form, naturalistic observations involve placing no constraints on a child's behavior other than those naturally occurring in the observational setting. However, sometimes it is necessary to place some restrictions on the observational setting to enhance the quality of the observations. For example, an observer who is in the home of a child to observe parent–child interactions may need to place some constraints on the child and parents to ensure that there are sufficient opportunities to observe interactions during the observational session. For example, one may wish to place restrictions that parents and children must stay in the same room and that there is no talking on the telephone, working on a computer, playing video games, or watching TV. Another example is an observational

system designed to observe a child's anxious behavior. The observer may wish the teacher to "create" a situation that seems to lead to anxiety in the child, such as being called on in class or taking a test, in order to observe the child's response.

In Box 8.3, we provide an example of a study by Ostrov and Keating (2004) in which observations of aggressive behavior of preschool children were observed in both free play and during several structured interactions. Both types of observations were conducted in a naturalistic setting (i.e., the child's school). The use of different observational situations within the same study allowed the authors to determine the types of settings in which aggression is most likely to occur.

Naturalistic observations are often preferred because they generally provide more ecologically valid data. However, time and cost constraints may prevent one from conducting a naturalistic observation. For many clinical assessments, it is often impossible for the assessor to make several

Box 8.3

An Observational Study of Preschool Aggression

Ostrov and Keating (2004) reported a study of aggression in preschool children using naturalistic observations in the child's school setting. This study provides a good example of two common types of naturalistic observational techniques, one in which no restrictions are placed in the natural setting (i.e., free play) and one in which the situation is structured (i.e., coloring task). The study involved 46 children (mean age of 64 months) in rural preschools.

What: The observation coding system focused behaviors in four main categories: (1) *physical aggression:* hitting, pushing, pulling, punching, forcibly taking objects; (2) *verbal aggression:* teasing, calling mean names, verbal threats of harm, insults; (3) *relational aggression:* excluding from play group; spreading rumors, withdrawing friendship; telling lies; ignoring peer; (4) *number of male and female playmates:* number of children of each sex the observed child directly interacted.

Where: Free play observations were conducted during regularly scheduled free play periods in large indoor playrooms, in classrooms, and outdoors on the playground. For the coloring task, pairs of children were given a series of three pictures to color. However, the potential for mild conflict was introduced by providing one colorful crayon and one white one.

How: Each observational session was 10 min and every instance of the specified behaviors were coded. Each child was observed for five sessions. Behavioral counts were summed across observational periods to determine a score in each of the four behavioral categories.

By Whom: Observers were three female and one male undergraduate students who were trained on the observational system. Prior to conducting observations, observers were introduced to the teacher and students and they spent a few days in the classroom to let the students adjust to their presence.

Results: Boys exhibited more physical aggression but girls displayed more relational aggression. Aggression was less overall and these gender differences were less pronounced during the coloring task. However, there was fairly high stability in a child's level aggression across contexts.

Summarized from: Ostrov, J. M., & Keating, C.F. (2004). Gender differences in preschool aggression during free play and structured interactions: An observational study. *Social Development*, *13*, 255–277.

home visits to observe a child or adolescent interacting with his or her parents. Also, for some behaviors, there may not be a way of obtaining unobtrusive observations in a child's natural environment. As a result, the level of reactivity would be so high that the data would be meaningless. In addition to these more practical considerations, sometimes there is a need to exert more control over the situation than is possible in a natural setting. For example, one may wish to observe a child's activity level in a free play situation by determining how many times a child passes from one part of a room to another. To code this reliably, one can divide the room into sections with tape and then code the number of times a child crosses over a tape divider (Milich, 1984). This type of control (e.g., dividing the playroom into grids) may not be feasible in a child's natural environment.

For these reasons, it is sometimes necessary or desirable to conduct analog observations in a laboratory or clinic. *Analog* refers to the creation of a contrived setting that approximates the natural environment. Dividing a clinic playroom into grids to observe a child's activity level is one example of an analog setting. However, the key to these observations is how well the analog situation approximates the natural environment. Staying with our example, it would be imperative that the playroom be similar to a play area that a child would be in outside of the clinic (e.g., with age-appropriate toys available). There are many other examples of analog settings for behavioral observations, but each involves the basic component of simulating a child's natural environment in a clinic setting.

Sometimes it is not feasible to have the clinic setting approximate the natural setting. In these cases, children may be asked to imagine themselves in a situation, and their behavior is observed in this role-play situation. An area in which role play observations have been frequently used is the assessment of children's social competence (e.g., Bornstein, Bellack, & Hersen, 1977; Dodge, McClaskey, & Feldman, 1985). For example, Dodge et al. (1985) had children pretend that they were in certain social situations and then pretend that the assessor was another child. An explicit coding system was developed to code the degree of social competence of a child's behavior in each of the imagined situations. An example of one of the role-play situations used in this study is included in Box 8.4.

Data Collection

The next stage in developing an observational system is to determine *how* one will code the target behaviors in the selected setting. There are several data collection methods that can be used, with the method of choice depending on the characteristics of the behaviors of interest. Although there are many variations of these basic data collection methods, the techniques can be largely placed into three categories: Event Recording, Duration Recording, and Time Sampling.

Event Recording

Event recording is the simplest of the data collection methods. It involves recording the number of times that a target behavior occurred during preset intervals or during an entire observational session. This method was illustrated in the study of preschool aggression described in Box 8.3 (Ostov & Keating, 2004). Due to its simplicity, event recording is the most frequently used method of direct observation. However, to use event recording, target behaviors must have discrete beginnings and endings, such as hitting another child, raising one's hand, and asking to play a game (Shapiro, 1987). In contrast, behaviors that are continuous and persist for long periods of time are more difficult to code using event recording because it

Box 8.4

A Role-Play Situation from the Dodge Study of Social Competence

Situation #2

Let's pretend that I'm playing blocks with some of my friends after lunch. We're building a really neat house. You come in the schoolroom and see us. Pretend that you really want to play blocks with us. what do you do and say?

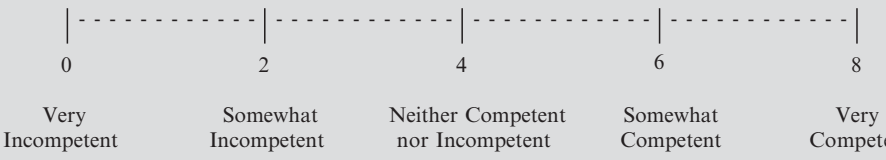
a. Was the child role playing?
b. How competent was the response? Score:
8-Complimentary or evaluative remark with a re-quest to play: "Boy, that's neat. Can I play?"

6—A simple request to play: "I'd ask, May I please play?"

4—Rhetorical question or evaluative remark: "What are you doing?" or "That's neat."

2—Suggestion for different activity: "Want to play a game?"

0—Aggressive responses: "I'd knock the blocks over", "I would say nothing or sit down at desk without speaking": "I do'nt know what to do", or no answer.



Reproduced with permission of authors from the Scoring System for Child Role Plays: Role Playing Criteria used in Dodge, K. A., McClaskey, C. L., & Feldman E. (1985). Situational approach to the assessment of social competence in children. *Journal of Consulting and Clinical Psychology, 53*, 344–353.

is difficult to distinguish the occurrence of one incident from the next. Examples of such continuous behaviors are off-task behavior, talking out loud, and engaging in solitary play. Event recording is especially useful for recording behaviors that occur only briefly and for recording low-frequency behaviors that only occur once or twice in an observational period, such as swearing or hitting another child (Keller, 1986; Shapiro, 1987).

Duration Recording

For some assessments, it may be more important to know how long a behavior occurs rather than the frequency of the behavior. In duration recording, the observer records the length of time from the beginning to the end of an instance of behavior. Duration and event recording can be combined to provide an even richer source of information (Shapiro, 1987). For example, in observing the temper tantrums of a young child, it may be helpful to record not only the frequency of tantrums within a given period, but to also record the duration of each tantrum episode.

Time Sampling

In both event and duration recording techniques, all instances of the behaviors are recorded during the observational period.

However, some behaviors occur too frequently to obtain accurate counts, or there are no clear beginning or end points to the behaviors, which prevents effective event and duration recording. For these types of behaviors, the observation period can be divided into predetermined intervals, and the behaviors are simply coded as being either present or absent during each interval. Therefore, rather than yielding an exact count of the number of times that a behavior occurred during the observational period, time sampling allows one to determine the proportion of intervals in which the behavior occurred.

Shapiro (1987) reviews three types of time-sampling techniques. In *whole-interval* recording one codes a behavior as present only if it occurs throughout a time interval. For example, an observational period in a child's classroom can be broken down into 20-s intervals and the number of intervals in which the child remained on task for the entire interval is recorded. In *partial-interval* recording, one records whether or not a behavior occurred at any time during the interval. Shapiro gives the example of a teacher dividing the day into 15-min segments and noting whether or not certain behaviors occurred during each segment. A final type of time sampling is *momentary* recording, in which one records whether a behavior was present or absent only during the moment when a time interval ends. For example, when observing the degree of social withdrawal of a child, one may divide an observational period into 60-s intervals and record whether or not a child was engaged in interactions with other children at the end of each interval.

Selecting the Observers

After determining what behaviors will be observed, where the observations will take place, and how the observations will be conducted, one still must determine *who* is best suited to conduct the observation. Having someone who is in the child's natural setting (e.g., teacher or parent) conduct the observation is often useful in naturalistic observations because it helps to maintain unobtrusiveness. However, it is often difficult to teach people in the child's natural environment how to use a coding system and to ensure that it is being used appropriately.

In order to exert more control over the observational methodology, many observational systems require rigorously trained and monitored observers. Barrios (1993) provides a summary of steps required in training and monitoring observers (see Table 8.2). As one can see from this summary, using specially trained observers is quite costly. Such stringent methodology is feasible only when a large number of children are being observed with the same observational method. Therefore, this methodology may not be optimal in many clinical settings.

One type of observation that is frequently used in clinical assessments involves training a person to observe his or her own behavior. The same steps of selecting target behaviors, determining the setting for the observation, and determining how the target behaviors will be recorded are followed. However, in *self-monitoring* the child is trained to record his or her own behavior. Although self-monitoring has been used largely with adults, children have used self-monitoring systems to monitor such diverse behaviors as classroom attending, class attendance, talking out in class, room cleaning, aggression, and inappropriate verbalizations (Mash & Terdal, 1988).

One method of self-monitoring that has begun to receive some attention in research is the use of electronic diaries. That is, with the advent of small computers and personal digital assistants (PDAs), children and adolescents can be taught to

Table 8.2 Steps in Training and Monitoring Observers

Step	Description
Orientation	Informing observers of the importance of objective assessment in the understanding and treating of childhood disorders. Informing observers of their duties and responsibilities, in particular their independent, unbiased, and faithful recording of the behavior of interest.
Education	Instructing observers in the response definitions and recording scheme through the use of written materials, filmed illustrations, and live demonstrations.
Evaluation	Assessment of observers' knowledge of the response definitions, coding system, and recording scheme through the use of written and oral examinations. Representation of materials until observers are thoroughly acquainted with all aspects of tracking and recording of the behaviors of interest.
Application	Graduated implementation of the observation system across a range of situations, beginning with analog ones and ending with actual setting of interest. Transition from one situation to the next contingent upon observers achieving a criterion level of agreement and accuracy.
Recalibration	Assessment of the accuracy and agreement of observers' recordings in the setting of interest. Identification and correction of any breakdowns in the fidelity of observers' recordings.
Termination	Questioning observers as to the merits of the observation system. Informing observers of their contributions to the understanding and treating of the behaviors of interest. Reminding observers of the need to maintain confidentiality.

Reproduced with permission from Barrios (1993). Direct observation. In T. H. Ollendick and M. Hersen (Eds.), Handbook of child and adolescent assessment (pp. 140–164). Boston: Allyn & Bacon.

carry a small PDA which cues the child to respond to certain prompts and questions. This allows the child to record responses in real time (i.e., when it is actually occurring). For example, Whalen et al. (2006) had 52 children (mean age of 10.58) report on their moods, behaviors, and social contexts every 30 min during nonschool hours. The children carried a PDA that beeped every 30 min to signal it was time to respond to certain questions. The children's responses were saved in the PDA for later analyses. Twenty-seven of the children were diagnosed with ADHD and the results showed that these children had more behavioral problems, negative mood, and conflict with parents.

Research suggests that children can self-monitor their behavior accurately if they are trained appropriately, have a clear and simple observational system, have an outside monitor of the accuracy of their recording, and are reinforced for the accuracy of their recording (Keller, 1986). The advantages of self-monitoring are that it is cost effective and is less intrusive than many other forms of behavioral observation. However, research clearly suggests that children change their behavior as they become more aware of it through self-monitoring (Keller, 1986; Shapiro, 1987). Whereas this change in behavior may be a beneficial aspect of self-monitoring in a treatment program, this reactivity limits the usefulness of self-monitoring as a means of obtaining objective information on a child's behavior for assessment purposes.

Examples of Observational Systems

In contrast to behavior rating scales, there are few observational systems that are widely used, have standardized procedures, and are readily available for clinical use. In this section, we review some notable exceptions. The goal of this overview is not to provide an exhaustive list of observational systems but to provide a carefully selected list of observational techniques that vary in terms of target behaviors, settings of observation, method of data collection, and degree of training needed to reliably use the observational system. We feel that, even if one does not choose to use one of the specific systems discussed here, these systems provide concrete examples of some of the issues discussed in this chapter and therefore can serve as a guide for the development and use of other observational systems. Also, some additional observational systems that focus on a specific areas of adjustment are reviewed in other chapters.

Achenbach System of Empirically Based Assessment: Direct Observation Form and Test Observation Form

The ASEBA system of assessments contains two observational systems. The first is the Direct Observation Form (DOF; Achenbach, 2001) which is designed for use with children and adolescents ages 5 to 14. It provides a method of coding observations in academic classrooms and other group activities. The Test Observation Form (McConaughy & Achenbach, 2004) is designed for use with children and adolescents ages 2 to 18 and allows for the coding of behavioral observations during the individual administration of standardized ability and achievement tests. Both of these systems are part of the ASEBA system and are designed to be interpreted in conjunction with the parent, teacher, and child self-report versions of the ASEBA, all of which have been discussed in previous chapters.

The DOF is designed to provide a direct observation of a child in a classroom or group setting during a 10-min period. There are three parts to the DOF. First, the observer is asked to write a narrative description of a child's behavior throughout the 10-min observational period, noting the occurrence, duration, and intensity of specific problems. Second, at the end of each minute the child's behavior is coded as being on- or off-task for 5 s. Third, at the end of the 10-min period, the observer rates the child on 96 behaviors that may have been observed during the observational period using a 4-point scale (from 0 = behavior was not observed to 3 = definite occurrence of behavior with severe intensity or for greater than 3 min duration).

The 96 problem behaviors on the DOF have a high degree of item overlap with the behaviors rated on the parent and teacher rating scales of the ASEBA system. Therefore, the DOF is nicely suited for a multimodal assessment of a child's or adolescent's emotional and behavioral functioning. Like the other parts of the ASEBA system, the DOF can be used to calculate a Total Problem score, which is a sum of the ratings of all 96 problems, two broadband scales (Internalizing and Externalizing), and six narrowband scales (Withdrawn-Inattentive, Nervous-Obsessive, Depressed, Hyperactive, Attention-demanding, Aggressive) (Achenbach, 2001). The DOF does report norms from a relatively small sample of 287 nonreferred children (Achenbach, 2001).

There is evidence that the DOF can be used reliably by observers with minimal training. Inter-observer correlations have been calculated on the Total Problems scale in several samples. Correlations between observers range from .96 in a

residential treatment center (Achenbach, 1986) to .92 in a sample of boys referred for special services in school (Reed & Edelbrock, 1983) to.75 in a sample of outpatient referrals to a child psychiatry clinic (McConaughy, Achenbach, & Gent, 1988). Inter-observer correlations for the On-task scores were .71, .71, and .88 in the three samples, respectively. Reed and Edelbrock (1983) reported inter-observer reliability in their sample of 25 boys for a selected set of individual items from the DOF. In general, most items showed high inter-observer correlations (most above .80), with the exceptions of Nervous, high-strung, or tense (.20); Picks nose, skin, or other parts of body (.52); and Compulsions, repeats behavior over and over (.53).

Reed and Edelbrock (1983) reported that the Total Problems scale and On-task scores from the DOF correlated in expected directions with teacher ratings of total problems and adaptive behaviors. In addition, the DOF Total Problems scale and On-task scores have been shown to differ in normal and disturbed children (McConaughy et al., 1988; Reed & Edelbrock, 1983). In terms of discriminating within disturbed children, the evidence for the DOF is less clear. McConaughy et al. (1988) reported that the Total Problems, On-task, and Externalizing scores differentiated children classified with internalizing or externalizing problems. However, the internalizing scale of the CBCL-DOF did not demonstrate discriminant validity in this study.

The TOF is designed to rate children's behavior during an individual standardized testing session. It has 125 items that are rated on a four point scale (0 = "no occurrence" – 3 = "definite occurrence with severe intensity or 3 or more minutes duration"). Items are rated by the examiner immediately after the testing session. Like the DOF, the TOF has a strong overlap in items with other measures in the ASEBA system. Thus, a Total Problem, an Internalizing, and an Externalizing composite can all be obtained from the TOF. Also, it includes five narrow band scales: Withdrawn/Depressed, Language/Thought Problems, Anxious, Oppositional, and Attentional Problems.

The TOF has a normative sample of 3,943 children between the ages of 2 and 18, most of which were obtained during the standardization of the Stanford-Binet Intelligence Scales-5th Edition (Roid, 2003). In general, the TOF scales show good test-retest reliability (.53–.87) over a period of 10 days, adequate interrater reliability (.42–.73), and good internal consistency (.74–.94) (McConaughy, 2005). Scores on the TOF are moderately correlated with corresponding parent completed ASEBA (.27–.43) and teacher completed ASEBA (.26–.38) scales (McConaughy, 2005). Also, the TOF scales have differentiated children with ADHD from normal control children (McConaughy, 2005).

In summary, the TOF and DOF are both time-efficient observational systems that require minimal observer training and fit into a multimethod assessment system. Both observational systems have information showing some basic levels of reliability. They both also provide norm-referenced scores, although the sample for the DOF norms is very limited. Also, both systems have proven to be useful for discriminating normal from clinic-referred children. However, their ability to differentiate within children with emotional and behavioral problems is less clear.

Behavioral Assessment System for Children-Student Observation System

The BASC-2-Student Observation System (BASC-2-SOS; Reynolds & Kamphaus, 2004) is a commercially available, short (15-min) observational system that is designed for use in a classroom setting. It is part of the comprehensive BASC-2 system, which

includes parent and teacher behavior rating scales and a child self-report form, all of which have been discussed in previous chapters. The BASC-2-SOS defines 65 specific target behaviors that are grouped into 13 categories (4 categories of positive/adaptive behaviors and 9 categories of problem behaviors). The 13 categories and examples of target behaviors in each category are provided in Table 8.3.

Table 8.3 Behavioral Categories from the Student Observation System of BASC2 Category/Definition

Category/Definition	Example of Specific Behaviors
Response to Teacher/Lesson (appropriate academic behaviors involving teacher or class)	Follows directions
	Raises hand
	Contributes to class discussion
	Waits for help on assignment
Peer Interaction (appropriate interactions with other students)	Plays with other students
	Interacts in friendly manner
	Shakes hand with other student
	Converses with others in discussion
Work on School Subjects (appropriate academic behaviors that student engages in alone)	Does seatwork
	Works at blackboard
	Works at computer
Transition Movement (appropriate nondisruptive behaviors while moving from one activity to another)	Puts on/takes off coat
	Gets book
	Sharpens pencil
	Walks in line
	Returns material used in class
Inappropriate Movement (inappropriate motor behaviors that are unrelated to classroom work)	Fidgeting in seat
	Passing notes
	Running around classroom
	Sitting/standing on top of desk
Inattention (inattentive behaviors that are not disruptive)	Daydreaming
	Doodling
	Looking around room
	Fiddling with objects/fingers
Inappropriate Vocalization (disruptive vocal behaviors)	Laughing inappropriately
	Teasing
	Talking out
	Crying
Somatization (physical symptoms/complaints)	Sleeping
	Complaining of not feeling well
Repetitive Motor Movements (repetitive behaviors that appear to have no external reward)	Finger/pencil tapping
	Spinning an object
	Body rocking
	Humming/singing to oneself
Aggression (harmful behaviors directed at another person or property)	Kicking others
	Throwing objects at others
	Intentionally ripping another's work
	Stealing
Self-Injurious Behavior (severe behaviors that attempt to injure one's self)	Pulling own hair
	Head banging
	Biting self
	Eating or chewing nonfood items
Inappropriate Sexual Behavior (behaviors that are explicitly sexual in nature)	Touching others inappropriately
	Masturbating
	Imitating sexual behavior
Bowel/Bladder Problems (urination or defecation)	Wets pants
	Has bowel movement outside toilet

From C. R. Reynolds and R. W. Kamphaus (2004). Behavior assessment system for children – 2nd Edition (BASC-2). Circle Pines, MN: American Guidance Services.

The BASC-2-SOS was designed to be completed during a 15-min observation of the child in an academic classroom.

The behaviors during the observation are coded in three parts. In Part A, each of the 65 behaviors are rated as being "Not Observed," "Sometimes Observed" or "Frequently Observed." Also, any behavior judged to be disruptive is noted. Part B uses a momentary time-sampling approach in recording data. The 15-min observational period is divided into 30 intervals. At the end of each 30-sec interval the child's behavior is observed for 3-sec. A checklist allows the observer to mark each category of behavior that occurred during the 3-sec observation interval. In Part C, the observer is asked to describe the teacher's interactions with the student, focusing on contingencies in the classroom that may be influencing the child's behavior.

The BASC-2-SOS is a simple and time efficient observational system that assesses, through direct observation, many behaviors that are crucial for the clinical assessment of children and adolescents. It is one of the few direct observational systems commercially available. Also, the BASC-2-SOS has an electronic coding format in which behavioral observations can be recorded directly into a PDA device.

However, the BASC2-SOS lacks a number of crucial psychometric elements. First, there is no information on the reliability of the system provided in the manual. Establishing inter-observer agreement is a crucial component in developing an observational system, to ensure that observations are objective and relatively free from bias. Second, there are no norms for the BASC-2-SOS, so norm-referenced interpretation of scores is not possible. Third, there is limited information on the validity of BASC-2SOS, such as whether or not the BASC-2-SOS code categories correlate with clinically important criteria (e.g., diagnoses, behavior rating scales, response to intervention). Lett and Kamphaus (1997) reported that scores on the BASC-2-SOS did differentiate children with ADHD from normal control children. However, the limited information on the psychometric properties of the BAS-2-SOS greatly limits its potential contribution to many clinical assessments.

Behavioral Avoidance Tests

Behavioral Avoidance Tests (BATs) have been used to observe a person's behavioral response to anxiety-producing stimuli since the early 1900s (Jersild & Holmes, 1935). Although there are many different versions of BATs (e.g., Morris & Kratochwill, 1983; Van Hasselt, Hersen, Bellack, Rosenblum, & Lamparski, 1979), they all involve exposing the child or adolescent to some feared stimuli (e.g., animal, dark, stranger, heights, blood), then requiring the child to approach the feared stimuli in graduated steps.

BATs provide explicit and objective criteria for observing a child's behavioral reaction to the feared stimulus, such as how closely the stimulus is approached, the number of steps in the gradual approach that are taken, or the time spent touching or handling the phobic stimulus. In quantifying these responses to anxiety-producing stimuli, BATs provide a measure of the severity of a child's anxiety and can help document changes brought about by interventions (Vasay & Lonigan, 2000).

Southam-Gerow and Chorpita (2007) provide a good summary of the advantages and disadvantages of BATs. They describe the primary disadvantage of BATs as the absence of a single standardized BAT. Instead, there have been numerous different BATs developed that vary widely on the number of steps in the graduated approach, the types of instructions given to the child, and how the feared stimulus is presented. As a result, it is impossible to compare the findings across studies and therefore it is impossible to develop a significant body of knowledge on the reliability, validity, and normative base

for any of the BATs. In addition, a hallmark of the BATs is the rigorous control over how the feared object is presented to the child and how the child approaches it. This degree of control may prevent the behavior observed in the contrived setting from generalizing to the child's natural environment. One final limitation of BATs is that they are most commonly used to assess specific fears and are more difficult to design for assessing more generalized anxiety (Vasay & Lonigan, 2000).

On the positive side, however, BATs are relatively simple and time efficient in their administration. Many of the BATs have been shown to have good inter-observer agreement with minimal observer training and to be sensitive to treatment effects (Barrios & Hartmann, 1988; Southam-Gerow & Chorpita, 2007). Also, scores from BATs are correlated with subjective ratings of fear and with phobia diagnoses (Vasay & Lonigan, 2000). Most importantly, they provide one of the only methods of assessing the behavioral components (i.e., a child's avoidance of a feared stimulus) of childhood anxiety.

Conflict Negotiation Task

The Conflict Negotiation Task was designed to assess peer interactions, especially those interactions that may be associated with childhood depression (Rudolph, Hammen, & Burge, 1994). Children are observed with an unfamiliar partner of the same age and gender. The system uses a task involving three points of potential conflict of interest between the child and his or her partner. First, child dyads are placed in a situation in which they are to build structures with colored blocks to match either of two models. They are informed that whoever constructs an identical model would win a prize. The dyad is given a set of blocks to share but the number of blocks is only sufficient to build one complete model. Second, after 10 min of observation, the dyad is asked to decide on how to distribute two prizes of unequal value. Third, the dyad participates in a 5-min interview and the child who received the less valuable prize is allowed to choose a new one.

The interactions during this task are coded by trained observers. Using an event recording system, behaviors are rated on a seven-point scale (0 = not at all present; 4 = moderately present; 7 = to a large degree present) and all ratings are scaled such that high scores indicate more negative peer interactions. The behaviors are grouped into four composites. Two composites are related to broad dimensions of social behavior displayed by the child being assessed. Conflict-resolution competence includes persistence in problem-solving efforts, positive assertiveness, positive conflict management, and general social competence. Emotional regulation includes conflict exacerbation, positive affect, and negative affect. The third composite consists of a dyadic quality code based on ratings of conflict or friction within the dyad, collaboration, problem-solving competence of the dyad, and mutuality/reciprocity. Finally, a peer response code is based on the ratings of the peer's behavior toward to the assessed child, including general response valence (negative or positive) to the target child, discomfort and embarrassment in response to the target child, and emotional state at the end of the interaction.

Rudolph et al. (1994) reported on the use of this system in a sample of 36 children (20 girls and 16 boys) between the ages of 7 and 13. The four composites from this observational system showed very high correlations between raters (.88 to .92) and the behaviors within each composite were highly intercorrelated (.82 to .97). Most importantly, when the 36 children were divided into those high on a measure of depression and those low on this measure, depressed children were rated as

significantly less competent on all four of the composites. Specifically, children high on depression were observed to be significantly worse in their conflict-resolution competence, in their emotional regulation, in mutuality and cooperation during the interaction, and in the negative valence placed on the interactions by their partners (e.g., their partner being more uncomfortable in the interactions). The authors point out that these data support the contention that children who score high on depression have significant interpersonal difficulties. Furthermore, the findings suggest that observational systems of children's interactions with peers can be useful for assessing important aspects of these social difficulties.

Dodge's Observation of Peer Interactions

Dodge (1983) developed a direct observation system that also assesses several components of peer interactions. However, this observation system focuses on behaviors associated with acceptance in a peer group. Dodge developed his system in a sample of 5-, 6-, 7-, and 8-year-old boys. Children were observed in 60-min play groups of eight boys each by three observers who were stationed behind a one-way mirror. The boys wore numbered T-shirts to aid in quick identification by the observer. Each observer coded the behaviors of one boy for a 6-min period and then coded a second child according to a prearranged schedule.

There were 18 target behaviors of five types that were defined for the observation system (see Table 8.4). A complex event recording system was used for the observations. Each time a target behavior was observed, the observer coded the time, the context (structured vs. unstructured), the target behavior observed, and the peer target (number of child). Observers received extensive training over a 4-week period and, with this training, the observational codes showed quite high inter-observer agreement (Dodge, 1983). Across the 18 target behaviors, 15 behaviors showed inter-observer agreement of 65% or better.

TABLE 8.4 Target Behaviors in the Dodge Observational System of Peer Interactions

Behavior Category	Target Behaviors
Solitary active	Solitary play
	Watching peers
	On-task behavior
	Off-task behavior
Interactive play	Cooperative play
	Aggressive play
	Inappropriate play (e.g., standing on table)
Verbalizations	Social conversations with peers
	Norm-setting statements (e.g., rule making)
	Hostile verbalizations (e.g., insults, threats)
	Supportive statements (e.g., compliments, offers of help)
	Exclusions of peers from play
	Extraneous verbalization (e.g., laughs, cheers)
Physical contact with peers	Hits
	Object possession (e.g., grabbing an object from peer)
	Physically affectionate behavior (e.g., holding hands, hugging)
Interactions with adult	Social conversation with adult leader
	Reprimanded by adult group leader

From K. A. Dodge (1983). Behavioral antecedents of peer social status. Child Development, 54, 1386–1399.

The three behaviors that showed poor inter-observer agreement were watching peers, norm-setting statements, and supportive statements.

Family Interaction Coding System

One of the most common uses of behavioral observations is to observe parent–child interactions, especially those associated with childhood conduct problems (Frick & Loney, 2000). For example, Patterson and colleagues at the Oregon Social Learning Center have developed a direct observational system designed to assess children's conduct problems in the home and to assess the interactional patterns in which the conduct problems are often embedded. The Family Interaction Coding System (FICS; Patterson, 1982) is composed of 29 code categories that include both child behaviors and parental reactions to the child behaviors. These categories are summarized in Table 8.5. The goal of the FICS was to observe children interacting with family members in natural home settings. However, as described by Patterson (1982), several restrictions had to be made in the home for the observational sessions. Specifically, to use the FICS, all family members must be present during the pre-arranged observation times with no guests present, and the family is limited to being present in two rooms of the house. There can be no telephone calls out (only brief answers to incoming calls) and no TV. Finally, there is to be no talking to the observers during coding.

Table 8.5 Target Behaviors in the Family Interaction Coding System

Approval	High rate[a]	Physical negative[a]
Attention	Humiliate[a]	Physical positive
Command	Ignore[a]	Receive
Command negative[a]	Laugh	Self-stimulation
Compliance	Noncompliance[a]	Talk
Cry[a]	Negativism[a]	Tease[a]
Disapproval[a]	Normative	Touch
Dependency[a]	No response	Whine[a]
Destructiveness[a]	Play	Yell[a]

Reproduced with author's permission from G. R. Patterson (1982). Coerceive family process. Eugene, OR: Castalia.

[a] Denotes aversive behaviors that are included in the Total Aversive Behavior (TAB) score.

The FICS was designed to have data coded continuously and to provide a sequential account of the interactions between a child and other family members. The behavior of the child and the person(s) with whom the child interacts are coded in sequence. After initial coding, many of the child's behaviors are summarized in a rate-per-minute variable that combines both frequency and duration of the behavior. However, the most frequently used score from the FICS is the Total Aversive Behavior (TAB) score which is a sum of the number of aversive behaviors which occurred during the observational session.

Patterson (1982) describes a moderate level of inter-observer agreement for most code categories of the FICS, with the categories of Negativism and Self-Stimulation showing the most questionable levels of agreement. One week test-retest reliability of the TAB was studied in a sample of 27 boys and was found to be quite high (.78). The TAB was also found to discriminate between families of children referred for behavior problems and nonreferred families and has proven to be sensitive to family-focused treatment for children's conduct problems (Patterson, 1982). Although most of the individual code categories of the FICS can be coded consistently by two observers, psychometric information is generally limited to the global index of aversive behavior, the TAB.

In summary, the FICS is an example of an observation system designed to assess a child's behavior in the home environment and to assess parent–child interactional patterns in which the behavior is embedded. The FICS is generally most useful for younger children (10 and under). Also, most of the psychometric information available for the FICS is for the aversive behaviors assessed by the TAB. This is not as severe a limitation as one might think, however, given that these aversive behaviors have proven to be quite important to understanding and treating children with conduct problems (Patterson, 1982).

Structured Observation of Academic and Play Settings

Milich, Roberts, and colleagues (Milich, 1984; Milich, Loney & Landau, 1982; Roberts, Ray, & Roberts, 1984) developed an observational system (Structured Observation of Academic and Play Settings [SOAPS]) to assess behaviors associated with ADHD in a clinic playroom analog setting. In this system, a clinic playroom is designed with age-appropriate toys, four tables, and a floor divided into 16 equal squares by black tape. The child is placed in two situations. Free Play involves the child being placed alone in the room and allowed to play freely with the toys. The Restricted Academic Playroom Situation involves the child being requested (1) to remain seated, (2) to complete a series of academic tasks, and (3) not to play with any of the toys.

Each observational situation lasts for 15 min. A combination of event recording and time sampling is used in this observation system. Event recording is used to determine the total number of grids crossed for the entire observational period. That is, the number of times that a child moves completely from one square of the divided room into another is counted. Event recording is also used to determine the number of times the child shifts his or her attention from one task to another during the entire observational period. A 5-sec time sampling procedure is used to observe other target behaviors. These include the proportion of 5-sec intervals that the child is out of his or her seat, fidgeting, noisy, and on task. In addition, a 5-sec time sampling is used to determine the number of intervals that the child was observed touching forbidden toys. Also during the academic task, the number of items completed is recorded.

This observational system is useful in clinical assessments of ADHD for a number of reasons. First, it is a relatively easy to use observational system. As a result, high inter-observer reliability has been obtained for most categories with minimal observer training. Second, categories from this system have been correlated with clinicians' diagnoses of ADHD, they have differentiated ADHD children from aggressive and other clinic-referred children (Milich et al., 1982), and they have been relatively stable over a 2-year period (Milich, 1984). Third, a modified version of this task has been shown to be sensitive to treatment with stimulant medication (Barkley, 1988).

Conclusions

In this chapter, we have attempted to summarize both the advantages and disadvantages of direct observations of behavior as part of a comprehensive clinical assessment. Although there are some limitations, direct observations are a useful component to many assessment batteries. Probably the biggest limitation to the clinical utility of direct observations is the time and cost involved in conducting behavioral observations appropriately. We have attempted to outline some of the major considerations in developing and using observational systems so that clinical assessors can (1) evaluate

existing observational systems appropriately, (2) develop their own observational systems as needed, and (3) recognize limitations in the data provided by observational systems that are not developed and used in a sound manner.

We concluded this chapter by providing an overview of several existing observational systems. This overview was not meant to be exhaustive. The observational systems included were specifically chosen to provide examples of the various domains of behavior that observational systems can assess, the different settings in which observations can be conducted, and the various methodologies that can be employed. Two of the systems are commercially available (ASEBA-DOF/TOF; BASC-2-SOS) and cover a broad array of behaviors. The other systems reviewed focus on more narrowly defined dimensions of behavior such as anxiety, social interaction, aggression, or ADHD. Whether or not a clinical assessor chooses to use these specific systems, we feel that concrete examples of observational systems help to illustrate the unique contribution that behavioral observations can make to an assessment battery.

Chapter Summary

1. Direct observations of children's and adolescents' behavior are an important part of an assessment because they provide an objective view of behavior that is not filtered through an informant.
2. Direct observations are also helpful in assessing environmental contingencies that affect a child's behavior.
3. Direct observations also have a number of limitations:
   a. Conducting observations in a way that provides valuable information is often an expensive and time-consuming process.
   b. Because of the cost of obtaining observations, the development of observational systems often has ignored basic psychometric considerations such as testing the reliability of the system or developing an adequate normative base.
   c. Even well-developed observational systems are subject to reactivity. That is, persons change their behavior when they are aware of being observed, which reduces the validity of the observations.
   d. Other factors affecting the validity of observational systems include the difficulty in observing an adequate sample of behavior and the inability to observe internal events.
4. The basic components of observational systems include:
   a. What-defining target behaviors to be observed.
   b. Where-selecting the most appropriate setting in which to observe the behavior
   c. How-determining how the target behaviors will be coded
   d. Who-determining who should observe the target behaviors
5. In defining target behaviors, one must consider the level at which behaviors will be defined and then clearly define the behaviors to be observed.
6. Observations conducted in a child's natural environment have greater ecological validity but allow less control over the observational setting than observations conducted in a laboratory or analog setting.
7. There are three basic ways in which behaviors can be recorded in an observational system:
   a. Event recording-the number of times a behavior occurred is recorded.

   b. Duration recording-the length of time from the initiation to the desistance of a behavior is recorded.
   c. Time sampling-behaviors are recorded as to whether or not they have occurred during preset time intervals.
8. Observations can be conducted by outside observers, people in a child's environment (e.g., parents, teachers), or by the child or adolescent himself or herself.
9. The ASEBA-DOF and ASEBA-TOF are two observational systems that can be used in conjunction with other components of the ASEBA system.
   a. The DOF allows for a direct observation of a child's classroom behavior during three to six 10-min observational periods.
   b. The TOF allows for a direct observation of a child's behavior during individual standardized testing, and it has strong normative data.
   c. Both the DOF and TOF have evidence supporting their reliability and for differentiating referred from nonproblem children.
10. The BASC-2-SOS is an observational system used to assess classroom behavior that can be integrated with other assessment components of the BASC-2 system.
   a. It specifies 65 target behaviors and uses a momentary time-sampling procedure for observations.
   b. There is no normative information nor is there any information on the reliability or validity of this observational system.
11. BATs allow one to observe a child's response to anxiety-provoking stimuli.
12. Observational systems have also been developed to assess peer interactions, parent–child interactions, and behaviors associated with ADHD.