Chapter 11

Structured Diagnostic Interviews

Chapter Questions

- What are the major differences between structured and unstructured interviews?
- What are the major similarities and differences among the most common structured diagnostic interviews used to assess children?
- What information can diagnostic interviews provide that cannot be obtained from other assessment techniques?
- What are some important guidelines for the appropriate use of diagnostic interviews in the assessment of children and adolescents?

History

Clinical interviews have a prominent place in the history of psychological assessment. The face-to-face verbal dialog between assessor and client is the prototypical format for most clinical enterprises. Historically, the most common type of clinical interview has been the unstructured interview. In the unstructured interview, the interviewer determines what questions should be asked, how the questions should be framed, what follow-up questions should be asked, and what are acceptable responses from the client. This unstructured format is quite consistent with the clinical approach to assessment, which was discussed in the previous chapter on projective techniques. It allows the assessment to be tailored to the individual needs of the client and relies heavily on the individual clinician's orientation and expertise.

However, the unreliability inherent in unstructured interviews generates some significant problems. The results and interpretation of such interviews tend to be highly idiosyncratic to the clinician conducting the interview. This unreliability is especially problematic for research. Hence, many clinical assessors have developed more

P.J. Frick et al., *Clinical Assessment of Child and Adolescent Personality and Behavior*,
DOI 10.1007/978-1-4419-0641-0_11,

structured diagnostic interview schedules that provide a clear and standardized format from which to conduct the clinical interview. This standardization helps address the key problems associated with unstructured interviews; however, one must recall that these interviews are valid only for certain purposes (e.g., collecting data on specific symptoms) and relatively less valid for others (e.g., treatment planning for a specific client; Mash & Hunsley, 2005).

Initially, most of these instruments were designed for use with adults and were primarily used in research. Two of the better known early interview schedules were the Feighner Research Diagnostic Criteria (Feighner et al., 1972), which later became the NIMH Diagnostic Interview Schedule (DIS; Robins et al., 1981), and the Schedule for Affective Disorders and Schizophrenia (SADS; Endicott & Spitzer, 1978).

Over the past two decades, structured diagnostic interviews have moved from being strictly research instruments to being a part of many clinical assessments. In addition, several interview schedules have been developed for use with children and adolescents. This chapter focuses on these interview schedules and their potential role in the clinical assessment of children and adolescents.

Overview

Structured diagnostic interviews consist of a set of questions that the assessor asks the informant (e.g., parent or child). There are explicit guidelines on how responses are to be scored. Interview questions generally start with a stem question (e.g., Have you been involved in many physical fights?). If the stem is answered affirmatively, then follow-up questions are asked to determine other relevant parameters such as frequency (e.g., How many fights have you been in, in the past year?), severity (e.g., Have you ever used a weapon in a fight?), onset (e.g., When was the first time you got into a fight?), and impairment (e.g., Has fighting caused problems for you at school, home, or with kids your age?). An example of the question format from the NIMH Diagnostic Interview Schedule for Children-Version 4 (DISC-IV; Shaffer et al., 2000) is provided in Box 11.1.

Box 11.1

Example of the DISC-IV Question Format

The following is an example of the stem/follow-up question format used by most diagnostic interview schedules. This question was taken from the questions assessing Major Depressive Disorder from the DISC-IV (Shaffer et al., 2000).

"I'm now going to ask some questions about feeling sad and unhappy."

1. In the last year, was there a time when you often felt sad or depressed for a long time each day?

   IF YES,

   A. Was there a time in the last year when you felt sad or depressed for a long time each day?
      IF NO, GO TO QUESTION 2
   B. Would you say that you felt that way for *most of the day?*
   C. Was there a time when you felt sad or depressed *almost every day*?
      IF NO, GO TO QUESTION 2
      IF YES,
   D. In the last year, were there two weeks in a row when you felt sad or depressed almost every day?
      IF NO, GO TO QUESTION 2
   E. When you were sad or depressed, did you feel better if something good happened or was about to happen to you?
   F. Now, what about the *last four weeks?* Have you felt sad or depressed?

Commonalities Across Interview Schedules

In Table 11.1, we summarize the basic characteristics of some of the most commonly used diagnostic interviews for children. It should be noted that the interviews are continuously updated to keep up with changing diagnostic criteria or modified for some specific application (e.g., outcome measure for treatment of childhood depression, screening device for an epidemiological study of child disorders). Therefore, many of the interviews have multiple versions and are continually being revised. The contents of these interviews are all based on the criteria specified in the *Diagnostic and Statistical Manual of Mental Disorders*, starting with the third edition (*DSM-III;* American Psychiatric Association, 1980) and continuing through its most recent revision (*DSM-IV-TR;* American Psychiatric Association, 2000).

Which *DSM* disorders are assessed is generally quite similar across the different interview schedules. All of the interviews can be used to assess for the disruptive behavior disorders, affective disorders, and anxiety disorders in children and adolescents. Each of these interviews also allows for at least a brief screening for schizophrenia. The majority of structured interviews also allow for an assessment of substance use, elimination, and eating disorders. Tic disorders are covered exclusively by CAPA, DISC-IV, and K-SADS, whereas the ADIS contains unique screening questions for mental retardation, learning disorder, and somatoform symptomatology. The CAPA is unique in its detailed assessment of sleep disorder symptoms. Most interviews organize questions by diagnosis. However, the Interview Schedule for Children and Adolescents (ISCA; Sherrill & Kovacs, 2000) provides a symptom-oriented interview format with items clustered by content (e.g., impaired concentration) and topic area (e.g., mental status) rather than specific diagnostic criteria.

In order to promote multi-informant assessments, most interviews contain parallel forms to ask identical questions to both the child and a parent. There is even an experimental teacher version of the DISC 2.3, an earlier version of the DISC, that was used in the *DSM-IV* field trials for the disruptive behavior disorders (Frick et al., 1994).

A recent trend in the most widely used structured interview schedules is the development of computer-administered versions. Currently, the DISC-IV, the DICA, the CAPA, and the Dominic-R all have formats that can be administered by computer. The computer format was designed to enhance the reliability of the interviews by increasing the ease of administration and data collection. Similar to the pencil-and-paper version, the computer format is designed to be administered by an examiner. The examiner reads items from the computer screen and enters the patient's responses. The computer quickly scores and stores responses, selects the appropriate follow-up questions, and skips out of diagnostic sections when the respondent fails to meet a certain threshold of severity. Also, the computer-administered interviews have programs that quickly score the interview and provide various summary scores (e.g., symptom indexes, number of diagnostic thresholds met) that aid in the interpretation of the results.

Most of the interviews were designed to assess children and adolescents between the ages of 8 and 17. Some interviews report applicability to younger children with parents as informants (e.g., Valla et al., 2000). However, there is some evidence that the reliability of children's self-report on diagnostic interviews is low before age 9 (Edelbrock et al., 1985; Hodges & Zeman, 1993). The length of time that it takes to administer a diagnostic interview is heavily dependent on the child being assessed. Because of the stem/follow-up question format, children with

TABLE 11.1 Overview of Structured Diagnostic Interviews for Children

Interview	Time	Ages	Time Frame	Informants	Comments
Anxiety Diagnostic Interview for *DSM-IV* (ADIS; Albano & Silverman, 1996; Silverman, Saavedra, & Pina, 2001)	60–90 min	7–17	Present	Child and Parent	Semi-structured interview focusing on childhood anxiety disorders. Categorical and dimensional symptom assessment with diagnostic summary that places diagnoses in hierarchical format. Support for reliability and validity for anxiety disorder diagnoses. Limited validity data for other diagnostic categories.
Child and Adolescent Psychiatric Assessment (CAPA; Angold & Costello, 2000)	60–120 min	9–17	Present	Child and Parent	Semi-structured interview with flexibility in wording and in probing for relevant symptoms. Can be used by trained lay interviewers. Includes Likert-type symptom severity scores. Separate section for impairment. Good reliability except for modest estimates for ODD and CD. Some validity evidence for diagnoses relative to external criteria.
Children's Interview for Psychiatric Symptoms (ChIPS; Weller et al., 2000)	30–60 min	6–18	Present	Child and Parent	Highly structured format. Can be used by trained lay interviewers. Brief with broad symptom coverage. Simplified language for use with younger children. Limited research, but some evidence of good sensitivity and specificity.
Diagnostic Interview for Children and Adolescents (DICA; Reich, 2000; Reich et al., 1982)	60–120 min	6–18	Lifetime	Child and Parent	Structured with changes toward more flexible format. Good reliability with some evidence of validity. Evidence supporting use categorically and dimensionally. Used cross-culturally and with a variety of populations.
Diagnostic Interview Schedule for Children (DISC-IV; Shaffer et al., 1993, 2000)	90–120 min	6–17	Present (separate lifetime module)	Child and parent (experimental teacher version available)	Highly structured format designed for use by trained lay interviewers. Good reliability at symptom and diagnosis level. Covers many diagnostic domains. Frequently used in epidemiological studies. Diagnoses from parent version related to greater impairment.

Dominic-R (Valla, Bergeron, & Smolla, 2000)	15–25 min	6–11	Present	Child	Highly structured pictorial interview with verbal symptom descriptions. Designed for use by lay interviewers with minimal training. Promising reliability and validity data. Should be interpreted with caution if used with young children.
Pictorial Instrument for Children and Adolescents (PICA-III-R; Ernst, Cookus, & Moravec, 2000)	40–60 min	6–16	Present	Child	Semi-structured pictorial interview with verbal symptom descriptors. Designed for use by experienced clinicians. Categorical and dimensional symptom assessment of 14 diagnostic categories. Should be interpreted with caution if used with young children. Limited psychometric data. Authors note the need for update to reflect *DSM-IV* criteria.
Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS; Ambrosini, 2000; Chambers et al., 1985; Tillman et al., 2003)	90 min	6–18	Present and lifetime (in separate and combined interview formats)	Child and Parent	Semi-structured interview originally designed to focus on depression. Variety of versions that vary in length. Low level of structure necessitates use by clinicians. Includes ratings of symptom severity. Widely used in a variety of populations with good psychometric evidence.

more symptoms will require more interview time because of the need to ask more follow-up questions. However, as is evident from Box 11.1, the average time to administer the interviews does not vary much across the different schedules and lasts typically from 60 to 90 min.

Major Sources of Variation Across Interview Schedules

From the previous discussion it is clear that the various interview schedules probably have more similarities than differences. However, one of the major differences across schedules is the *degree of structure* inherent in the interview format. All of the interviews provide some degree of structure and give guidelines for standardized administration and scoring. However, there is substantial variation in the amount of "leeway" given the assessor across the various interviews. For example, the K-SADS is one of the least structured of the interviews. The original manual for administration includes the following instructions:

> "The K-SADS supplies a series of questions addressed to the child for each item to be rated. The aim is not to oblige the rater to ask all of the questions. They serve as a guide for questions which have been found most helpful and informative. The rater should ask as many questions as necessary to arrive at a well-documented rating. Needless to say, probing should be as neutral as possible and leading questions should be avoided" (Puig-Antich & Chambers, 1978, p. 2).

In contrast, the DISC was designed to have a high degree of structure in administration. The manual for its administration includes the following instructions:

> "The DISC symptom questions are designed to be read *exactly* as written. There is very limited scope for independent questioning. DO NOT deviate from the prescribed question sequence. DO NOT make up your own questions because you think you have a better way of getting at the same information, or because you think the question is poorly worded" (Fisher et al., 1992, p. 31).

The trade-off between leeway and structure is obvious. Less structure allows the assessor to tailor the interview according to the needs of the individual client. However, these interviews generally require a greater degree of experience to administer and often have lower levels of reliability (Gutterman, O'Brien, & Young, 1987; Hodges & Zeman, 1993).

Another major variation among the structured interviews for children is the time frame used to assess symptoms and diagnoses. All of the interviews assess whether problems are currently evident. This is called *a present episode* frame of reference. Most interviews consider present episodes to be within the previous six months, although in some instances the time frame may be as short as within the last two weeks (e.g., ISCA for emotional disorders) or as long as within the last year (e.g., DISC-IV for Conduct Disorder). Of note, the CAPA restricts the assessment of symptoms to the previous three months due to concerns with the reliability of memory in children and adults over longer time intervals. Similarly, the Dominic-R does not obtain any temporal information, such as onset and duration of symptoms, given concerns as to the validity of such information in young children. Nevertheless, the major source of variation is whether or not interview schedules are limited to present episodes. A number of interviews restrict the focus of assessment to the present episode time frame (e.g., CAPA and ChIPS). However, an increasing number of interviews provide for the assessment of both present and *lifetime* diagnoses. For example, the DISC-IV provides a more recently incorporated whole-life module assessing for whether or not a

child has exhibited symptoms of diagnoses since age five but prior to the current year (Shaffer et al., 2000). Similarly, there are lifetime formats for both the ISCA and K-SADS. The DICA is unique in its exclusive focus on lifetime diagnoses.

A third source of variation within the interview schedules is the answer format. For most interviews, the interview responses are coded into a categorical format (yes, no). This categorical format is consistent with the *DSM* orientation in which symptoms are considered either present or absent. In contrast, the ADIS, CAPA, ISCA, PICA-III-R, and K-SADS have answer formats that can be placed on a Likert-type scale that allows one to rate the severity of a symptom. While this format makes it more difficult to translate responses into *DSM* diagnoses, it does not create an artificial dichotomy between the presence or absence of a symptom and allows symptom scores to reflect gradations in severity.

Evaluation of Diagnostic Interviews

Advantages

Structured interviews share with behavior rating scales the goal of obtaining a detailed description of a child's emotions and behaviors from multiple informants. The logical question is: what advantages do the time-consuming structured interviews offer in comparison to the more time-efficient behavior rating scales? Some of the more important advantages of structured diagnostic interviews follow.

(1) Structured interviews are useful in obtaining important parameters of a child's behavior that are not typically assessed by most behavior rating scales. Specifically, most interview schedules provide questions that elicit information on the duration of a child's behavioral difficulties and the age at which the problems began to emerge. This temporal information allows one to take a developmental perspective in understanding a case, a perspective that has proven to be crucial for assessing many forms of childhood psychopathology (e.g., Silverthorn & Frick, 1999).

(2) Interviews also allow one to determine the *temporal sequencing* among behaviors. For example, it is important in the assessment of childhood depression to determine whether periods of sadness occurred contiguously with other behaviors associated with depression, such as sleep disturbances, eating disturbances, or thoughts of death (Kazdin, 1988).

(3) Most structured interviews assess the *level of impairment* associated with behaviors being reported. Most interviews have questions that elicit information on the degree to which a child's difficulties are affecting his or her functioning in major life arenas (e.g., at home, at school, and with peers).

(4) Diagnostic interviews also enhance the *correspondence between assessment techniques and diagnostic criteria*. As mentioned previously, the most common structured interviews were specifically designed and revised to correspond to the changing *DSM* system. Therefore, the usefulness of the interview is, in part, dependent on the usefulness of the diagnostic definitions that are being assessed.

This tie between assessment and diagnosis can be advantageous for several reasons. First, it promotes revisions of the interviews to correspond with advances in our knowledge of the basic characteristics of child and adolescent psychological disorders. Second, it allows one to make a diagnosis based on strict adherence to diagnostic criteria. Either due to theoretical, empirical, or practical reasons

(e.g., insurance reimbursement), many clinicians attempt to make *DSM* diagnoses as a result of their assessments. Too often, diagnoses are made based on information (e.g., rating scales, projective tests) that do not directly assess the diagnostic criteria or through techniques that are unsystematic in their assessment of symptomatology and associated features (e.g., unstructured interviews; Klein et al., 2005). As a result, the meaning of the diagnosis is ambiguous. Structured diagnostic interviews do not have this problem. They tend to result in more reliable diagnoses (Silverman & Ollendick, 2005) and can actually provide important information that may alter diagnostic impressions and treatment recommendations (Kashner et al., 2003).

(5) Diagnostic interviews are also helpful in *training* clinical assessors. As assessors are developing their competence in interviewing, it is often helpful to have an explicit format from which to conduct the interview. It gives the assessor a good way to learn the basic characteristics of childhood emotional and behavioral disorders. After being trained in administration procedures and after conducting several interviews with actual clients, assessors often begin to internalize the diagnostic criteria for the most common disorders of childhood. This knowledge can then be applied in situations in which a structured interview is not possible.

Disadvantages

Diagnostic interviews also have a number of weaknesses of which the clinician should be aware.

(1) The time consuming nature of the interviews coupled with some question as to whether they provide incremental validity in the assessment of some disorders such as ADHD (Pelham et al., 2005) suggest that structured interviews may not be practical or useful in many situations.

(2) In addition, structured interviews depend on DSM criteria, which is a strength for assessing well-validated syndromes but a weakness for assessing disorders with a weak empirical basis.

(3) Diagnostic interviews are subject to the same potential *reporter biases* that were discussed in previous chapters on self-report inventories and parent and teacher behavior rating scales.

(4) Making differential diagnoses with the assistance of structured interviews still does not directly translate to plans for intervention (Mash & Hunsley, 2005).

(5) There is great difficulty in making *norm-referenced interpretations* from interviews. Clinically significant levels of symptoms are often based on *DSM* criteria rather than based on a comparison with a representative normative sample. Therefore, the appropriateness of clinical elevations for a given age depends on the appropriateness of the diagnostic criteria for that age. For example, Barkley (1997) has questioned the validity of applying the same diagnostic criteria (i.e., the same number of symptoms) for ADHD across the early-childhood and adolescent years based on a documented decline in hyperactivity and impulsivity symptoms as children move into adolescence. As a result, the uniform diagnostic threshold, while appropriate for elementary school-aged children, may be too conservative for adolescents.
Several community studies using an early version of DISC (e.g., Anderson, et al., 1987; Costello, 1989) allow one to view the base rates of disorders assessed by the DISC in community samples of school-age children. Such work has continued with the current version with various populations (e.g., Roberts,

Roberts, & Xing, 2007). Although these studies provide some information on how a child who meets criteria would compare to others in the general population, this type of normative information is still much more limited than the type provided by other assessment techniques, most notably behavior rating scales.

(6) Another weakness of most diagnostic interviews is the failure to provide a format for obtaining information from a child's teacher. This source of information is crucial in the clinical assessment of elementary school-aged children (Loeber et al., 1991). As a result, information from teachers must be obtained by some other method, thereby making it difficult to determine if discrepancies between a teacher's report and the report of others are due to real differences in a child's classroom behavior or whether they are due to differences in the assessment format.

Recommendations for Use of Structured Interviews

Based on the strengths and weaknesses of structured interviews, several recommendations can be made on their appropriate use in the clinical assessment of children and adolescents. First, like any assessment technique, the diagnostic interview should never be used alone in a clinical evaluation. It should be one part of a comprehensive assessment battery. For example, information from diagnostic interviews should be supplemented by assessment techniques that provide better norm-referenced scores (e.g., behavior rating scales) and by assessment techniques that provide information on a child's classroom functioning (e.g., behavioral observations in the school).

In addition, the diagnoses derived from the diagnostic interviews should be viewed within the context of the overall assessment. A diagnosis can be viewed similarly to the way an elevation on a behavior rating scale is interpreted. Specifically, it is one piece of information that needs to be integrated with other sources of information to develop a good case formulation. Stated simply, diagnoses based solely on diagnostic interviews should not be considered final clinical diagnoses. Such final diagnoses should be based on an assessor's integration of multiple sources of. A case example in which a diagnostic interview was used as part of a comprehensive assessment battery is provided in Box 11.2.

The child's age is also an important consideration in the use of structured interviews. Generally, the reliability for most interview schedules is low before the age of 9 for child self-report (Hodges & Zeman, 1993). It seems that the structured, face-to-face dialogue is not appropriate for assessing very young children. Several interviews have been developed using pictorial stimuli, rather than relying purely on question-and-answer format, in an effort to increase the reliability of the interview for younger children.

A description of two such pictorial interviews is provided in Box 11.3.

Another important issue in interpreting information obtained from standardized interview schedules is the consistent finding that, if the interview is repeated, parents and children uniformly report fewer symptoms on the second administration (Jensen, Watanabe, & Richters, 1999; Piacentini, et al., 1999). For example, Piacentini, et al. (1999) found that when the DISC-IV was readministered to the same sample of 245 parent–child pairs (age 9–18) 12 days later, parent-reported symptoms dropped 42%, and child-reported symptoms dropped 58%,

The reasons for this symptom attenuation have not been conclusively shown. However, they can include (1) sensitization to clinical

Box 11.2

Case Example: The DISC-IV in the Evaluation of a 9-Year-Old Girl with Attention-Deficit Hyperactivity Disorder

Alexis who is 9 years and 3 months old was referred for a comprehensive psychological evaluation by her parents upon the recommendation of her teachers. Her teachers had reported to Alexis' parents that she was having difficulty paying attention and was daydreaming, interrupting others, and making careless mistakes in her work. Her parents requested a comprehensive evaluation to determine the severity and possible cause of these difficulties and to get recommendations for possible interventions to aid in her school adjustment.

Alexis's background, developmental, and medical history were unremarkable. During the testing, Alexis had great difficulty concentrating and was easily distracted. She was also very fidgety and restless. Intellectually, Alexis had much better verbal comprehension abilities, especially in the area of verbal reasoning, than nonverbal perceptual-organizational abilities. Consistent with her verbal abilities, Alexis scored in the above average range on measures of reading and math achievement.

Alexis's emotional and behavioral functioning were assessed through the use of structured interviews con-ducted with Alexis, her parents, and her teachers as well as through rating scales completed by her parents and teacher. The structured interviews were the parent version of the DISC-IV and the experimental teacher version used in the DSM-IV field trials (Frick et al., 1994). The child version was given to Alexis.

The following is an excerpt from the report of Alexis's evaluation that illustrates how information from the DISC-IV was integrated with other assessment information.

The only problematic areas that emerged from this assessment of Alexis' emotional and behavioral functioning were significant problems of inattention, disorganization, impulsivity, and overactivity that seem to be causing her significant problems in the classroom. On the DISC-IV, Alexis's parents agreed that such difficulties were noticeable by the time she was age 5. Alexis herself endorsed many symptoms of inattention, but not overactivity, on the DISC-IV. Alexis's teachers described her as being very restless and fidgety, being easily distractible, having very disorganized and messy work habits, having a hard time completing tasks, and making many careless mistakes. Results from teacher rating scales further suggested that these behaviors are more severe than would be typical for children Alexis's age. These behaviors are consistent with a diagnosis of Attention-Deficit Hyperactivity Disorder (ADHD). Also consistent with this diagnosis were the age of onset of these symptoms as reported by Alexis's parents. These behaviors associated with ADHD seem to be causing significant problems for Alexis in school, affecting the amount and accuracy of her schoolwork.

issues leading to a heightened threshold for symptom reporting, (2) a circumscribed focus on only the interval between assessment periods, (3) statistical regression to the mean, and (4) knowledge that brevity of responding will shorten the duration of the interview (Piacentini et al., 1999). This symptom attenuation is not much of an issue in most clinical uses of structured interviews that do not involve multiple administrations. However, there is also evidence that the number of symptoms reported declines *within* an interview schedule, such that parents and child tend to report more symptoms for diagnoses assessed early in the interview, even if the order of assessment is varied (Jensen et al., 1999).

This type of symptom attenuation within an interview is of much greater concern because it clearly can influence the results from a typical clinical use of structured interviews. In an attempt to solve this problem, Edelbrock, Crnic, and Bohnert (1999) modified the administration of DISC-2.3

Box 11.3

Research Note: Developing Structured Interviews for Young Children

To increase the usefulness of structured interviews for assessing younger children, several authors have attempted to develop interviews that involve pictorial content to either replace or augment the typical question-and-answer format for structured interviews. One example of this approach is the Dominic-R by Valla et al. (2000). The Dominic-R is designed to assess *DSM-III-R* criteria for anxiety disorders, mood disorders, and the disruptive behavior disorders in children ages 6–11.

The interview involves pictures of a child named Dominic facing situations that are common in children's daily lives. The pictures are accompanied by questions about the visual image (e.g., "Do you feel sad and depressed most of the time, like Dominic?"). There is a version of the Dominic-R, the "Terry Questionnaire," that employs an African American character named "Terry," and there are translations of the Dominic-R in French, Spanish, and German. There is also an Interactive Dominic Questionnaire that is a CD-ROM-based interactive cartoon. The Dominic-R takes about 15 to 25 minutes to complete and is highly structured, which allows it to be administered by lay interviewers. The reliability of the Dominic-R was assessed in a sample of 340 community children aged 6 to 11, and it revealed reliability coefficients that were much improved over other structured interviews with very young children. An adolescent version has also been developed (Smolla et al., 2004).

For the child version, test-retest reliability over 7 to 12 days for diagnoses from the Dominic-R ranged from a kappa of .44 to a kappa of .69, with most being above .60 (Valla et al., 1997). Also, the diagnoses based on the Dominic-R were strongly associated with diagnoses made by experienced clinicians with kappa values ranging from .64 to .88 (Valla et al., 2000). Furthermore, research has shown the Dominic-R to accurately designated children meeting *DSM-IV* criteria for Conduct Disorder (Arseneault et al., 2005). Therefore, it appears that the combination of pictorial stimuli and verbal stimuli can enhance the reliability and validity of responses in young children.

A similar assessment system, described by Ernst, Cookus, and Moravec (2000), is called the Pictorial Instrument for Children and Adolescents (PICA-III-R). The PICA-III-R is a semi-structured interview that includes 137 pictures assessing anxiety disorders, mood disorders, disruptive behavior disorders, psychotic disorders, and substance abuse. Therefore, the content of the PICA-III-R is somewhat broader than that of the Dominic-R. Like the Dominic-R, the PICA-III-R combines pictorial stimuli with verbal questions. However, the verbal questions on the PICA-III-R are more extensive, often including many follow-up questions (e.g., "Do you ever get like him (e.g., sad)?" "How much?" "Do people tell you that you look sad?"). Also, the PICA-III-R verbal questions are not meant to be read verbatim. The type and degree of questioning is left somewhat up to the interviewer. Because of all of these characteristics, the PICA-III-R may be more useful for older children with more severe forms of psychopathology compared to the Dominic-R, and it must be administered by a experienced clinician. Ernst et al. (2000) reported that, in a sample of 51 inpatient children and adolescents (aged 6 to 15), the PICA-III-R scales were generally internally consistent (i.e., most over .80), although the internal consistency of the Mania (.69) and Obsessive-Compulsive symptoms (.54) were somewhat low. There is minimal information on the validity of the PICA-III-R other than the fact that scores from this interview changed over the course of hospitalization for the inpatients, which presumably reflected improvement brought about by treatment.

Both the Dominic-R and PICA-III-R provide examples of ways to enhance the usefulness of structured interviews in obtaining child self-report. It is important to note, however, that both of these interviews are in the very early stages of development, and much more information is needed on their

(Continues)

Box 11.3 (Continued)

reliability across samples and their validity for assessing *DSM* diagnoses.

Finally, another example of a unique approach to the assessment of young children is the Berkeley Puppet Interview (see Measelle, Ablow, Cowan, & Cowan, 1998). This interview calls for two puppets to make statements about themselves and then ask the ask the child a question with two choices based on the same attribute (e.g., "Are you good at making friends or are you not good at making friends?"). This interview includes academic, social, and symptom-related (i.e., aggression-hostility and depression-anxiety) domains. Some evidence of convergent validity has been found based on ratings by parents and teachers (Measelle et al., 1998) and the interview has been able to differentiate clinic-referred from community samples of children (Ablow et al., 1999). However, much more research on the validity, feasibility, and utility of this assessment tool is needed.

to produce more stable symptom endorsements that are less susceptible to attenuation effects. They included a more detailed introduction with an overview of all areas to be covered on the interview, and they provided definitions of key concepts used throughout the interview. Furthermore, they employed a flexible order of administration that allowed parents to select the order in which the various diagnoses were assessed. Using this methodology in a sample of 24 parent-child pairs with children ranging in age from 6 to 15, there was very minimal reduction in the number of symptoms reported when the interviews were repeated one week later. However, none of the standardized administration procedures that accompany structured interview schedules provide for this type of administration, and therefore, the possible drop in number of symptoms reported for disorders assessed later in the interviews must be considered when interpreting the results for individual children and adolescents.

A final consideration in using structured interviews concerns when to administer diagnostic interviews in the assessment battery. There is no research on this issue, and these recommendations come from clinical experience. Diagnostic interviews should not be the first assessment administered to parents. The structured format does not facilitate the development of rapport between the interviewer and parent, and some parents become frustrated in trying to fit their main concerns and descriptions of their child's behavior into the confines of the interview. Therefore, it is often helpful to precede diagnostic interviews with less structured questions that allow parents to express, in their own words, their concerns for their child. However, for children and adolescents, we actually find that the structured format enhances rapport in many cases. Children often enter the assessment situation nervous because they are unsure about what is expected of them. The clear and explicit response format of diagnostic interviews makes the demands of the situation apparent for the child and thereby reduces his or her anxiety in many cases.

Up to this point, we have tried to give an overview of structured interviews, looking at the various formats that are available, highlighting some of the advantages and disadvantages of using interview schedules in a clinical assessment, and providing guidelines for appropriate use. In the next section, we provide a more in-depth look at one particular interview schedule, the DISC-IV. We chose the DISC-IV as a prototypical example of a structured interview

because it is one of the most widely used interview schedules for children and adolescents, and it has been one of the most systematically developed. However, one must be aware that the DISC-IV is one of the most structured interview schedules, and therefore, it has all of the advantages and disadvantages that accompany a high degree of structure.

Focus on the NIMH Diagnostic Interview Schedule for Children (DISC-IV)

Development

The original version of the DISC (DISC-1; Costello, 1983; Costello et al., 1984) was designed to be a downward extension of the adult-oriented Diagnostic Interview Schedule (Robins et al., 1981). The DISC-1 was developed as part of an initiative by the NIMH Division of Biometry and Epidemiology that focused on obtaining a greater understanding of the prevalence of childhood mental disorders (Shaffer et al., 2000). The DISC-1 was designed for use in epidemiological studies and was explicitly tied to the version of the *DSM* being used at the time *(DSM-III*; American Psychiatric Association, 1980).

In 1985, Dr. David Shaffer at the New York State Psychiatric Institute and his colleagues undertook a revision of the interview to (1) improve its reliability for use with children and for use by lay interviewers and (2) provide diagnostic compatibility with the *DSM-III-R* (American Psychiatric Association, 1987) and anticipated *DSM-IV* and ICD-10 criteria (Fisher et al., 1992). Modifications to the DISC have been greatly informed by field-testing conducted as part of a large NIMH-funded multisite study titled the Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) study.

The current version of the DISC, the DISC-IV (Shaffer et al., 2000), assesses approximately 30 diagnoses of childhood and adolescence and is fully compatible with the *DSM-IV* (American Psychiatric Association, 2000) and ICD-10 (World Health Organization, 1993) classification systems. The development of the DISC has expanded to other languages and to alternative administration formats (i.e., computerized).

Structure and Content

The DISC-IV (Shaffer et al., 2000) contains 358 core questions and approximately 1,300 questions that are asked contingent on a child's responses to the core questions. There are two parallel versions of the DISC-IV, the youth version (DISC-Y), to be administered to children of ages of 9–17, and the parent version (DISC-P), to be administered to the parents of children of ages 6–17. There is also an experimental teacher version (DISC-T), which was developed for use in the *DSM-IV* field trials (Frick et al., 1994). The DISC-IV was designed with a primary focus on current psychological functioning. It assesses for symptoms occurring within two overlapping time intervals: the past twelve months and the past four weeks. The DISC-IV contains an optional whole life module designed to measure symptoms occurring as early as age 5. An alternative present-state version has been developed targeting only the four-week time interval (see Shaffer et al., 2000).

The DISC-IV is organized in "diagnostic modules." There are six modules that comprise sets of related disorders. A summary of the modules is provided in Box 11.4. For each diagnosis, DISC-IV is designed to obtain information about the presence of symptoms included in *DSM*

Box 11.4

Organization and Content of the DISC-1V

Module	Disorders Covered
Anxiety	Social Phobia Separation Anxiety Specific Phobia Panic Agoraphobia Generalized Anxiety Selective Mutism Obsessive-Compulsive Posttraumatic Stress
Mood	Major Depressive/Dysthymic Mania/Hypomania
Psychosis	Schizophrenia
Disruptive	Attention-Deficit/Hyperactivity
Behavior	Disorder Oppositional Defiant Disorder Conduct Disorder
Substance	Alcohol Abuse/Dependence
Use	Nicotine Dependence Marijuana Abuse/Dependence Other Substance Abuse/ Dependence
Miscellaneous	Anorexia Nervosa/Bulimia Nervosa Elimination Disorders Tic Disorders Pica

Source: Shaffer et al., 2000.

criteria. If a certain threshold, usually below *DSM* diagnostic threshold, is met, the questions regarding the age of first onset, impairment, and past treatment are asked. (See Box 11.1 for an example of the DISC-IV question format.)

Administration

The DISC-IV was designed to be administered by interviewers without clinical experience after approximately two to six days of training. Use of the computerized DISC-IV is accompanied by less stringent training requirements. Training includes (1) instruction on standard DISC-IV administration procedures, (2) viewing an actual administration of the DISC-IV, and (3) supervised practice in administration with a confederate in a controlled situation.

At the beginning of the interview, the interviewer completes an introductory module consisting of several pieces of demographic information (e.g., age and sex of child) that are necessary to properly administer the interview. The interviewer also establishes a time line with the interviewee to assist in his or her recall for the onset and duration criteria contained in the interview. The time line establishes salient events (e.g., birthdays, vacations, start of the school year, holidays) that occurred in the year preceding the interview. These anchors help the child or parent remember the time frame for diagnostic questions.

The verbal instructions given to the respondent are semi-structured. That is, several points that must be covered are provided, but verbatim instructions are not required. The points include:

1. There are no right or wrong answers. The best answer is the one that tells the most about the child.
2. The informant should try to answer "yes" or "no" to each question.
3. The time frame is within the last year, unless otherwise specified.
4. Some of the questions on the form will be left out.
5. Some questions maybe asked more than once.
6. It is possible to take breaks, if needed.

Unlike the instructions, the administration of the actual DISC-IV questions is quite structured. The questions are designed to be read exactly as written and in the sequence prescribed. Interviewers are explicitly instructed not to make up their own questions or to ask for an example unless it is requested in the inter-

view format. If a respondent does not understand the question, the interviewer should repeat the question, emphasizing the words that seemed to cause confusion. The interviewer is not allowed to answer interpretive questions for the respondent (e.g., "What do you mean by often?" *or* "Is one or two times considered frequent?"). The interviewer is instructed to simply ask the respondent to interpret the question "whichever way s/he thinks is best."

Reliability

The DISC-IV (and other interview schedules) relies heavily on psychometric data derived from prior versions of the instrument. Initial reliability data on the DISC-2.3 was obtained from a series of articles by Shaffer and colleagues (Piacentini et al., 1993; Schwab-Stone et al., 1993; Shaffer et al., 1993). These authors tested the psychometric properties of the DISC-2.3 in a sample of 75 clinic-referred children ages 11–17. In 41 cases, the child and/or parent were reinterviewed one to three weeks later by a second interviewer. There were sufficient cases to calculate the test-retest reliability for five *DSM-III-R* diagnoses (i.e., Attention-Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, Conduct Disorder, Major Depression, and Separation Anxiety Disorder). The kappa statistics for the test-retest agreement are reported in Table 11.2. Also reported in this table are the intraclass correlations, which provide an estimate of the test-retest reliability of the symptom clusters that form the criteria for *DSM-III-R* diagnoses. Importantly, the tendency for parents and children to report substantially fewer symptoms on repeated administrations of structured interview can substantially reduce test-retest coefficients. Therefore, Table 11.2 also includes another index of reliability, the Cronbach's alpha, as an estimate of the internal consistency of the symptom clusters.

On a diagnostic level, all diagnoses showed relatively high test-retest agreement for the parent interview, the child interview, and the combined parent and child interview. When combining a parent and child report, a symptom was considered present if either parent or child endorsed it. The one exception to the generally good reliability was the low reliability of the Oppositional Defiant Disorder diagnosis by child report. Another

TABLE 11.2 Test–Retest and Internal Consistency Estimates from the DISC-2.3

	Kappa	ICC	Alpha
Parent only (n = 39)			
Attention-Deficit Hyperactivity	.55	.87	.87
Oppositional Defiant	.88	.82	.75
Conduct	.87	.86	.56
Major Depression	.72	.82	.88
Separation Anxiety		.77	.61
Child Only (n = 41)			
Attention-Deficit Hyperactivity		.72	.83
Oppositional Defiant	.16	.44	.67
Conduct	.55	.60	.59
Major Depression	.77	.68	.85
Separation Anxiety	.72	.66	.71
Combined Parent and Child (n = 37)			
Attention-Deficit Hyperactivity	.56	.84	
Oppositional Defiant	.59	.65	
Conduct	.50	.80	
Major Depression	.71	.66	
Separation Anxiety	.80	.72	

NOTE: Kappa is the agreement between diagnoses at Time 1 and diagnoses at Time 2 with a 1- to 3-week interval between interviews. ICC is the intraclass correlation between symptoms at Time 1 and Time 2. Alpha is Chronbach's alpha calculated for the symptom at Time 1.SOURCE: M. Schwab-Stone, P. Fisher, J. Piacentini, D. Shaffer, M. Davies, & M. Briggs (1993). "The Diagnostic Interview Schedule for Children-Revised Version (DISC-R): II. Test–Retest Reliability," *Journal of the American Academy of Child and Adolescent Psychiatry*, *32*, 651–657.

exception was the low internal consistency of the Conduct Disorder symptoms. However, the low internal consistency of these symptoms is not surprising for two reasons. First, only three symptoms are required for the diagnosis of Conduct Disorder and most of these symptoms tend to have a relatively low base rate. Second, these symptoms tend to be indicative of discrete—perhaps independent—problem behaviors or delinquent acts, making internal consistency estimates a less than optimal method of determining the reliability of this symptom domain.

Although these reliability data are based on an earlier version of the DISC, the DISC 2.3, Shaffer et al. (2000) provide initial reliability estimates for the DISC-IV from a sample of 84 parents and 82 children, ages 9–17, selected from several outpatient psychiatric clinics. These data were derived from the computer-administered version of the DISC-IV, the C-DISC-IV. Interviews were conducted by lay interviewers with an average retest interval of seven days. The preliminary findings are consistent with the results for the DISC-2.3, with kappa coefficients ranging from .43 (Conduct Disorder) to .96 (Specific Phobia) for the parent report and from .25 (Simple Phobia) to .92 (Major Depressive Episode) for the child report.

Validity

The validity of diagnostic interviews is often assessed by comparing the results of structured interviews to diagnoses made by experienced clinicians. For example, Piacentini et al. (1993) reported moderate to strong agreement between the results of DISC interviews and clinician diagnoses when the Parent DISC-2.3 was used (average kappa = .50) but low agreement based on the Child DISC-2.3 (average kappa = .34). Combining the two interviews gave agreement estimates between those of either informant alone (average kappa = .41). These authors reported that most of the cases with disagreements between clinician diagnoses and the DISC-2.3 were cases that were close to the diagnostic threshold. For example, several disagreements emerged in which children had seven symptoms of ADHD (rather than the required eight symptoms in *DSM-III-R* criteria) and were not given the diagnosis according to the DISC-2.3 but were given the diagnosis of ADHD by the clinician.

However, another study (Lewczyk et al., 2003) found relatively poor correspondence between diagnoses based on the DISC-IV and diagnoses by clinicians. The DISC-IV yielded higher rates of diagnoses on anxiety disorders, ODD, CD, and ADHD, but clinician diagnoses of depression were more common.

Friman et al. (2000) provided a unique investigation of the predictive validity of the DISC by comparing interview data to behavioral observations recorded in a residential treatment program. The researchers examined both convergent and discriminant validity across a lengthy time interval (i.e., one year). Validity data were obtained on 369 children, aged 9–17, who were administered a computerized version of the DISC-Y 2.3, the C-DISC-Y 2.3, upon enrollment in the residential program and at a one year follow-up. Diagnoses of both Oppositional Defiant Disorder and Conduct Disorder were compared to daily observations of disruptive behavior that were coded by program staff and summed to form monthly behavior ratings of both oppositional and conduct problem behaviors. Youth meeting criteria for a DISC-2.3 diagnosis of ODD or CD upon enrollment exhibited significantly greater observed behavioral difficulties on program entry than youth not meeting a diagnosis for either disorder. Furthermore, change in diagnostic status across the two assessment periods predicted changes in observed disrupted behavior across the same time interval. For example, youth who met criteria for an ODD/CD diagnosis at Time

1 but not at Time 2 were characterized by a downward pattern of observed antisocial behavior in the months separating the interviews. This is contrasted with youth whose observed antisocial behavior increased as they moved from no diagnosis at Time 1 to diagnosis at Time 2.

As further evidence for the validity of the DISC interview, Edelbrock and Costello (1988) found strong associations between the diagnoses of ADHD, Conduct Disorder, and Depression/Dysthymia from the DISC-P and the Hyperactive, Delinquent, and Depressed scales of a previous version of the Child Behavior Checklist (Achenbach & Edelbrock, 1983) in a sample of 270 clinic-referred children between the ages of 6 and 16. High rates of agreement were also found between the original DISC-P and the CBCL in another study of 40 psychiatric referrals and 40 pediatric referrals (Costello, Edelbrock, & Costello, 1985). In contrast, the relation between the CBCL and the child version of the DISC tended to be much lower. However, it is impossible to determine whether the low correlations with the DISC-C were due to differences in informants (i.e., parent-completed CBCL and child-respondent DISC) or to differences in the assessment instruments themselves.

Conclusions

Structured diagnostic interviews have become an important part of many clinical assessments of children and adolescents. Like behavior rating scales, diagnostic interviews provide a reliable means of assessing a child's emotional and behavioral functioning. In this chapter, we have attempted to highlight the advantages and disadvantages of using diagnostic interviews. Diagnostic interviews enhance clinical assessments by providing a format for determining how long a child's problems have been occurring, for determining the temporal sequencing of behaviors, and for estimating the degree of impairment associated with a child's emotional or behavioral problems. These important parameters of a child's emotional and behavioral functioning are often not assessed by other assessment modalities. In addition, diagnostic interviews are typically tied to the most recent revisions of the *DSM*, which closely links assessment with this system of classification.

On the negative side, diagnostic interviews are often time intensive, and they typically do not provide any norm-referenced information on a child's functioning above that which is accorded by *DSM* criteria. In addition, diagnostic interviews typically do not include a format for obtaining information from a child's teacher, and their reliability in obtaining self-report information for young children (i.e., below age 9) is somewhat questionable, although pictorial or other formats provide a promising method of enhancing their usefulness in this young age group. As a result of these weaknesses, diagnostic interviews are best used as part of a more comprehensive assessment battery. We have attempted to provide guidelines for their use in this capacity. We have also attempted to provide an overview of the most commonly used diagnostic interviews for assessing children and adolescents, highlighting the major commonalities and differences across interviews. We concluded the chapter with a more detailed discussion of the DISC-IV as an example of a typical diagnostic interview schedule designed for use with children and adolescents.

Chapter Summary

1. Structured diagnostic interviews consist of a set of questions to be asked of a child or adolescent with explicit guidelines on how the youth's responses are to be scored.

2. Most of the commonly used structured interviews are designed to assess diagnostic criteria from one of the recent versions of the *Diagnostic and Statistical Manual of Mental Disorders*. Therefore, if a goal of the evaluation is to make or clarify a *DSM* diagnosis, structured interviews are an important assessment tool.
3. Interviews vary on the degree of structure inherent in the interview format and whether or not the interview assesses only current episodes of the disorders.
4. Structured interviews, like behavior rating scales, obtain detailed descriptions of a child's emotions and behaviors from multiple informants.
5. Unlike rating scales, however, structured interviews allow for the assessment of important parameters of a child's behavior, such as the duration of the behavioral difficulties, the temporal sequencing among problems, and the degree of impairment associated with the difficulties.
6. Most structured interviews are time consuming and are not good for making norm-referenced interpretations.
7. Diagnoses derived from structured interviews should be viewed within the context of other assessment instruments.
8. Child self-report from diagnostic interviews is typically not reliable before age 9.
9. The NIMH Diagnostic Interview Schedule for Children-Version IV (DISC-IV) is a prototypical structured diagnostic interview for use with children and adolescents.
   a. The DISC is highly structured so that it can be administered by trained lay interviewers.
   b. There are child, parent, and experimental teacher versions of the DISC.
   c. The DISC contains 358 core questions and 1,800 follow-up questions that are asked contingent on a child's responses to the core questions.
   d. The DISC is organized in six modules: (1) Anxiety Disorders, (2) Mood Disorders, (3) Psychosis, (4) Disruptive Behavior Disorders, (5) Alcohol and Substance User Disorders, and (6) Miscellaneous Disorders (i.e., Eating Disorders, Elimination Disorders, Tic Disorders).
10. Diagnostic interviews should only be used as part of a more comprehensive assessment battery.