Parent and Teacher Rating Scales

Chapter Questions

- How reliable are parent and teacher ratings of child behavior problems?
- What domains of behavior are assessed by parent and teacher rating scales?
- How are parent and teacher rating scales used in the typical psychological evaluation?
- Why are teachers important sources of information about a child's emotional and behavioral adjustment?
- To what extent do parents and teachers agree in their ratings of children and adolescents?
- What factors influence this agreement?
- What factors should play a role in the use and selection of parent and teacher rating scales?

Evaluating Children via Parent Ratings

It has long been recognized that children are often less-than-accurate reporters of their own behavior. Furthermore, children may not have sufficient reading or oral expression skills for self-report purposes (Lachar, 1990). Problems with underreporting and response sets have always been well-recognized by clinicians and, to some extent, have been documented by research (see previous chapter). These concerns about child self-reports have undoubtedly contributed to the popularity of parent rating scales. Furthermore, the parental perspective is often invaluable when conceptualizing a case; that is, because children are often referred for an evaluation because of a parent's concerns, information on the parent's perspective of a child's problems is critical.

Parent ratings of child behavior possess additional advantages, including brevity and cost efficiency (Hart & Lahey, 1999). The time-efficient nature of parent ratings makes it easy to collect additional information about child behavior. Given the importance of parental influence on child behavior, parental perceptions of behavior should routinely be collected in clinical assessments.

Today, the commonly used parent rating scales routinely provide a broad coverage of problems. For example, while the unstructured interview may allow the clinician to carefully evaluate a specific area of functioning, other important behavioral problems or areas of concern may be missed (Witt, Heffer, & Pfeiffer, 1990). Parent or other caretaker ratings also foster objectivity and clarity in the assessment process. Because of the behavioral specificity of the typical item content of these measures, parents are required to operationally define their concerns and provide specific and objective ratings of hyperactivity, depression, nervousness, and the like (Witt et al., 1990).

All rating scales, including parent ratings, can be influenced by bias and rater response sets (Witt et al., 1990). Even biased reporting, however, can be of value. If, for example, parent ratings provide very different results when compared to the ratings of others, the clinician can develop some important insights into the child's family functioning. If a child's father rates his son as having significantly more behavioral problems than the mother, the clinician can explore the dynamics behind the ratings. A straightforward explanation may be that the father is doing the majority of the child care. This information could be important to acquire if the presumption had been that the child's mother was providing most of the caretaking.

Factors Influencing Parent Ratings

As discussed in more detail in Chap. 15, research has indicated that parental, specifically maternal, distress may influence the ratings of child functioning in a negative way. Although the issue of whether or not maternal distress is directly influential on the ratings of a child's symptoms is far from settled, it stands to reason that stressful home environments would be positively correlated with parent reports of child symptoms.

The construct being evaluated and the child's developmental level are two additional factors that may influence parents' reports. Teachers have traditionally been considered superior to parents as reporters of a child's ADHD-like symptoms (Loeber, Green, & Lahey, 1990; Loeber et al., 1991; Tripp, Schaughency, & Clarke, 2006). However, parents are still considered necessary and useful in providing information about inattention and hyperactivity (Tripp et al., 2006) and in documenting the effects of treatment for ADHD (Biederman, Gao, Rogers, & Spencer, 2007). In addition, parents may be in a unique position to understand the antecedents of a child's disruptive behaviors, as they can observe their child more closely than a teacher who works with several children simultaneously. Parents have also been discussed as particularly important observers and informants of child anxiety and depression (Klein, Dougherty, & Olino, 2005; Silverman & Ollendick, 2005).

As reviewed by De Los Reyes and Kazdin (2005), research has reached mixed conclusions about the degree to which the child's age influences agreement in ratings across informants. Parents, in particular, may be useful informants of a child's functioning throughout childhood and adolescence, although there

may be discrepancies between their reports and the reports of others. The information used in conjunction with parent reports may vary with age. More specifically, parents are obviously vitally important sources of information for young children in such areas as conduct problems, whereas the children themselves would not be reliable (and thus, not valid) informants. Teachers, however, could offer useful perspectives of the young child's social, academic, and behavioral functioning. For adolescents, parents may still provide valid and useful information, but their knowledge of the child's conduct and behavior problems may be more limited, as the behaviors may sometimes occur outside of the parent's awareness. The adolescent – provided that he or she is willing to provide such information – would be the most knowledgeable informant of these behaviors, and the teacher's contribution would also presumably diminish.

Finally, parent ratings are more likely to attribute the child's problems to dispositional factors in the child, whereas youth self-reports are more likely to indicate the family environment as a factor in need of intervention (see De Los Reyes & Kazdin, 2005). Thus, informants (including parents, teachers, and children) may base their ratings of a child's functioning on the attributions that they make regarding the genesis and maintenance of the child's problems.

That parent ratings may be influenced by factors that are not necessarily directly tied to the child's actual functioning does not render parent ratings questionable. Instead, it calls to mind the many potential variables to consider in understanding the child's presenting problems – an understanding that is critical for case conceptualization and subsequent recommendations for intervention.

Evaluating Children via Teacher Ratings

Although teachers have traditionally been considered an important source of information about children's academic performance, they have not often been used in the assessment of children's behavioral and emotional functioning. However, knowing how a child behaves in the classroom is important for several reasons. First, school is a setting in which the child spends several hours a day. Therefore, a child's adjustment to the school setting can have a dramatic impact on his or her overall psychological functioning. Second, the multiple demands of the school environment (e.g., to stay seated, to follow the demands of adults, to interact with classmates) present many challenges to the child— challenges that may not be present in other settings. Third, the demands of the school setting change as a child progresses through school (e.g., demands for organization, the importance of social acceptance). Therefore, understanding school-related problems that are unique to a given period can provide clues to specific problems in adaptation that a child or adolescent might experience.

On the basis of these considerations, there is increasing interest in assessing a child's behavioral and emotional functioning in the school setting. Given the many advantages of behavior rating scales, such as time-efficiency and objectivity, it is not surprising that the primary assessment instruments for children's school behavior have been teacher-completed behavior rating scales. In addition to suggestions for appropriate use of rating scales in general discussed previously, there are several considerations for interpreting information from teachers that warrant special attention.

Factors Influencing Teacher Ratings

As described above for parents, the usefulness of teacher information may vary depending on what type of behavior is being assessed. Teachers are often considered the best source of information about a child's attention problems and overactivity because they have the opportunity to observe the child in a situation that demands sustained attention and inactivity. In contrast, teachers' ratings tend to be less useful in assessing many types of antisocial behavior that are unlikely to occur in the school environment (e.g., setting things on fire, being cruel to animals) or for internalized types of problems that may not be readily observable in the classroom setting (Loeber et al., 1991).

The usefulness of teacher information may also vary according to the age of the child. Children in early elementary school frequently have one teacher who observes a child across several class periods, if not the entire school day. In contrast, high school teachers frequently have students for one class period during the day. Therefore, the usefulness of information may decrease as a child advances in school and contact with any single teacher decreases (Edelbrock et al., 1985).

A final issue in interpreting teacher rating scales is understanding the frame of reference or standard used by teachers. As discussed previously (e.g., Piacentini, 1993), a number of characteristics of a rater can influence his or her judgment of the intensity, quality, and/or frequency of a child's behavior. In the case of teacher ratings, a characteristic of teachers that can influence their ratings is their experience with many children of the same age. Experience allows the teacher to make some internal normative comparison of a child's behavior with the behavior of other children the teacher has taught. This internal norm is a double-edged sword. It often gives the teacher a unique perspective of knowing both the individual child and the behaviors that are age-appropriate. However, some teachers, such as teachers who work in special education classrooms, may have a skewed base of comparison that could influence their ratings. That is, their ratings of a child's behavior may be influenced by a comparison of the child with other disturbed children.

Despite these cautions and limitations, we feel strongly that teacher ratings are an essential element of a comprehensive clinical assessment of children's behavioral and emotional functioning. Carlson and Lahey (1983), in an early review of teacher ratings, reported that most of the teacher rating scales available at that time suffered from significant psychometric problems in development and inadequate norming. As a result, the available scales were severely limited in their usefulness for clinical evaluations. Fortunately, since that 1983 review, there have been numerous advances in the teacher rating scales and the emergence of new scales, with many of the inadequacies of earlier scales eliminated or greatly reduced.

Overview of Omnibus Parent and Teacher Rating Scales

Parent and teacher rating scales are not interchangeable and, with the seemingly exponential growth of such instruments, psychologists have to make many decisions about the utility of various measures. This chapter attempts to aid the clinician in decision making by providing an overview of the variety of scales available, with particular attention devoted to defining the strengths and weaknesses of each measure. Writing such a chapter requires selectivity. Hence, if a scale is not

mentioned in this chapter, it should not be construed as a judgment of the quality of the scale. As with our chapter on self-report rating scales, we have attempted to review those instruments that are widely used and/or part of a long-standing system of rating scales used for child and adolescent assessment. This broad overview of the various scales is not designed to replace information provided in the technical manuals that accompany these instruments, to be reviewed by any user of the scales. Optimally, however, the principles applied to evaluating parent and teacher rating scales in this chapter can be used by psychologists to evaluate other scales as well.

The parent and teacher rating scales reviewed in this chapter are highlighted in Tables 7.1 and 7.2, respectively. Commonly used omnibus measures that assess many different domains, as opposed to single construct measures, are the focus. Although these scales are discussed in isolation to balance clarity and specificity, it should be recalled that they are often part of larger multimethod assessment methods that are discussed in various chapters of this book. The integration of components and information from different informants and methods is discussed in the context of interpretation in Chap. 15 and in subsequent chapters that address specific syndromes.

Behavior Assessment System for Children, 2nd Edition (BASC-2)

Parent Rating Scale (PRS)

The BASC-2 Parent Rating Scale (BASC-2-PRS; Reynolds & Kamphaus, 2004) is part of the larger BASC system. The PRS was published concurrently with the SRP (discussed in Chap. 6) and TRS (see below) as well as other components of the BASC assessment system (Reynolds & Kamphaus, 2004). The PRS has three forms composed of similar items and scales that span the preschool (2–5 years), child (6–11 years), and adolescent (12–21 years) age ranges. The PRS takes a broad sampling of a child's behavior in home and community settings.

Content

As with its predecessor, the BASC-2-PRS was developed using both rational/theoretical and empirical means in combination to construct the individual scales. The benefit of this approach is that the resulting scales have relatively homogenous content. The uniqueness of the scales was also enhanced by not including items on more than one scale. Table 7.3 provides item examples for each scale. There are four composites: Externalizing Problems, Internalizing Problems, Adaptive Skills, and a Behavioral Symptoms Index that includes some internalizing and externalizing scales (i.e., Atypicality, Attention Problems, Hyperactivity, Aggression, Depression, and Withdrawal).

Two types of scales are included at each age level: clinical and adaptive. Clinical scales of the PRS are designed to measure behavior problems much like other measures discussed below in that behavioral excesses (e.g., hitting others) are the focus of assessment. The PRS also includes critical items that are thought to warrant follow-up or clinical attention in their own right. These items (e.g., "Has a hearing problem.") are not necessarily indicative of the most severe pathology; instead, they may be worthy of further questioning or recommendations by the clinician. The adaptive scales measure behaviors (e.g., compliments others) or skills that are associated with

TABLE 7.1   Overview of Parent Rating Scales

Inventory	Ages	Content	Reading Level	Time to Administer	Norm Sample	Reliability	Validity	MI
Behavior Assessment System for Children, 2nd edition (BASC-2; Reynolds & Kamphaus, 2004)	2–5; 6–11; 12–21	134–160 items; Behavioral Symptoms Index, Adaptive Skills, Internalizing Problems; 3 validity scales; 9 clinical scales, 5 adaptive scales; 7 optional content scales	4th grade	10–20 min	E	Internal consistency (good); 1–7 week test–retest (good); Interrater (moderate to high)	Moderate to high with similar tools; Support for factors	E
Achenbach Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000; Achenbach & Rescorla, 2001)	1½– 5; 6–18	100–113 items; Internalizing, Externalizing, Competence; 8 problem scales; 6 DSM-oriented scales	Not given	15–20 min	G	Internal consistency (good for school age version); 1 week test–retest (good); interrater (moderate for preschool version; good for school age version)	Good validity (clinical vs. nonclinical groups); limited info. on relations with other tools; limited info. on current version	E
Child Symptom Inventory-4 (Gadow & Sprafkin, 1998) ESI/CSI/ASI	3–5; 5–12; 12–18	97 items corresponding to 17 disorders	Not given	10–15 min	F	Internal consistency (moderate to good); Test-retest (moderate to good); content (moderate to good)	Content derived from DSM criteria; Evidence of construct validity (good)	G

TABLE 7.1 (Continued)

Inventory	Ages	Content	Reading Level	Time to Administer	Norm Sample	Reliability	Validity	MI
Conners 3rd Edition, Parent Rating Scale (Conners-3; Conners, 2008a)	6–18	110 items; 43-item short form; 3 validity scales; 6 empirical scales; 1 rational scale; 5 *DSM-IV-TR scales*	4th–5th grade	20 min	E	Internal consistency, 2–4 week test–retest, and interrater reliability all good	Good evidence of criterion-related and differential validity	E[a]
Personality Inventory for Children, 2nd edition (PIC-2; Lachar & Gruber, 2001)	5–19	275 items, 96 item short form; 9 clinical scales (several sub-scales); 3 validity scales	4th grade	40 min	E	Internal consistency; test–retest; interrater; Good reliability, somewhat lower for subscales	Validity (based on teacher and self-reports of same system); lack of research on relations with other criteria	G

NOTE: Norm Sample evaluated as E = Excellent, G = Good, F = Fair; MI = Degree of enhancement of multi-informant assessment; E = Excellent, G = Good; ESI = Early Symptom Inventory; CSI = Child Symptom Inventory; ASI = Adolescent Symptom Inventory.
[a] We view the Conners-3 as providing an excellent means of assessment of externalizing problems; however, if an extensive assessment of internalizing or learning difficulties is desired, the Conners Comprehensive Behavior Rating Scales (CBRS; Conners, 2008b) is recommended.

TABLE 7.2   Overview of Teacher Rating Scales

Inventory	Ages	Content	Reading Level	Time to Administer	Norm Sample	Reliability	Validity	MI
Behavior Assessment System for Children, 2nd edition (BASC-2; Reynolds & Kamphaus, 2004)	2–5; 6–11; 12–21	100–139 items; Behavioral Symptoms Index, Adaptive Skills, Externalizing Problems, Internalizing Problems, School Problems; 3 validity scales; 10 clinical scales, 5 adaptive scales, 7 optional content scales	Not given	10–20 min	E	Internal consistency (good); 1–9 week test–retest (good); interrater reliability (good with exception of Somatiz)	Moderate to high relations with similar tools; Support for factors	E
Achenbach Teacher Rating Form (TRF; Achenbach & Rescorla, 2000; Achenbach & Rescorla, 2001)	1½–5; 6–18	100–113 items; Internalizing, Externalizing, Adaptive Functioning; 8 problem scales; 6 DSM-oriented scales	Not given	15–20 min	G	Internal consistency (good); 7 month test–retest (moderate); interrater (low to mode-rate)	Good validity (clinical vs. nonclinical groups); limited info. on relations with other tools	E
Child Symptom Inventory-4 (Gadow & Sprafkin, 1998) ESI/CSI/ASI	3–5; 5–12; 12–18	77 Items corresponding to 13 disorders	Not given	10 min	F	Test–retest; limited evidence available	Content derived from DSM criteria; Classification rates match well with epidemiology	E
Conners 3rd Edition, Teacher Rating scale (Conners-3-T; Conners, 2008a)	8–18	115 Items; 39-item short form; 3 validity scales, 6 empirical scales, 1 rational scale, 5 *DSM-IV-TR* symptom scales	4th–5th grade	20 min	G	Internal consistency and test–retest reliability (good). Interrater reliability (moderate to good)	Good evidence of criterion-related and differential validity	E[a]

(Continues)

TABLE 7.2 (continued)

Inventory	Ages	Content	Reading Level	Time to Administer	Norm Sample	Reliability	Validity	MI
Student Behavior Survey (SBS; Lachar, Wing-enfeld, Kline, & Gruber, 2000)	5–18	102 items; Academic Resources, Health Concerns, Emotional Distress, Unusual Behavior, Social, and Aggression composites	Not given	10–15 min	E	Good reliability (internal consistency; test–retest)	Good convergent validity for most scales; DBD scales do not overlap fully with DSM criteria; Limited validity evidence for some scales	F

NOTE: Norm Sample evaluated as E = Excellent, G = Good; MI = Degree of enhancement of multi-informant assessment; E = Excellent, G = Good, F = Fair; ESI = Early Symptom Inventory; CSI = Child Symptom Inventory; ASI = Adolescent Symptom Inventory; DBD = Disruptive Behavior Disorder.
[a]We view the Conners-3 as providing an excellent means of assessment of externalizing problems; however, if an extensive assessment of internalizing or learning difficulties is desired, the Conners Comprehensive Behavior Rating Scales (CBRS; Conners, 2008b) is recommended.

TABLE 7.3   BASC-2-PRS Scale Definitions and Key Symptoms as Indicated by Items with the Highest Factor Loadings Per Scale

Activities of Daily Living	Skills associated with performing everyday tasks; "Acts in a safe manner", "Sets realistic goals", "Attends to issues of personal safety"
Adaptability	Ability to adapt to changes in the environment; "Adjusts easily to new surroundings", "Adjusts well to changes in family plans", "Recovers quickly after a setback"
Aggression	Tendency to act in hostile or threatening manner; "Is cruel to others", "Loses temper too easily", "Annoys others on purpose"
Anxiety	Tendency to be nervous, fearful, or worried; "Worries about making mistakes", "Worries about what other children think", "Is nervous"
Attention Problems	Tendency to be easily distracted or have difficulty concentrating; "Has a short attention span" "Listens carefully" (reverse scored); "Is easily distracted"
Atypicality	Tendency to behave in odd manner; "Acts strangely." "Says things that make no sense", "Seems out of touch with reality"
Conduct Problems	Tendency to engage in antisocial and rule-breaking behavior; "Breaks the rules"; "Deceives others"; "Gets into trouble"
Depression	Feelings of unhappiness, sadness, or stress; "Is negative about things", "Says 'I don't have any friends'", "Seems lonely"
Functional Communication	Ability to communicate ideas and express oneself clearly; "Communicates clearly", "Responds appropriately when asked a question", "Accurately takes down messages"
Hyperactivity	Tendency to be overly active and act without thinking; "Acts out of control", "Interrupts others when they are speaking", "Disrupts other children's activities"
Leadership	Possessing skills needed to accomplish goals, ability to work with others; "Gives good suggestions for solving problems", "Is creative", "Is a 'self-starter'"
Social Skills	Having the skills necessary to interact successfully with peers and adults; "Encourages others to do their best", "Offers to help other children", Congratulates others when good things happen to them"
Somatization	Tendency to be sensitive to, and complain about, minor physical ailments; "Complains about health"; "Gets sick"; "Complains of being sick when nothing is wrong"
Withdrawal	Tendency to avoid others; "Makes friends easily" (reverse scored), "Avoids other children", "Quickly joins group activities" (reverse-scored)

NOTE: Adapted from Tables 7.7 and 10.3 of the BASC-2 manual (Reynolds & Kamphaus, 2004).The clinical norms may be especially important when assessing a child in a residential setting to be able to compare him or her to others with relatively severe difficulties. That is, it may be understood that a child is functioning poorly compared to most other children his/her age (i.e., elevations on general norms), but it may be informative to consider how the child functions (e.g., "How severe are his conduct problems?") in comparison to other children with emotional and behavioral difficulties for treatment planning purposes.

good adaptation to home and community (see Table 7.3).

Each of the parent forms of the BASC-2 includes seven optional content scales: Anger Control, Bullying, Developmental, Social Disorders, Emotional Self-control, Emotional Self-control, Executive Functioning, Negative Emotionality, and Resiliency. As with the BASC-2-SRP (see Chap. 6), the content scales for the BASC-2-PRS were constructed via theoretical and empirical methods. These scales were developed based on current theoretical perspectives about important domains of youth functioning (Reynolds & Kamphaus, 2004). There exists very little research on these scales, yet their labels and item content are intriguing and warrant further investigation of their reliability, validity, and clinical utility. Initial analyses indicate that the PRS content scales possess adequate (i.e., 0.70 and higher) internal consistencies (Reynolds & Kamphaus, 2004).

Administration and Scoring

The PRS uses a four-choice response format (i.e., never, sometimes, often, almost always) with no space allowed for parent elaboration. According to the authors, the scale takes about 10–20 min for parents to complete.

A variety of derived scores and interpretive devices are offered. Linear T-scores are available for all scales and composites, meaning that the original distributions for these indices in the norming sample were maintained. Other scores available include percentile ranks, confidence bands, and statistical methods for identifying high and low points in a profile. Both hand-scoring and computer entry scoring are available for the PRS.

Norming

PRS provides three norm-referenced comparisons depending on the questions of interest to the clinician. Some examples of

questions and their implications for norm

Is Daniel inattentive in comparison to children of the same age?	General national sample
Is Daniel inattentive in comparison to a large sample sample of children who are currently diagnosed and receiving treatment?	Clinical national sample
Is Daniel inattentive in comparison to boys of the same age?	Male national

group selection include the following.

Question Norm Group

These various norm-referenced comparisons are more than are typically offered for such scales. The general national norming sample is advised as the starting point for most purposes (Reynolds & Kamphaus, 2004). Of course, as just noted, depending on the question, the clinician may opt for gender-specific or clinical norms. Gender-specific norms may be useful in trying to convey a child's current level of functioning to others, such as parents. In other words, one might present the child's scores relative to the general population and then emphasize how the child compares to other boys/girls on areas of concern in order to provide a more complete picture. However, too much information may cause confusion for some parents. Gender-based norms may also help answer some specific research questions (i.e., correlates of inattention and hyperactivity among girls and among boys).

The general norm sample for the PRS included 1,200 preschoolers, 1,800 children, and 1,800 adolescents. Cases were collected at test sites in 40 states. Across age groups, the PRS sample closely matches US Census statistics (Current Population Survey, 2001) in terms of sex, race/ethnic-

ity, and socioeconomic status (SES). The norming sample also represents a good fit to census data on geographic region (see Reynolds & Kamphaus, 2004 for more details). The clinical norming sample for the PRS included responses from 1,975 parents, with most cases being classified as having a learning disability or ADHD.

Reliability

The median reliability coefficients provided in the manual suggest good evidence for the reliability of the individual scales and composites. All scales and composites have median reliability estimates of 0.80 and above, with the exception of the Activities of Daily Living and Atypicality Scales. The BASC-2 manual also provides information on 1–7-week test-retest reliability and inter-rater reliability between parents. Test-retest reliability coefficients were 0.70 and higher, with the exception of Depression for the preschool form which was .66. Interrater reliability was generally good, with coefficients in the same range, with the exception of Aggression on the preschool and child forms (i.e., 0.59 and 0.58, respectively) and Anxiety on the preschool form (.56).

Validity

The PRS appears to have a broad content coverage. The PRS assesses a variety of externalizing behavior problems (McMahon & Frick, 2002) and has an expanded assessment of adaptive skills. In addition, the PRS enjoys considerable factor analytic support for a three factor model consisting of externalizing problems, internalizing problems, and adaptive skills. The strongest measures of the externalizing factor are the Aggression, Conduct Problems and Hyperactivity. The Internalizing factor is marked by loadings by the Atypicality, Depression, Anxiety, Somatization, and Withdrawal scales. Adaptive skills scales that load highly on this factor include Activities of

Daily Living, Functional Communication, Leadership, and Social Skills.

Some of the secondary loadings for the scales may also have implications for interpretation. Specifically, the factor-analytic data suggest that the following profiles are reasonable:

- Poor Adaptive Skills with Attention Problems
- Good Adaptive Skills with Anxiety
- Internalizing Problems accompanied by Poor Adaptability

Criterion-related validity analyses produced consistent associations between the PRS and other parent rating scales. This pattern particularly holds for the composites and for the externalizing problem scales. Generally, the internalizing problem scales (e.g., Anxiety, Depression, Somatization) show moderate correlations with analogous scales from other measures (see Reynolds & Kamphaus, 2004).

Interpretation

The same logical interpretive steps that were outlined for the BASC-SRP (discussed in Chap. 6) also apply to the BASC-PRS. Briefly, the clinician should:

1. Assess validity using validity indexes and informal means (e.g., inspect for a high number of items with no response).
2. Inspect critical items and follow-up as appropriate.
3. Interpret scores on scales and composites, with particular attention to elevations (T-scores of 65–70 and higher) on clinical scales and low scores (T-scores of 35 and below) on adaptive scales.
4. Attend to items that appeared to have led to scale elevations (or low adaptive scores).
5. Integrate score with information from other informants.

6. Integrate data with information from other assessment tools (e.g., interview, behavioral observations, intelligence testing).
7. Set objectives for treatment/intervention

As was the case with the SRP, we again recommend a focus on interpretation at the scale level, as the reliabilities of the PRS scales are generally good, and elevations on scales are more specifically informative than would be the case for elevated composite scores.

The original PRS enjoyed a great deal of research support and research use. There is quite limited information available to date on the BASC-2-PRS. Nevertheless, the combined rational and empirical approach to scale development has intuitive appeal for use in clinical situations. Clinicians are still urged to keep abreast of the research literature discussing the strengths and limitations of any assessment tool.

Strengths and Weaknesses

The BASC-2-PRS has a number of apparent strengths and weaknesses as follows:
The strengths of the PRS are:

1. Good psychometric properties based on the information reported in the BASC-2 manual.
2. A variety of scales that may be useful for differential diagnosis (e.g., Attention Problems vs. Hyperactivity, and Anxiety vs. Depression).
3. The availability of validity scales and critical items.
4. An expanded group of norm-referenced adaptive scales

Among the weaknesses of the PRS are:

1. A response format that does not allow parents to provide additional detail about their responses

2. Cross-informant and cross-scale comparisons not as readily made as on other measures, as different forms (e.g., parent vs. self-report) include different item content and scales
3. Limited research on the latest edition of the PRS.

Teacher Rating Scale (TRS)

The BASC-2 Teacher Rating Scale (BASC-2-TRS; Reynolds & Kamphaus, 2004) allows the clinician to gather information on a child's observable behavior from the child's teacher and place that information in the context of other information obtained in the overall BASC system (e.g., self-report scale, parent rating scales, classroom observation system). As with the PRS, there are three forms of the BASC-2-TRS: preschool (ages 2–5), child (6–11), and adolescent (12–21). The three forms contain behavioral descriptors that are rated by the teacher on a four-point scale of frequency, ranging from "Never" to "Almost Always." The three forms have 100 items for the preschool version and 139 for both the child and adolescent versions.

Content

As with the other BASC-2 rating scales, the items of the BASC-2-TRS were chosen to measure multiple aspects of a child's personality and behavior. The TRS includes both positive (adaptive) and pathological (clinical) dimensions. For the most part, the BASC-2-TRS has maintained the content areas of the original BASC. The only scale additions to the current version of the TRS were the Functional Communication scale for all age groups and the Adaptability scale for the adolescent version. The BASC-2-TRS consists of five composites (i.e., Behavioral Symptoms Index, School Problems, Externalizing Problems, Internalizing Problems, Adaptive Skills)

across all age ranges, with 11 scales in the preschool version and 15 scales in the child and adolescent versions. The scales grouped into the composites – except for the Behavioral Symptoms Index which includes the Hyperactivity, Atypicality, Depression, Aggression, Attention Problems, and Withdrawal scales – are provided in Table 7.4. The TRS also has the same optional content scales as those provided for the PRS (see above). Because these are a new feature of the BASC-2, very limited information is available on their psychometric properties or clinical utility.

The content coverage of the BASC-2-TRS scales has several unique features relative to other teacher rating scales. First, it provides comprehensive coverage of several areas of adaptive behavior. Second, the current version of the TRS continues the strategy of including separate scales for motor hyperactivity and attention problems, which aids in the differentiation of subtypes of Attention-Deficit Hyperactivity Disorder (Vaughn, Riccio, Hynd, & Hall, 1997). Third, there are separate scales for anxiety, depression, and withdrawal, which aid in the assessment of emotional difficulties. Fourth, the BASC-2-TRS

TABLE 7.4   Composites and Scales of BASC-2-TRS

Composite	Scales
Externalizing Problems	Aggression Hyperactivity Conduct Problems
Internalizing Problems	Anxiety Depression Somatization
School Problems	Attention Problems Learning Problems
Adaptive Skills	Adaptability Functional Communication
	Leadership
	Social Skills
	Study Skills

includes items that screen for learning problems that often accompany emotional and behavioral problems in children.

Administration and Scoring

The BASC-TRS takes approximately 10–20 min to complete. The cover of the record provides instructions to the teacher for completing the form and space for recording background information about the child and teacher (e.g., age, gender, type of class, length of time in class). Both hand scoring and computer scoring are available. Norm tables in the BASC manual are provided so that any of four sets of norms can be used: general, male, female, and clinical (see above for discussion of the uses of these different types of norms). Both T-scores and percentile ranks are listed for each set of norms, with linear T-scores again being utilized for the TRS. As with the other BASC-2 rating scales, the BASC-2-TRS scoring sheet highlights critical items (e.g., "I want to kill myself") that are clinically important or that warrant further follow-up.

Norming

The norming group included 1,050 preschoolers, 1,800 children (ages 6–11), and 1,800 adolescents (ages 12–21) with equal sex distributions in all age groups. Respondents were recruited from sites throughout the USA. As described previously, the sampling procedures for obtaining the normative sample were designed to closely mirror US Census statistics in terms of race/ethnicity, SES, and geographic region, and this goal was accomplished (see Reynolds & Kamphaus, 2004). Details regarding the 1,779-member clinical sample for the TRS are also provided in the BASC-2 Manual.

Reliability

The manual for the BASC-2-TRS (Reynolds & Kamphaus, 2004) provides evidence

on three types of reliability: internal consistency, test-retest reliability, and interrater reliability. With very few exceptions, the scales of the BASC-2-TRS proved to be quite reliable in the normative sample. More specifically, internal consistency coefficients tended to average well above 0.80 across all age groups, and all were 0.75 or higher. Similarly, test-retest reliability over one to nine weeks was high, with the exception of the Anxiety scale, with coefficients ranging from 0.64 for the adolescent version to 0.77 for the adolescent version. Still, these coefficients are adequate. Finally, the consistency of ratings between two teachers was tested in samples of preschool-age children (n = 74), school-age children (n = 38), and adolescents (n = 58), with moderate reliability estimates emerging across age group samples (median coefficients of 0.69, 0.60, and 0.52, respectively). Correlation coefficients tend to be somewhat higher for externalizing than for internalizing problems consistent with past research (Achenbach, McConaughy, & Howell, 1987). It is also worth noting that the coefficients tended to be lower for adolescents, which may be associated with the limited contact an individual teacher may have with students of that age group. Interrater agreement for Somatization (r = 0.25), Withdrawal (r = 0.24), and Atypicality (r = 0.31) was particularly low for teacher ratings of adolescents.

Validity

The TRS is closely, but not exactly, aligned with the item content of the PRS. However, the TRS has additional scales (i.e., Study Skills, Learning Problems) that seem particularly valid for use with a teacher rating scale. The BASC-2 manual provides factor analytic support for the construct validity of the scales and composites of the TRS. In addition, initial research on the TRS shows generally high correlations with analogous scales from other teacher rating scales. However, the correspondence to analogous scales is somewhat lower for internalizing types of problems than for the indices of externalizing problems (see Reynolds & Kamphaus, 2004). One notable finding was the lack of a correlation (i.e., r = .03) between the TRS Somatization scale and the Somatic Complaints scale of the Achenbach Teacher Report Form. A significant limitation of the latest version of the TRS is the very limited research on its validity and utility outside of what was conducted by the developers.

Interpretation

The BASC-2-TRS includes validity scales that provide a useful and efficient first point of interpretation. More specifically, it contains a "fake bad" index (F), which helps to assess the possibility that a teacher rated a child in an overly negative pattern. Therefore, interpretation of this scale, in particular, should be the first step in the interpretative process, keeping in mind that a high score on the F index may actually indicate significantly problematic functioning. Therefore, this validity index should be interpreted in the context of other assessment data. The Consistency Index and the Response Pattern Index available for the TRS (as are available for the PRS and SRP) provide another initial point of interpretation. Critical items should be reviewed promptly, because these items tend to be clinically important indicators that deserve careful follow-up assessment.

The reliability estimates at the scale level of the TRS are good; therefore, we again recommend focusing interpretation mainly at the scale, rather than composite, level since more specific information is available through the TRS scales. Interpretations at the item level must be made quite cautiously because of the low reliability of individual items. It is often informative to see which items led to a child's or adolescent's elevation on a given clinical scales. For example, it may be informative for a child with an elevation on the Adaptabil-

ity scale to determine if this elevation was largely due to problems specifically within the interpersonal domains or due to more general problems in adapting to changes in routine. Finally, interpretation at the scale level for the parent form is a viable early step in interpretation (see above); therefore, interpretation at the scale level of the TRS facilitates integration of information across parent and teacher ratings. In addition, considering individual items within elevated scales on both rating forms may help determine the source of consistencies and inconsistencies across parent and teacher ratings, further informing case conceptualization and recommendations.

Strengths and Weaknesses

Like its predecessor and its companion parent rating scale, the BASC-2-TRS has a number of strengths and weaknesses. Notable strengths include:

1. It is part of a multimethod, multi-informant system that aids in a comprehensive clinical evaluation with item content that covers important problematic and adaptive domains of classroom behavioral and emotional functioning.
2. The assessment of adaptive functioning is enhanced on this version of the TRS.
3. The preschool age range of the BASC-2-TRS is expanded from the age range available from the original TRS.
4. The BASC-2-TRS has a large nationwide normative sample on which norm-referenced scores are based, allowing for one to confidently make many norm-referenced interpretations of scores.

Weaknesses of the BASC-2-TRS include:

1. The limited research base for the current edition.
2. The relatively lower correlations between internalizing scales on the TRF

and analogous scales from other teacher rating scales.
3. The different item content across informants, especially with the SRP, makes integration of BASC-2 information somewhat more challenging.

A sample case using the BASC-2-PRS and BASC-2 TRS is provided in Box 7.1.

ACHENBACH SYSTEM OF EMPIRICALLY BASED ASSESSMENT (ACHENBACH & RESCORLA, 2000, 2001)

Parent Report: Child Behavior Checklist (CBCL)

The Child Behavior Checklist (CBCL; Achenbach, 2001) and its predecessors have a long history of prominence in child assessment. The CBCL scale is the product of an extensive multiple-decade research effort, and it has a distinguished history of research usage. The current version of the CBCL is much like its predecessors with some item changes, response format changes, and the introduction of DSM-Oriented scales (see below; see Achenbach & Rescorla, 2001).

The CBCL is part of an extensive system of scales including teacher rating (TRF), self-report (YSR), and classroom observation measures. The newest version of the CBCL (Achenbach & Rescorla, 2001) has two separate forms: one for ages 1½–5 and the other for children of ages 6–18.

The development of the CBCL and its revisions reflects the author's belief that parent reports are an important part of any multi-informant system of child evaluation. In Achenbach's (1991) own words:

> Parents (and parent surrogates) are typically among the most important sources of data about children's competencies and problems. They are usually the most knowledgeable about their

Box 7.1

Sample Case Using the BASC-2 PRS and TRS

Johnny was referred for a psychological evaluation by his mother because of his academic difficulties and distractibility. He is a 12-year-old sixth-grader. Johnny has reportedly had trouble completing schoolwork since the first grade. Johnny was retained in the second grade because of poor work completion and academic progress. He was reportedly placed in special education in the third grade in a learning problem class. He was reportedly placed in a regular class again in the fourth grade. A psychological evaluation conducted at the end of the fourth grade resulted in the conclusion that he was a "slow learner," according to his mother. This year, he is again having difficulty concentrating, and he rarely completes his assignments. Academic progress is still unacceptable to his mother and teachers.

Johnny's developmental milestones were slightly delayed. He still reportedly has problems drawing and using scissors. Last year, he was diagnosed with juvenile diabetes, according to his mother. His family psychiatric history is significant for depression (mother), and Johnny's father reportedly had difficulty academically when he was in school.

Johnny was exceedingly cooperative during the evaluation. He addressed the examiner politely and would occasionally answer questions by saying "yes sir." He had considerable difficulty comprehending instructions on an intelligence test. He was reluctant to admit to not knowing an answer, and he worked extremely slowly. The test session had to be conducted over two days because of his slow response style.

He responded impulsively to items on occasion. He also wiggled in his seat and frequently looked around the room and asked questions of the examiner. His full scale IQ score in this evaluation was 85. His achievement test scores were similarly in the Low Average range.

His BASC-2-PRS (completed by his mother) and BASC-2-TRS (completed by his current teacher) results are highly consistent with background information.

Hyperactivity

$T = 63$ (parent report)
$T = 60$ (teacher report)

The reports of relatively mild levels of hyperactivity are consistent with Johnny's history which indicates no history of disruptive or impulsive behaviors.

Attention Problems

$T = 76$ (parent)
$T = 77$ (teacher)

The parent and teacher reports of significant attention problems (e.g., has trouble concentrating, is easily distracted, daydreams) is consistent with reports of Johnny's difficulties dating back to the first grade.

Somatization

$T = 67$ (parent)
$T = 63$ (teacher)

Mild to moderate concerns in the area of Somatization appear to be related to Johnny's history of diabetes and its attendant difficulties.

Learning Problems

$T = 72$ (teacher)

Johnny's teacher reported significant learning problems for Johnny, indicating concerns in all academic areas. It was recommended that he be further evaluated through a Response to Intervention (RTI) procedure at his school, and services available to him should be planned accordingly.

Adaptability

$T = 39$ (parent)
$T = 32$ (teacher)

Functional Communication

$T = 30$ (parent)
$T = 32$ (teacher)

Johnny's mother and teacher reported concerns with his ability to adjust to changes in plans and to adequately communicate his need for help. These difficulties may interfere with his academic performance.

Johnny's results were strikingly similar for both teachers and parents. The diagnosis of ADHD, Predominantly Inattentive Type was made based on Johnny's history as reported by his mother and teacher during interviews and based on these scores on the PRS and TRS.

child's behavior across time and situations. Furthermore, parent involvement is required in the evaluation of most children, and parents' views of their children's behavior are often crucial in determining what will be done about the behavior. Parents' reports should therefore be obtained in the assessment of children's competencies and problems whenever possible (p. 3).

Content

The CBCL includes 100 items for the preschool version and 113 items for the school age version. Responses are made on a three-point scale (i.e., Not True; Sometimes/Somewhat True; Very True/Often True).

The CBCL syndrome scales are primarily empirically derived, with substantial use of factor-analytic methods. The CBCL scales were also derived separately by gender and age group (see Achenbach & Rescorla, 2000, 2001). Throughout the test development process, the CBCL develop-

TABLE 7.5 Sample Content of CBCL (6–18-Year-Old Version) Syndrome Scales

Anxious/Depressed: Cries a lot, is fearful, feels too guilty, talks of suicide

Withdrawn/Depressed: Would rather be alone, shy/timid, sad

Somatic Complaints: Feels dizzy, constipated, has headaches, nausea

Social Problems: Dependent, lonely, gets teased, prefers to be with younger kids

Thought Problems: Cannot get mind off of certain thoughts, sees things, stores things, strange behavior

Attention Problems: Cannot concentrate, daydreams, impulsive, cannott sit still

Rule-breaking Behavior: Drinks alcohol, lacks guilt, breaks rules, sets fires, prefers to be with older kids

Aggressive Behavior: Argues a lot, destroys others' things, gets in fights, mood changes, attacks people

From Achenbach & Rescorla (2001).

ers also emphasized the derivation of scales that were common across raters (e.g., parents and teachers). The CBCL parent and teacher scales have closely matched items and scales that make it easier for clinicians to make cross-informant comparisons. Sample item content from the CBCL scales is shown in Table 7.5.

The item content for the preschool (ages 1½–5) version of the CBCL is notably different from the version for 6–18-year olds, with somewhat different syndrome scales. The syndrome scales for the 1½–5-year-old version are Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior (Achenbach & Rescorla, 2000).

On both versions, there is a Total Problems score (the most global score available on the CBCL) as well as composites for Internalizing Problems and Externalizing Problems (see Achenbach & Rescorla, 2000, 2001). The CBCL also includes competence scales (except for the preschool version) that are designed to discriminate significantly between children referred for mental health services and non-referred children (Achenbach & Rescorla, 2001).

DSM-Oriented scales were formed based on experts' ratings (see Achenbach, Dumenci, & Rescorla, 2001) of how well the items fit DSM criteria for relevant disorders or groups of disorders (e.g., Major Depression and Dysthymia for the Affective Problems scale). For the school-age version of the CBCL, the DSM-Oriented scales are Affective Problems, Anxiety Problems, Somatic Problems, Attention/Hyperactivity Problems, Oppositional Defiant Problems, Conduct Problems. The five DSM-oriented scales on the preschool version of the CBCL are Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention Deficit/Hyperactivity Problems, and Oppositional Defiant Problems. The DSM-Oriented scales are a new feature to the Achenbach system and

were designed to more closely align scores that were available from these instruments to current diagnostic nomenclature.

Administration and Scoring

The CBCL is easily administered in 15–20 min. The CBCL is somewhat unique in that adaptive behavior is assessed with a combined fill-in-the-blank and Likert scale response format. In addition, some of the problem behavior items require the parent to elaborate on or describe the problem endorsed. This format is advantageous in that it allows the parent to respond in an open-ended format. Clinicians can gain access to qualitative information of value using this format. Open-endedness, however, also has a disadvantage: It may extend administration time and requires more decision making on the part of the parent.

Hand scoring and computer scoring are available for the CBCL. The CBCL offers normalized T-scores as the featured interpretive scores. Percentile ranks are also provided. T-scores are available for all scales and three composites: Externalizing, Internalizing, and Total. T-scores are now also offered for the Competence scales.

The advantages and disadvantages of using normalized versus linear T's are debatable (see Kline, 1995). On the one hand, the advantage of comparable percentile ranks across scales was recognized by the MMPI-A author team (see Chap. 6). Normalized scores, however, clearly change the shape of the many skewed raw score distributions forcing the T-score distribution to take a shape that it does not actually take in the general population (see Chap. 2). In addition, the reporting of T-scores on the CBCL is truncated for the Syndrome and DSM-Oriented scales such that low scores are reported simply as $T \leq 50$. For the Competence scales, the distribution is truncated above a T-score of 65 and below a T-score of 35.

Norming

The norming of the school age CBCL is based on a national sample of 1,753 children aged 6 through 18 years. This sample was collected in 40 states and the District of Columbia (Achenbach & Rescorla, 2001). Relevant stratification variables such as age, gender, ethnicity, region, and SES were recorded in an attempt to closely match US Census statistics on these variables. The respondents were mothers in 72% of the cases and fathers in 23% of the cases (5% of the cases used "others"). Sixty percent of the respondents were classified as White, with Hispanics appearing to be somewhat underrepresented (9%). Fifty-one percent of cases were from a middle SES background, 33% were from an upper SES background, and 16% were from a lower SES background. Forty percent of respondents were from the southern part of the USA (see Achenbach & Rescorla, 2001). From this sample, separate norms were developed for ages 6–11 and 12–18, with each of these groups further delineated by gender.

The norming sample for the preschool CBCL version for ages 1½–5 was also recruited in an attempt to match US Census statistics on the same variables. This sample consisted of 700 respondents (76% mothers, 22% fathers, 2% "others"). Fifty-six percent of respondents were White, 21% African American, 13% Latino, and 10% Mixed or Others. In the preschool norming sample, 33% of respondents were from an upper SES background, 49% from middle SES, and 17% from a lower SES background. Again, 40% of these respondents were from the southern USA (Achenbach & Rescorla, 2000).

Children were excluded from the sample if they had "received mental health or special education classes during the previous 12 months" (Achenbach & Rescorla, 2001, p. 76). Separate clinical norms are not offered for the CBCL.

Reliability

The CBCL has good evidence of reliability with internal consistency coefficients ranging from 0.78 to 0.97 on the Syndrome scales, 0.72 to 0.91 on the DSM-oriented scales, and somewhat lower internal consistency on the Competence scales (i.e., 0.63 to 0.79; Achenbach & Rescorla, 2001). On the preschool version of the CBCL, the internal consistency coefficients for the Syndrome scales and composites ranged from 0.66 to 0.95. For the DSM-Oriented scales, internal consistencies ranged from 0.63 to 0.86 (Achenbach & Rescorla, 2000). The data from a test-retest study for a sample of 73 (*mean* interval = 8 days) children yielded coefficients ranging from 0.80 for the Anxiety DSM-Oriented scale to 0.93 for the DSM-Oriented Conduct Problems scale on the 6–18-year-old version (Achenbach & Rescorla, 2001). For the preschool CBCL (*n* = 68), 8-day test-retest reliabilities were good, ranging from 0.74 for the Attention-Deficit/Hyperactivity Problems DSM-Oriented scale to 0.92 for the Sleep Problems scale (Achenbach & Rescorla, 2000).

A two-year stability study of 67 children yielded coefficients ranging from 0.45 for Somatic Problems to 0.81 for Aggressive Behavior for the school age CBCL. Twelve-month test-retest coefficients for the preschool version (*n* = 80) ranged from 0.52 for two of the DSM-Oriented scales to 0.62 for the Anxious/Depressed Syndrome Scale. These coefficients are indicative of strong reliability in light of the lengthy interval and the expected natural instability in some of these areas over time. Lastly, mother–father interrater agreement on the CBCL was generally good on the school age version with all coefficients except the Activities scale being 0.63 or higher. The interrater agreement on the preschool version was lower, with coefficients ranging from 0.48 to 0.67 (see Achenbach & Rescorla, 2000).

Validity

Much of the validity evidence reported by the authors of the CBCL focuses on the ability of the scale to differentiate clinical from nonclincal samples. Results of these analyses indicated good differential validity across scales for both boys and girls. As noted in Chap. 2, however, evidence of differential validity must now also show differentiation between clinical samples. To date, such evidence is lacking for the latest versions of the CBCL.

The factor structure of the CBCL continues to raise some conceptual issues regarding the content validity of the scales. For example, it is unusual for depression and anxiety items to be included on the same scale. In addition, the Attention Problems scale includes items that appear more indicative of hyperactivity and impulsivity than inattention (see Table 7.5). High scores on these scales still require a great deal of clinical judgment as to what characteristics led to the high scores and should be the focus of further attention.

Validity studies as well as basic and applied research investigations using the previous versions of the CBCL are legion. Although the research base of the current CBCL is not as well-established, some evidence on its validity is promising. For example, the preschool version of the CBCL has been found to be useful in screening for Autism Spectrum Disorders based on the Withdrawal and Pervasive Developmental Problems scales (Sikora et al., 2008), and the CBCL has been touted for its ability to screen for a variety of problem areas and its strong convergent and divergent validity (Scholte, Van Berckelaer-Onnes, & Van der Ploeg, 2008). However, the correspondence of the DSM-Oriented Anxiety scale on the CBCL to *DSM* criteria for anxiety disorders has been called into question (Ferdinand, 2008). Clearly, more research is needed on the latest rating scales in the Achenbach system, but just as clearly, the CBCL has

enjoyed and continues to enjoy a great deal of empirical support.

Interpretation

Interpretation of the CBCL is bolstered by many articles by Achenbach and colleagues devoted to its clinical use dating to McConaughy and Achenbach's (1989) informative work on this subject. The CBCL user is fortunate to have many interpretive resources available.

More specifically, McConaughy and Achenbach (1989) provide an assessment methodology for the identification of severe emotional disturbance in the schools. Their multi-axial empirically based assessment model proposes five axes for such assessment situations: (1) parent reports (Achenbach, CBCL), (2) teacher reports (Achenbach Teacher Report Form), (3) cognitive assessment, (4) physical assessment, and (5) direct assessment of the child (i.e., Achenbach Direct Observation Form and Youth Self-Report). McConaughy and Achenbach assist the psychologist working in schools further by linking each CBCL scale to the accepted criteria for severe emotional disturbance. High scores on the Anxious/Depressed scale may, for example, indicate the presence of a general pervasive mood of unhappiness, which, in turn, may qualify a child as severely emotionally disturbed and document eligibility for special education and related services. Of course, these examples fit best with previous versions of the CBCL and are only as useful as their degree of correspondence with eligibility categories used by school systems.

For interpreting the CBCL specifically, because there are no established validity scales and because there are some scales that include heterogeneous content, we recommend more attention to item-level interpretation than we have for other measures. That is, the clinician should pay close attention to scale elevations and draw most conclusions at the scale level; however, it would behoove practitioners to determine any concerning aspects of the parent's item response style that would render the protocol invalid. Additionally, interpretation should not stop at the scale level. Rather, one should inspect the items that led to the scale elevations to determine the best way to describe the child's difficulties. Fortunately, the scoring methods available for the CBCL make linking item responses to scale elevations a straightforward process.

Strengths and Weaknesses

The CBCL has many strengths that continue to make it a popular choice for clinicians. Noteworthy strengths include:

1. An ever-growing research base on the current CBCL, as well as a wealth of validity research on its predecessors.

2. Its popularity among professionals which facilitates communication about its results

3. Several writings that provide interpretive guidance above and beyond that provided by the manual

4. Improved approach to assessing competence and available of new DSM-Oriented scales that are aligned to *DSM* criteria.

5. Some response flexibility in that parents are asked to elaborate on their answers to some items.

Weaknesses of the CBCL include:

1. Lack of validity scales which are now common among behavioral rating scales.

2. Lack of close correspondence between the empirically derived scales and common diagnostic criteria (e.g., *DSM*; Hart & Lahey, 1998)

3. Heterogeneous content within some scales.

The CBCL continues to be a preferred choice of many child clinicians because of its history of successful use and popularity with researchers. The continuing development of the CBCL database bodes well for its future. The most recent versions of the CBCL would benefit from research aimed at assessing the construct validity of its scales, particularly the DSM-Oriented scales which were not part of the previous CBCL. Such research efforts are necessary to define further the degree of confidence that a clinician can place on specific scales for making differential diagnostic decisions.

Teacher Report: Teacher Report Form (TRF)

The Achenbach Teacher Report Form (TRF; Achenbach & Rescorla, 2001) is designed to be completed by teachers of children between the ages of 6 and 18 for the school-age version and between 1½ and 5 for the preschool version which is labeled the "Caregiver-Teacher Report Form" (Achenbach & Rescorla, 2000). The item content of the TRF is very closely matches the CBCL item content.

Content

The school age version of the TRF includes several background questions (e.g., "How long have you known this pupil?" "How well do you know him/her?"), a teacher's rating of a child's academic performance, and a four-item screening of a child's adaptive behavior with scoring on a 1–7 scale (e.g., "How hard is he/she working?" "How appropriately is he/she behaving?" "How much is he/she learning?" "How happy is he/she?"). The preschool TRF includes the same background questions but does not include the other items. These background questions have associated norms and fall under the "Adaptive Functioning" domain of the TRF. The major portion of the TRF consists of 100 items for the preschool version and 113 items for the school age version. These items describe problematic behaviors and emotions that the teacher rates as being Not True, Somewhat True/Sometimes True, or Very True/Often True of the child. The problem behavior items cover a broad array of both internalizing (e.g., anxiety, depression, somatic complaints) and externalizing (antisocial behavior, aggression, oppositionality) behaviors. As with the CBCL, the TRF now includes DSM-Oriented scales (6 for the school age version; 5 for the preschool version). The only scale difference is that the TRF does not include a Sleep Problems scale.

Administration and Scoring

The TRF takes approximately 15–20 min to complete. The instructions to the teacher are printed on the front of the answer sheet. Scoring of the TRS can be done by hand using the TRF Profile Sheets with separate profile sheets available for boys and girls. However, a computer-scoring system is available that greatly facilitates scoring by automatically calculating raw scale scores and converting them to norm-referenced scores appropriate for the child's age and gender.

Both the Profile Sheets and the computer-scoring program provide raw scores and norm-referenced scores for several scales. As with the CBCL, a Total Problem score, which is an overall indicator of a child's classroom adjustment, and two broadband scores consisting of Internalizing and Externalizing behaviors are included. These broad dimensions are further divided into the eight Syndrome scales.

The TRF allows for raw scores on all scales to be converted to T-scores and percentile ranks based on the standardization sample. The T-scores are normalized standard scores. That is, the raw score distributions are transformed to a normalized distribution. This procedure allows T-scores on all scales to have similar dis-

tributions and corresponding percentiles based on the assumptions of a normal distribution. However, as noted previously for the CBCL, this transformation assumes that the dimensions assessed by the scale *should* be normally distributed in the general population, an assumption that is questionable because most children tend to cluster in the normal end of the distribution. Norm-referenced scores are based on gender and age-specific norms. In addition, as with the CBCL, T-scores on the TRF are truncated, such that the lowest score provided is ≤ 50.

Norming

The norming sample for the school-age version of the TRF consisted of 2,319 children, 72% of whom were White. Fourteen percent were identified as African American, and 7% were identified as Latino; thus, both ethnic minority groups appear to be underrepresented in this sample. The TRF normative sample appears to be geographically representative. Thirty-eight percent of children in this sample were from an upper SES background, 46% from a middle SES background, and 16% from a low SES background.

For the preschool version, the norming sample consisted of 1,192 children. This sample was geographically diverse, but only 10% of the sample came from a lower SES background. The sample represented Whites and African Americans well (i.e., 48% and 36%, respectively), with only 8% of the sample identifying as Latino.

For both versions of the TRF, the normative sample excluded children who had received mental health or special education services within the preceding 12 months. Therefore, as with the other rating scales in the Achenbach system, the sample should be considered a normal comparison group, rather than one that is normative and representative of the general population.

Reliability

Achenbach and Rescorla (2000, 2001) provide three types of reliability information on the TRF. Internal consistency estimates were provided for the Syndrome and DSM-Oriented scales. Coefficients indicated good internal consistency, ranging from 0.72 to 0.95. For the preschool version, coefficients were quite variable, ranging from 0.52 for Somatic Complaints to 0.96 for the Aggressive Behaviors scale. Test-retest reliability over an average of a 16-day interval is presented on a sample of 44 children in the age range of 6–18. The test-retest coefficients were generally high (i.e., 0.80s and higher), with the exception of the Withdrawn/Depressed scale ($r = 0.60$) and the Affective Problems scale ($r = 0.62$). Four-month test-retest reliability was variable, ranging from 0.31 (Affective Problems) to 0.72 (Hyperactivity-Impulsivity). The 8-day test-retest reliability for the preschool version of the TRF was somewhat variable, ranging from a coefficient of 0.57 for Anxiety Problems to 0.91 for Somatic Complaints (the scale with the lowest internal consistency). Three-month test-retest reliability for the preschool TRF was variable with coefficients ranging from 0.22 for Somatic Complaints to 0.71 for Emotionally Reactive.

Correlations between ratings from two different teachers are provided for 88 children (Achenbach & Rescorla, 2001). The correlations were modest across scales, with the mean coefficient for the full sample being 0.49 for the Competence scales, 0.60 for the Syndrome scales, and .58 for the DSM-Oriented scales. Similar analyses for 102 preschoolers revealed an overall mean coefficient of 0.62, with a range of 0.21 (Somatic Complaints) to 0.78 (Aggressive Behavior).

Validity

There is relatively limited validity information available for the current versions of the TRF. The validity data reported in the manuals (Achenbach & Rescorla, 2000,

2001) mainly focus on the ability of the scales to differentiate non-referred from clinical samples within gender. The TRF scales generally show such differential validity. However, the Somatic Complaints scale of both versions does not appear to consistently differentiate the two groups (Achenbach & Rescorla, 2000, 2001). As noted above, validity studies for the BASC-2-TRS demonstrated good correspondence with the TRF, especially for externalizing problems. An exception was the lack of correlation between scales on each form assessing somatic complaints. Research has supported the predictive validity of the TRF at age three in predicting problems, especially externalizing problems, at age five (Kerr et al., 2007). The TRF, like the CBCL, was based heavily on its well-researched predecessor. There is a wealth of research on the previous versions of the TRF which supports its validity in (1) differentiating clinic-referred children from non-referred children, (2) correlating with classroom observations of children's behavior, and (3) correlating with independent clinical diagnoses (see Achenbach, 1991; Casat, Norton, & Boyle-Whitesel, 1999; Piacentini, 1993).

Interpretation

Information on the reliability and validity of the adaptive functioning component of the TRF is lacking; therefore, interpretations of these scales should be done very cautiously, if at all. Subsequently, although it may be useful to next consider the TRF Total Problems score and composites, more specific information can be gleaned from interpretations of the eight syndrome scales and the six DSM-Oriented scales. The reliability of these scales (with the exception of the Somatic Complaints scale) is good. However, because the initial development of the TRS item pool was done in an attempt to be atheoretical, the item content of the TRS scales tends to be more heterogeneous than other rating scales that used a more explicit guiding theory for scale development. For example, the Attention Problems scale consists of items traditionally associated with inattention (e.g., difficulty concentrating) but also includes items associated with immaturity, overactivity, poor school achievement, and clumsiness. Therefore, it is imperative that the items that led to a clinical scale elevation be reviewed to understand the meaning of the elevation. For example, a child may show an elevation on the Attention Problems scale because of problems with immaturity, clumsiness, or academic problems, or a child may have an elevation due to problems of inattention and/or hyperactivity. However, because of the unreliability of individual items, this item-level analysis should be conducted only when there is an elevation on the clinical scale. The Attention-Deficit/Hyperactivity Problems DSM-Oriented scale can be quite helpful in this type of scenario as it is more closely aligned with characteristics indicative of ADHD.

Interpretation of the Thought Problems scale on both the TRF and CBCL deserves several cautionary notes. This scale has an especially heterogeneous content, consisting of items describing obsessions (e.g., "Cannot get mind off certain thoughts"), compulsions (e.g., "Repeats acts over and over"), fears (e.g., "Fears certain animals, situations, or places"), and psychotic behaviors (e.g., "Hears sounds or voices that are not there"), many of which are fairly ambiguous (e.g., strange behavior, strange ideas). For this scale, it should be apparent that item level interpretation and integration with other information collected during the assessment are crucial before drawing conclusions.

One final note is in order for interpreting TRS scales. Because the norm-referenced scores of the TRS are based on a normal sample and not a normative sample, it is recommended that a more conservative cut-off score be used than would be

Box 7.2

Sample Case Using the CBCL and TRF

Doug is an 11-year-old, fifth-grade boy who was referred for an assessment because of parental and teacher concerns about his school performance. He is suspected of having significant attention problems. Doug also has significant trouble in peer and other relationships. He often fights and argues with peers, resulting in his often playing by himself.

Doug has a history of significant medical difficulties. He is reportedly the product of an at-risk pregnancy. Although he reportedly reached most developmental milestones within normal age limits, he has a history of motor delays. In second grade, he was diagnosed with muscular dystrophy. He does not tolerate many foods well, and consequently, his appetite is poor.

In the first grade, Doug was diagnosed with a learning disability in reading. He is in a resource special education program. He is described by his teacher as having significant social difficulties. He is reportedly often disrespectful toward teachers and peers. His grades deteriorated significantly toward the end of the last academic year. His teachers consider him to be a capable underachiever with behavior problems such as inattention, excessive talking, fighting, arguing, and poor work completion.

Doug's performance on intelligence and achievement testing indicate that his cognitive functioning is consistent with what would be expected for his age, whereas his writing and reading skills were slightly below what would be expected for his age.

Doug's ratings from the CBCL and TRF completed by his mother and teacher reflect the multitude of his difficulties as follows:

Comparisons across these two raters, behavioral observations, and background information indicate that Doug is experiencing difficulties in a number of behavioral, emotional, and social domains. In particular, reports by Doug's mother and teacher of his tendency to daydream, his difficulty sustaining attention, his impulsive behavior, and his restlessness, as well as the fact that such behaviors were reported for Doug at an early age, suggest that he meets criteria for a diagnosis of Attention-Deficit

Scale	Mother/Teacher (T)
Withdrawn/Depressed	70/61
Somatic Complaints	91/64
Anxious/Depressed	79/55
Social Problems	80/75
Thought Problems	79/67
Attention Problems	78/69
Rule-Breaking Behav.	73/76
Aggressive Behav.	85/65

Hyperactivity Disorder (ADHD)- Combined Type. Ratings on the CBCL and TRF Attention-Deficit/Hyperactivity Problems scale also indicated significant problems with behaviors related to ADHD. In addition, Doug's reported argumentativeness, disrespect toward others, and tendency to break rules at home and at school indicate that he meets criteria for Oppositional Defiant Disorder as well. Some concerns regarding depression were also reported, particularly by Doug's mother. However, these issues do not warrant another diagnosis, but they do warrant continued monitoring and some interventions.

Indications of thought problems were not corroborated by other findings. The problems reported by Doug's mother on this scale (e.g., trouble sleeping, cannot't get his mind off certain things) seemed particularly tied to his reported problems with depression.

CBCL and TRF Social Problems scores were corroborated by reports of his difficulty getting along with others. Doug's social interaction skills are in need of intervention.

Although the CBCL and TRF clearly indicate some areas of concern and in need of intervention, this case also highlights the need, more pressing in a case like this, to complement the CBCL and TRF with other assessment strategies. In this case, background information was particularly important for corroborating rating scale information and indentifying treatment objectives.

the case for other rating scales. Any elevations, regardless of the degree of elevation, should still be considered in conjunction with other assessment results (e.g., parent report, self-report, history, observations, etc.). The sample case that follows gives a brief example of an interpretive approach to the CBCL and TRF (Box 7.2).

Strengths and Weaknesses

The TRF remains one of the most widely used of the teacher-completed behavior rating scales. In addition to its popularity and familiarity with a large number of professionals, the strengths of the TRF include:

1. The large research literature on the TRF and its predecessors which demonstrates good correspondence between the TRF and other indicators of child functioning, particularly on externalizing behaviors.
2. The inclusion of DSM-Oriented scales aids the clinician in interpreting teacher reports in terms of diagnostic categories.
3. A larger normative sample than was available for the previous versions of the TRF.

Some weaknesses of the TRF include:

1. An underrepresentation of Hispanics in the normative sample.
2. The exclusion of children with mental health or special education services in the normative sample, indicating that such children are not represented.
3. The questionable reliability and validity of the Somatic Complaints scale.
4. A relatively limited assessment of adaptive functioning.

Child Symptom Inventory-4 (CSI-4)

Parent and Teacher Report Checklists

The Child Symptom Inventory-4 (CSI-4; Gadow & Sprafkin, 1998) is a standardized rating scale designed to assess the symptoms of over a dozen childhood disorders. This content is unique from other rating scales in that it is the only omnibus rating scale whose entire content is explicitly tied to the diagnostic criteria specified in the *DSM-IV* (American Psychiatric Association, 1994). Therefore, its content reflects the research that went into developing these diagnostic criteria, which is excellent for some disorders but more suspect for others especially for children (Widiger et al., 1998).

The CSI-4 has both parent and teacher report versions that contain analogous scale content, which enhances its usefulness for comparing and combining ratings across informants. The CSI-4 was designed for use with children of ages 5–12, but there is an analogous Adolescent Symptom Inventory-4 for youth ages 12–18 (ASI-4; Gadow & Sprafkin, 1998) that has both parent and teacher versions and an adolescent self-report checklist, the Youth Symptom Inventory-4 (YSI-4). As part of the same system, the Early Childhood Inventory-4 (ECI-4; Gadow & Sprafkin, 1997) assesses *DSM-IV* symptoms in preschool children of ages 3–5.

The content of these forms is mostly identical; however, they also each include some domains that may be particularly developmentally relevant. For example, the ASI-4 includes assessments of Antisocial Personality Disorder, Anorexia, and Bulimia. The ESI-4 omits items screening for psychosis but includes items for Selective Mutism, Reactive Attachment Disorder, sleep problems, and elimination problems.

A fairly unique aspect of this system is the inclusion of a symptom checklist specifically for ADHD (ADHD-SC4). This inventory includes 50 items that assess the core symptoms of inattention and hyperactivity as well as other areas of interest related to ADHD. More specifically, the ADHD-SC4 includes a Peer Conflict scale to assess the social difficulties that often accompany ADHD and a Stimulant Side Effects Checklist as a means to monitor side effects of medication a child may be taking for the management of his/her ADHD symptoms.

Content

The CSI-4, because of its explicit link to the *DSMIV* system for classifying mental disorders (Table 7.7), covers many symptom domains that are not assessed by other rating scales (e.g., tic disorders), especially symptoms of more severe types of childhood psychopathology (e.g., Obsessive-compulsive Disorder, PosttraumaticStress Disorder, schizophrenia, autism, Asperger's Disorder). As a result, the CSI-4 may be especially useful in the assessment of more severely disturbed children. The items on the CSI-4 were designed to approximate symptoms from the *DSM-IV* with rephrasing done to eliminate jargon, to emphasize observable behavior, rather than making inferences about internal processes, and to eliminate descriptions of frequency (e.g., "often" acts without thinking). The CSI-4 is fairly long (i.e., 97 items for the parent form, 77 items for the teacher form), but the scales are grouped according to each individual diagnosis and, as a result, the whole scale need not be given. Instead, symptoms of certain disorders can be selected based on the specific purpose of the evaluation (e.g., Frick, Bodin, & Barry, 2000).

Administration and Scoring

The 97 items on the CSI-4 are rated on a 0 ("Never") to 3 ("Very Often") scale.

TABLE 7.6   Prevalence of *DSM-IV* Disorders in a Normal Sample using the CSI-4 Screening Criteria

	Parent Checklist		Teacher Checklist	
DSM-IV Category	**Boys**	**Girls**	**Boys**	**Girls**
Attention-Deficit Hyper.	(n = 134)	(n = 129)	(n = 662)	(n = 661)
Inattentive	6.4	2.4	11.2	4.2
Hyperactive-Impulsive	4.1	3.2	3.5	0.5
Combined	4.1	0.8	4.7	1.2
Oppositional Defiant	9.2	7.0	6.3	1.8
Conduct	6.8	2.3	3.5	1.1
Generalized Anxiety	3.7	2.3	0.8	0.8
Separation Anxiety	3.0	3.1		
Schizophrenia	0.0	0.0	0.0	0.0
Major Depression	0.0	0.0	0.0	0.0
Dysthymia	2.2	0.0	0.8	0.6
Autism	0.7	0.8	0.4	0.3
Asperger's	0.0	0.0	0.0	0.0

Like other rating scales, quantitative scale scores corresponding to each diagnostic category (e.g., Conduct Disorder) can be determined by simply summing the ratings across items, and this score is called the "symptom severity" index. However, a "symptom count" score can be used to more closely approximate the *DSM-IV* method of considering symptoms as either present or absent. Using this method, any item rated as being present "Often" or "Very Often" is considered to indicate the presence of the symptom, and any item rated as "Never" or "Sometimes" is considered to indicate the absence of the symptom.

Norming

The normative sample of the CSI-4 included 552 parent ratings (272 boys, 280 girls) and 1,323 teacher ratings (662 boys and 661 girls) in three states (Gadow & Sprafkin, 2002). The children were elementary school-age. Children receiving special education services were not included, making this sample a normal rather than normative sample.

In addition to being somewhat geographically limited, there was great overrepresentation of Caucasian children, particularly for the teacher rating sample, with that sample being 95% Caucasian, 2.8% African American, and 0.7% Hispanic. Because of these limitations in the CSI-4 normative samples, norm-referenced interpretations should only be made very cautiously. However, because the CSI-4 was not designed primarily to be used as a norm-referenced instrument but instead was designed as a screener for *DSM* diagnoses, the more critical psychometric consideration is its reliability and validity for this purpose.

Reliability

One study reporting on the reliability of the parent CSI-4 found moderate to good internal consistency for both symptom-severity scores and symptom-count scores

(Sprafkin et al., 2002). More specifically, internal consistency coefficients ranged from a low of 0.45 for the symptom-severity index for schizophrenia to 0.92 for symptom severity of ADHD-Predominantly Inattentive Type. Four-month test-retest reliability coefficients ranged from 0.35 for Major Depression to 0.88 for ADHD Predominantly Hyperactive-Impulsive Type, with all but two coefficients being 0.65 or higher (Sprafkin et al.).

Relatively limited information on the reliability of the teacher version of the CSI-4 is available. For example, the CSI-4 manual describes test-retest reliability for the ADHD and ODD categories during a medication trial for children with behavioral problems. One-week test-retest coefficients for these two diagnoses averaged 0.62 for ADHD and 0.90 for ODD (Gadow & Sprafkin, 1998).

Validity

There are several pieces of evidence for the validity of the CSI-4 as a screener for *DSM-IV* diagnoses in school-aged children. First, the prevalence of the diagnoses, based on the symptom-count scoring method of the CSI-4 in the norm sample, seemed to approximate those found in community samples of children using structured diagnostic interviews (Frick & Silverthorn, 2001). These estimates, computed separately for boys and girls, are provided in Table 8.4. Second, when these prevalence estimates were compared to a clinic-referred sample of school-aged children, the prevalence of *DSM* diagnoses was higher in the clinic-referred sample for almost all diagnoses. The exceptions were ADHD Hyperactive/Impulsive Type for both boys (7.5% clinic vs. 4.1% norms) and girls (4.7 vs. 3.2%), Asperger's Disorder for both boys (2.7 vs. 0%) and girls (1.3 vs. 0%), and Schizophrenia for boys (1.1 vs. 0%). The primary concern is the finding for the one ADHD subtype, because the failure to find signifi-

cant differences for the latter two disorders seems largely due to the very low base rate of these disorders in both samples.

Third, and probably most importantly, Gadow and Sprafkin (1998) reported on a clinic-referred sample of 101 referrals (between the ages of 6 and 12 years) to an outpatient child psychiatry service, in which they tested the correspondence between CSI-4 diagnostic cut-offs and clinical diagnoses made by mental health professionals. In general, the sensitivity and specificity rates for the disorders assessed by the CSI-4 generally indicated quite good correspondence with clinical diagnoses. This correspondence was especially good when parent and teacher ratings were combined, such that a disorder was considered present if either the parent or teacher ratings led to a CSI-4 screening diagnosis. For this multi-informant composite, the Sensitivity rates ranged from 0.87 to 1.00, and the specificity rates range from 0.40 to 0.92. For example, a diagnosis of Generalized Anxiety Disorder (GAD) showed a sensitivity rate of 0.93 indicating that, of those in the sample who had a clinical diagnosis of GAD, 93% crossed the screening cut-off for a diagnosis on the CSI-4. The specificity rate of 0.71 indicates that, of those without the diagnosis of GAD in the sample, 71% did not cross the screening cut-off on the CSI-4.

Sprafkin and colleagues (2002) found good convergent validity for the CSI domains (parent form) based on their relations with the CBCL Syndrome scales. Of note, virtually all CSI domains were moderately correlated with the Anxious/Depressed scale of the CBCL, which may speak more to the general distress nature of that CBCL scale than the lack of discriminative validity of the CSI domains. They also concluded that the CSI-4 is a good screener of a variety of child disorders based on the high correct classification rates found in their sample.

In a separate study, the teacher form of the CSI-4 showed similarly good diagnostic accuracy with diagnoses made from structured interviews and moderate relations with parent ratings (Gadow et al., 2004).

It should be noted that the research on the parent and teacher forms of the CSI-4 far outpaces the research available on their companion measures, the ESI-4 and ASI-4. However, this issue is of less concern given the highly similar framework under which these measures were developed and the true intent of these measures (i.e., to screen for symptoms included in a widely used diagnostic nosology).

Interpretation

Although the CSI-4 content is designed to correspond to the symptoms of *DSM-IV* disorders, the authors of the scale are very clear in stating that the scale should never be used in isolation to make diagnoses (Gadow & Sprafkin, 1998). Instead, the CSI-4 is a screener that could indicate the need for a more complete diagnostic assessment. Rather than being a significant limitation, it highlights some very important uses of the CSI-4. As mentioned in Chapter 3, there is a great deal of overlap and co-occurrence among the various forms of childhood disorders. The CSI-4 provides an efficient way of screening for a large number of potential comorbidities that can allow for a more focused and intensive assessment in the specific areas of concern indicated by this screening. Also, such a screening, because it is time- and cost-efficient, may be quite beneficial for defining smaller samples at high risk for psychopathology from larger non-referred samples (see Frick et al., 2000).

Given the fairly substantial limitations in the normative samples for the CSI-4 and its companion measures, norm-referenced interpretations are not recommended. Instead, the symptom-count method of

scoring is recommended to provide the best approximation of *DSM-IV* disorders. Although the normative data suggest that the symptom-severity method of scoring is somewhat more reliable, it is not as consistent with the structure of the *DSM* criteria that relies on the presence or absence of symptoms to make diagnoses. Also, without good normative data, it is difficult to judge when symptom severity should be considered "significant," unless one is simply trying to make relative comparisons between groups of children. In addition, the symptom-count method provides a very easy method for combining information from multiple informants, which as the available data clearly suggest also provides the best correspondence to clinical diagnoses. Specifically, a symptom can be considered present if endorsed by any informant (e.g., either teacher or parent), and the rate of symptomatology based on this multi-informant procedure can be compared to *DSM-IV* criteria (see Piacentini, Cohen, & Cohen, 1992).

Strengths and Weaknesses

Strengths of the CSI-4 system include:

1. Its uniqueness in attempting to assess content that directly corresponds to *DSM-IV* classifications of childhood psychopathology.
2. Efficiency in gaining diagnosis-relevant information.
3. Good correspondence with clinical diagnoses, especially when using both parent and teacher informants.

Weaknesses of note include:

1. The lack of a large normative base; thus, norm-referenced interpretations should not be made from this rating scale system.

2. A relative lack of research, particularly on the ESI-4 and ASI-4, as well as the self-report component of this system.

The CSI-4 and its related measures offer a potentially useful component to child assessment, particularly when preliminary diagnoses are needed for reimbursement/insurance purposes. However, as the authors note, the CSI-4 (or any other assessment technique) should not be used as the sole criterion for making a clinical diagnosis. Instead, such decisions must be based on a combination of many sources of information.

Conners, 3rd Edition (Conners-3)

Parent Rating Scale

The Conners-3 (Conners, 2008a) Parent Rating Scale (Conners-3-P) is the most recent revision to a widely used behavior rating scale system. The Conners Parent Rating Scale is designed similarly to the BASC and Achenbach systems in that it includes a number of clinically relevant domains for which normative scores are derived. The parent rating scale is designed for ages 6 through 18. The Long Form contains 110 items and the Short Form contains 45 items. There is also a 10-item Global Index form. The Conners-3-P takes 10–20 min to complete, depending on which form is used. The following discussion will focus on the Long Form.

As noted in Table 7.1, we recommend the Conners Comprehensive Behavior Rating Scales (Conners, 2008b) for an assessment that covers externalizing, internalizing, and academic issues. However, as the information below indicates, the Conners-3 is unique in its detailed

evaluation of ADHD and other externalizing issues.

Scale Content

The Conners-3-P includes five empirically-derived scales: Hyperactivity/Impulsivity, Executive Functioning, Learning Problems, Aggression, and Peer Relations. An Inattention scale developed theoretically is also available, as are five *DSM-IV-TR* Symptom scales for each of the Disruptive Behavior Disorders (i.e., 3 ADHD subtypes, ODD, and CD). The Conners-3-P includes screening items for depression and anxiety, as well as impairment items for home, school, and social relationships. Like the BASC, the Conners-3 includes critical items that may signal the need for further follow-up. These critical items are particularly geared toward severe conduct problem behaviors (e.g., uses a weapon, is cruel to animals). Consistent with its predecessors, the Conners-3 includes a brief ADHD Index. This scale is based on items that best differentiate ADHD from non-clinical samples. As described in Chap. 6 for the Conners-3 SR, the Conners-3-P has three validity scales: Positive Impression (or "fake good"), Negative Impression ("fake bad"), and the Inconsistency Index. These scales are new to the Conners system. Two open-ended questions regarding other concerns and particular strengths/skills are also included. Detailed information on the generation and selection of items is provided in the Conners-3 manual (Conners, 2008a).

Administration and Scoring

The Conners-3-P uses a four-choice response format where 0 = not at all true (never, seldom), and 4 = very much true (very often, very frequently). A Spanish translation is available. Both hand scoring and computer scoring are available.

Raw scores are transformed to linear T-scores, meaning that each scale maintains its natural distribution in the conversion to norm-referenced scores. Separate norms are used for boys and girls, as is the case for the other versions of the Conners-3. Norms are also computed by age.

Norming

The normative sample of 1,200 cases was collected mainly in the USA, with a small number of cases coming from Canada. Recruitment of the normative sample was aimed at reflecting US Census data regarding ethnicity/race. Data reported by Conners (2008a) indicate that the normative sample closely reflects the Census statistics. This representativeness is a notable improvement over previous versions of the Conners rating scale, in that the previous samples were predominantly Caucasian. As noted for the Conners-3 SR, the Western USA appears to have been somewhat underrepresented. The majority of the parents (63.5%) in the normative sample had at least some post-secondary education. A clinical sample of 718 participants was also collected for validation purposes, with over 35% of that sample being diagnosed with ADHD or one of its subtypes.

Reliability

Internal consistency coefficients for the content and *DSM* scales of the Conners-3-P are all 0.80 and higher, and many are 0.90 and higher for the overall sample. The Peer Relations scale for girls had somewhat lower coefficients (i.e., 0.72 for 6–9-year olds; 0.78 for 10–13-year olds) Two to four week test-retest coefficients were good (i.e., all higher than .70). Interrater reliability for the parent form was also good, with adjusted *r*s all 0.74 and higher for the content and *DSM* scales (see Conners, 2008a).

Validity

Because the content scales were empirically-derived, it is not surprising that confirmatory factor analyses supported the five-factor model for those scales. All intercorrelations among content and *DSM* scales were moderate to high in magnitude (i.e., ranging from 0.36 to 0.98). Correlations with analogous scales from the teacher and self-report forms of the Conners-3 were all moderate (i.e., *r*s = 0.49 to 0.67). Criterion-related validity was demonstrated through moderate to high correlations between Conners-3-P scales and analogous scales from the BASC-2-PRS, CBCL, and BRIEF (see Conners, 2008a). The associations between the Conners-3 and CBCL were particularly high. Differential validity evidence also indicates that the Conners-3-P was successful in distinguishing both a general population sample and within clinical samples. That is, scores on scales such as those tied to ADHD tended to be elevated for clinical groups relative to non-clinical groups and higher for individuals with ADHD relative to others within a clinical population. The correct classification rate based on content and *DSM* scale elevations were also relatively high (i.e., 57 to 86%). More validity evidence for the Short Form and the Indexes are available in the manual (Conners, 2008a).

Interpretation

As discussed in Chap. 6, Conners (2008a) provides a clear step-by-step approach for interpreting ratings on the various forms of the Conners-3. This approach involves (a) examining the validity scales; (b) evaluating scale elevations; (c) examining the overall profile of scores (i.e., determining the constructs that seem to be represented across elevations); (d) item-level interpretation, including critical items and screener items;

and (e) integration with other assessment information.

Strengths and Weaknesses

Some of the strengths of the Conners-3-P are:

1. The improved representativeness of normative sample
2. Availability of complementary teacher and self-report forms that provide a comprehensive assessment of externalizing problems
3. Good initial reliability and validity evidence
4. Brevity of Short and Index Forms

Some characteristics that may be considered weaknesses are:

1. Limited assessment of internalizing problems and adaptive functioning (an issue that is addressed through use of the Conners CBRS)
2. Uniform negative wording of items, which may result in a negative response set
3. A lack of available validity research conducted by persons other than the developers

Teacher Rating Scale

The Teacher Rating Scale in the Conners-3 system is very similar to the parent rating scale. In fact, the teacher ratings scale and parent rating scale include the same scales. The Long Form of the teacher rating scale is slightly longer than that of the parent rating scale (i.e., 115 items), whereas the Short Form is slightly shorter (i.e., 41 items). Two 10-item Hyperactivity Index forms are also available. The following discussion will focus

primarily on the Long Form of the Con-ners-3-T. As with the self-report and par-ent forms of the Conners-3, we discuss the Conners-3-T because of its relatively unique focus on ADHD and behavioral problems. The companion teacher rating scale from the Conners CBRS (Conners, 2008b) provides a more extensive assess-ment of broader domains of functioning. Therefore, the selection of one set of rat-ing scales versus the other in the Conners family should be dictated by the purpose of the evaluation.

Content

The Conners-3-T has some item overlap with the Conners-3 parent rating scale, but there are also unique items in each form. The same four-point response scale used for the self-report and parent-report ver-sions of the Conners-3 is also used for the teacher-report scale. As noted above, the scales are the same as those for the par-ent rating scale, including validity scales, impairment items, and critical items.

Administration and Scoring

The Conners-3-T can be completed in 10–20 min, or less if the Short Form is used. The Conners-3-T has both hand-scoring and computer-scoring formats that allow for easy calculation of norm-referenced scores. As with the self-report and parent report forms, only sex-specific T-scores can be calculated. The scores are Linear T-scores and are based on each age group, which allows it to capture potential vari-ability in discrete developmental stages.

Norming

The norming process for the Conners-3-T was essentially the same as that used for the Conners-3-P and Conners-3 SR. Specifically, the norming sample for the Conners-3-T consists of 1,200 teach-ers from throughout the USA, with a few respondents from Canada. Recruitment was aimed at a sample that would reflect U.S. Census data on ethnicity/race. The students rated by teachers in the norming sample do appear to match the Census data on ethnicity/race (Conners, 2008a). However, the sample appears to be somewhat skewed toward middle to high SES – based on par-ent education – as 76.9% of the students rated by teachers in the norming sample had parents with at least some post-secondary education. Almost as many cases came from Canada as came from the western USA in the Conners-3-T norming sample.

Reliability

Internal consistency coefficients for the teacher report version Conners-3 were quite high. Specifically, the coefficients for each of the content and *DSM* scales were 0.90 or higher, with the exception of the Conduct Disorder scale (0.77). Two- to four-week test-retest reliability coefficients were also good, ranging from 0.72 to 0.83 (see Conners, 2008a). Lastly, and perhaps particularly importantly for teacher ratings, interrater reliability coef-ficients for pairs of teacher raters were moderate to high. The Peer Relations and Oppositional Defiant Disorder scales had the lowest adjusted coefficients (i.e., 0.52 and 0.55, respectively), whereas the Hyperactivity/Impulsivity and Conduct Disorder scales had the highest coeffi-cients (i.e., 0.77). It should be noted that the lower coefficients from these analyses may reflect less-than-ideal rater agree-ment, or they may reflect real differences in a child's behavior from one classroom context to another. Additional analy-ses, particularly in determining whether teacher agreement might change as a function of the child's age, are needed.

Validity

Factor analyses revealed a four-factor solution for the Conners-3-T: Hyperactivity/Impulsivity, Aggression, Peer Relations, and a combined Learning Problems/Executive Functioning scale. Conners (2008a) also found support for considering the Learning Problems/Executive Functioning as consisting of two subscales consisting of items intended to load on a Learning Problems and an Executive Functioning scale. As noted above, the Conners-3-T scales were moderately correlated with the same scales from the parent and self-report versions. The scales on the Conners-3-T were all moderately interrelated. Criterion-related validity for the Conners-3-T was supported through moderate to high correlations between Conners-3-T scales and analogous scales on the BASC-2-TRS, Achenbach TRF, and BRIEF Teacher Form. Similar to the parent version of the Conners-3, the teacher version demonstrated good differential validity in that scales were elevated for individuals from a clinical sample relative to a general sample, and scale scores tended to differ within the clinical sample in intuitive ways. For example, ADHD scale scores tended to be higher for youths diagnosed with ADHD than for youths with other difficulties who did not have ADHD diagnoses (see Conner, 2008a).

Interpretation

At the very least, the Conners-3-T appears to be useful as a screening for problems in classroom adjustment, particularly in terms of learning or externalizing problems, and as part of a comprehensive assessment battery. The recommended approach for interpreting the Conners-3-T mirrors that described for the Conners-3 SR and Conners-3-P.

Strengths and Weaknesses

The strengths of the Conners-3-T include:

1. Content that allows for an extensive assessment of ADHD symptoms and other behavioral problems.
2. Good correspondence across scales on the parent and teacher versions, which facilitates comparisons in a multi-informant assessment.
3. The presence of several short screening scales which may be more feasible for many teachers.

Apparent weaknesses of the Conners-3-T include:

1. Minimal assessment of depression and anxiety, as well as adaptive functioning. The Conners CBRS should be used if extensive assessment of these domains is desired.
2. Lack of research on reliability and validity conducted by persons other than the developer.
3. The normative sample is not quite as diverse as that for the parent and self-report forms of the Conners-3, yet it is still diverse in terms of race/ethnicity.

Personality Inventory for Children-2 (PIC-2); Student Behavior Survey (SBS)

Parent Report *PIC-2*

The Personality Inventory for Children-2 [(PIC-2); Lachar & Gruber, 2001] is based closely on its predecessor, the PIC-R (Wirt, Lachar, Klinedinst, & Seat, 1990). The original development of the PIC followed closely on the heels of the MMPI, with much of the early development work taking place in the 1950s. The PIC-2 is a 275-item rating scale designed for use with

parents of children between the ages of 5 and 19 years (Lachar & Gruber, 2001).

Scale Content

The PIC-2 scales, although revised, have a long clinical history. The PIC-2 includes scales that were developed via a mixture of empirical means with considerable use of external validation techniques and scales

TABLE 7.7 PIC-2 Clinical Scales, Subscales, and Internal Consistency Coefficients

Cognitive Impairment (r = 0.87)

Inadequate Abilities (r = 0.77)

Others think my child is talented

My child seems to understand everything that is said.

Poor Achievement (r = 0.77)

It is hard for my child to make good grades.

Reading has been a problem for my child.

Developmental Delay (r = 0.79)

At one time my child had speech difficulties.

My child could ride a tricycle by the age of 5.

Impulsivity and Distractibility (r = 0.92)

Disruptive Behavior (r = 0.91)

My child jumps from one activity to another.

My child cannot keep attention on anything.

Fearlessness (r = 0.69)

My child will do anything on a dare.

Nothing seems to scare my child.

Delinquency (r = 0.95)

Antisocial Behavior (r = 0.88)

My child has been in trouble with the police.

My child has run away from home.

Dyscontrol (r = 0.91)

When my child gets mad, watch out!

Many times my child has become violent.

Noncompliance (r = 0.92)

My child often breaks the rules.

My child tends to see how much he/she can get away with.

Family Dysfunction (r = 0.87)

Conflict among members (r = 0.83)

There is a lot of tension in our home.

Our family argues a lot at dinner time.

Parent Maladjustment (r = 0.77)

One of the child's parents drinks too much alcohol.

The child's parents are divorced or living apart.

Reality Distortion (r = 0.89)

Developmental Deviation (r = 0.84)

My child often gets confused.

My child needs protection from everyday dangers.

Hallucinations and Delusions (r = 0.81)

My child thinks others are plotting against him/her.

My child is likely to scream if disturbed.

Somatic Concern (r = 0.84)

Psychosomatic Preoccupation (r = 0.80)

My child is worried about disease.

My child often has an upset stomach.

Muscular Tension and Anxiety (r = 0.68)

Recently my child has complained of chest pains.

My child often has back pains.

Psychological Discomfort (r = 0.90)

Fear and Worry (r = 0.72)

My child will worry a lot before starting something new.

My child is often afraid of little things.

Depression (r = 0.87)

My child has little self-confidence.

My child hardly ever smiles.

(Continues)

TABLE 7.7   (continued)

Psychological Discomfort (r = 0.90)

   Sleep Disturbance/Preoccupation with
   Death
   (*r* = 0.76)

   My child's sleep is calm and restful.

   My child thinks about ways to kill himself/
   herself.

Social Withdrawal (r = 0.81)

   Social Introversion (*r* = 0. 78)

   My child is usually afraid to meet new
   people.

   Shyness is my child's biggest problem.

   Isolation (*r* = 0.68)

   My child does not like to be close with
   others.

   My child often stays in his/her room for
   hours.

Social Skill Deficits (r = 0.91)

   Limited Peer Status (*r* = 0.84)

   My child often brings friends home.
   (reversed)

   My child is very popular with other chil-
   dren. (reversed)

   Conflict with Peers (*r* = 0.88)

   My child seems to get along with everyone.
      (reversed)

   Other children make fun of my child's
   ideas.

NOTE: From Lachar & Gruber (2001).

developed through rational/ theoretical
approaches.

   Many changes and improvements have
been made in the PIC-2 scales. Content
overlap was either reduced or eliminated
between scales, item-total correlation had
to be high, and validity scales were added
(Lachar & Gruber, 2001). Scale content was
also better articulated with that of the PIY
in order to enhance score comparisons. A
Spanish translation was developed as well.

The PIC-2 also includes a 96-item short
form (the first 96 items of the Standard
Form) called the "Behavioral Summary."

   An overview of the PIC-2 clinical scales
is provided in Table 7.7. In addition to these
scales, the PIC-2 provides three validity
scales (i.e., Inconsistency, Dissimulation,
and Defensiveness) and critical items.

Administration and Scoring

It takes a parent about 40 min to complete
the 275 true–false statements of the PIC-
2. All administrations require at least two
components, an administration booklet
and hand-scoring or computer-scoring
answer sheets. The hand-scoring process
involves the use of four forms with a Criti-
cal Items Summary Sheet as an option.
The use of either PC or mail-in computer
scoring limits the number of components
to only two (administration booklet and
answer sheet).

Norming

The norming sample included 2,306 chil-
dren in the kindergarten through 12th
grades. The normative sample appears to
represent 1998 US Census data – which
were the data available at the time of the
PIC-2 norming – well in terms of ethnicity,
parents' education level, and geographic
region of residence.

   Linear transformations of T-scores were
utilized. The range of derived scores is
limited to T-scores based only on within-
sex comparisons. Therefore, as alluded to
in the discussions of other tests, one is not
able to determine how a child's behavior
compares to that of children in general.
Percentile ranks are also not available.

Reliability

Internal consistency coefficients for the
scales are for the most part acceptable and
are shown in Table 7.7. The results of one-

week test-retest studies are also generally supportive (see Lachar & Gruber, 2001). Interrater reliability between mothers and fathers was generally very good, with coefficients mostly 0.75 and higher for non-clinic-referred children. One exception was the Somatic Complaints scale and its subscales, with coefficients of 0.49 to 0.54 (Lachar & Gruber, 2001).

Validity

Several types of validity evidence are reported in the PIC-2 manual including criterion-related, differential diagnosis, and factorial validity. Factors corresponding to the Externalization, Internalization, Social Adjustment, and Total composite scores are described.

The relations between PIC-2 scores and external indicators of adjustment are described in detail in the manual (see Lachar & Gruber, 2001). Some of the indicators include teacher SBS and child self-report PIY ratings. Unfortunately, such studies, by being limited to the PIC "family" of measures, do not allow clinicians to determine the degree to which PIC-2 results will differ from CBCL, BASC-2, MMPI-A, or other results. Evidence of this nature is important, as clinicians often use multiple measures and frequently have to describe their findings in comparison to previous evaluation results. The extent of PIC-2 criterion-related validity evidence to be found in the manual is sometimes difficult to discern. Considerable reference is made to SBS and PIY validity studies.

Children with diagnoses in the clinical samples were used to compare PIC-2 results for several diagnostic groups using MANOVAs. Many significant effects were found. However, sensitivity, specificity, and other typical indices of diagnostic accuracy are not provided.

As is the case with the PIY, independent evidence of validity is difficult to obtain at this time. Several aspects of validity remain to be assessed in order to support clinicians'

use of the scale. First priority for further validation is to assess the criterion-related validity of the PIC-2 with widely used scales, such as the CBCL and BASC-2 PRS because many clinicians will be faced with having to interpret PIC-2 results in tandem with these measures.

Interpretation

Chapter 3 of the PIC-2 manual provides considerable guidance to the user. In fact, the sheer amount of tabular information presented is potentially overwhelming. The frequency of item endorsements for various samples, for example, is presented for each scale. The value of such information is questionable because it is based on the assumption that an item response is a reliable and valid indicator of some construct, which is a dubious assumption. Nevertheless, the manual provides numerous useful case studies and correlates of profiles. In addition, the meaning of various T-scores for the individual scales is thoroughly described in an additional set of tables. Clinicians will probably find these descriptions of T-score outcomes to be valuable for deriving score meaning.

Otherwise, we reiterate our recommended sequential approach to interpretation (i.e., checking validity scales, critical items, scale elevations, subscale elevations, relevant item endorsements, considering primary vs. secondary concerns, integration with other information).

Strengths and Weaknesses

PIC-2 strengths include:

1. A thorough manual by Lachar and Gruber (2001) that summarizes important studies of scale development.

2. A great variety of subscale scores that may be of value for specialized uses.

3. The inclusion of valuable interpretive guidance in the manual.

4. Norming sample that closely matches census data at the time of the scale's development.

Weaknesses of the PIC-2 may include:

1. Test length.
2. A lack of criterion-related validity studies and shortage of validity studies independent of the test developers.
3. Limited score options (i.e., absence of general norm-referenced comparisons and percentiles).

The PIC-2 represents a significant upgrade of the PIC-R. The most important improvements are a reduction of item overlap between scales and the collection of new norms. Both independent validation research and clinical experience are necessary to determine the ultimate utility of the scale.

Teacher Report: The Student Behavior Survey

The Student Behavior Survey (SBS; Lachar, Wingenfeld, Kline, & Gruber, 2000) is the teacher version of the rating scale system that includes the parent-completed PIC-2 youth self-report PIY. As a result, SBS rounds out a rating scale system with a long and distinguished history in the assessment of children and adolescents by providing a source of information on a child's classroom adjustment based on teacher report. The SBS is not as long as its parent-report and self-report siblings, containing 102 items that are rated on a four-point Likert scale. This rather moderate length allows most teachers to complete the form easily in 15–20 min. The scale has normative information for children of ages 5 through 18.

Content

Despite being developed to complement the PIC-2 and PIY scales, the SBS was not beholden to the item content of the parent-report and self-report scales. Instead, the content of the SBS was developed based on teacher endorsements of statements that seem to reflect important dimensions of classroom adjustment. The content of the SBS can be divided into three major categories. The first category is Academic Resources, which contains four subscales: Academic Performance (eight items), Academic Habits (thirteen items), Social Skills (eight items), and Parent Participation (six items). These subscales are adaptive scales focusing on potential strengths of the child in the academic environment, and therefore, items on these subscales are worded in a positive direction. The second category is Adjustment Problems, which includes seven subscales: Health Concerns (six items), Emotional Distress (fifteen items), Unusual Behavior (seven items), Social Problems (twelve items), Verbal Aggression (seven items), Physical Aggression (five items), and Behavior Problems (fifteen items). These two areas include the main clinical scales of the SBS focusing on emotional, social, and behavioral areas of concern for the child's classroom adjustment.

The third section is a Disruptive Behavior Disorders category that includes three subscales: Attention-Deficit Hyperactivity (16 items), Oppositional-Defiant (16 items) and Conduct Problems (16 items). As the names of the subscales imply, these scales were developed to provide a screening for the major disruptive behavior disorder categories specified in the *DSM-IV*. However, the individual items were not specifically developed to tap *DSM* criteria. Instead, three clinicians chose items from the existing 102-item pool that were judged to be most indicative of the *DSM-IV* criteria, a similar approach to that employed for the Achenbach measures

(discussed earlier). This procedure led to some criteria not being assessed (e.g., "Is spiteful and vindictive") and other items included that are not part of the *DSM* criteria (e.g., "Demonstrates polite behavior and manners" reverse-scored). This issue is especially relevant to the Conduct Problems scale, which is fairly divergent from the content of the *DSM-IV* definition of Conduct Disorder, including such items as "uses drugs or alcohol" and "preoccupied with sex."

Administration and Scoring

The items on the SBS are grouped according to their subscales, such that the 8 items for the Academic Performance subscale are items 1 through 8, the 13 items for the Academic Habits subscale are items 9 through 21, and so on. In addition, the subscale titles document this explicit grouping to the teacher raters. This is a somewhat unique format in that other rating scales have items for the subscale intermixed throughout the scale. There could be both positive and negative consequences of this format. For example, it makes scoring much easier and reduces the likelihood of clerical errors in computing raw scores, because it is readily apparent which items are included on each subscale. Also, it makes inspection of items that led to subscale elevations a very simple process. Alternatively, it opens the possibility that teachers may be influenced by the name of the construct (e.g., social skills) and rate children according to their overall perceptions of a child's adjustment for that domain rather than basing their ratings on their perceptions of the individual behaviors. For example, a teacher who views a child as socially unskilled may rate items under that heading as more problematic than if he or she was not explicitly informed about the overall domain being assessed.

However, there is no empirical evidence that this item format affects ratings in any systematic way, and as mentioned previously, it greatly simplifies the scoring process. There are two "Auto-Score" forms for the SBS: one for children of ages 5–11 and one for adolescents of ages 12–18. Raw scores are simply computed by summing the ratings within each of the 11 subscales included in the Academic Resources and Adjustment Problem domains. Between the two sides of the ratings is carbon paper that copies ratings on only those items that correspond to the three disruptive behavior subscales. Raw scores are based on a sum of these items as well. These 14 raw scores are then transferred to a cover Profile page with separate columns for boys and girls. These profiles reflect a conversion to T-scores and show the relative elevations among subscales based on this norm-referenced score. Importantly, the conversions and profiles can only be computed for separate male and female norms, and not for both sexes combined.

Norms

The primary normative sample for the SBS includes 2,612 children from regular education classrooms from 22 schools in 11 states. The sample was fairly evenly divided between boys and girls and had substantial representation at each year of age from 5 to 18. Also, the regional breakdown, parental educational level, and ethnic composition (e.g., 70% Caucasian, 15% African American, 10% Hispanic American) was fairly representative of US Census Bureau statistics (see Lachar et al., 2000). The one relatively minor exception was the somewhat high rate of college graduates in this norm sample (i.e., 35 vs. 26.9% cited for the US Census).

One of the unique features of the SBS is that, in addition to the regular education norm sample on which T-score conversa-

tions were based, the manual also reported on a large referred sample ($n = 1,315$) that obtained teacher ratings on children from 41 different sites in 17 states in the USA. These children included those in special education classes, those referred to both inpatient and outpatient mental health clinics, and those referred to juvenile justice facilities. This large sample allows for a comparison of the psychometric properties of the SBS in both a large normal sample of children and in a large disturbed sample. Overall, each of the SBS scales differentiated the referred and normal samples with Cohen's d ranging from 0.23 (Parent Participation) to 0.98 (Academic Performance; see Lachar et al., 2000).

Reliability

The information provided in the manual (Lachar et al., 2000) on the reliability of the SBS is exemplary. Internal consistency estimates for the 14 subscales across both the normal and referred samples ranged from 0.85 to 0.95, indicating uniformly excellent internal consistency. Test-retest correlations are provided for four samples of children ranging in age from 5 to 18 and with retest intervals ranging from 2 to 30 weeks. Again, all scales showed quite good temporal stability, with the test-retest of the Unusual Behavior scale over a 20-week period in adolescents being the only index to be somewhat low (i.e., $r = 0.29$). A third type of reliability, inter-rater agreement, was tested in two samples of 30 children, one sample including fourth and fifth grade regular education students and a second sample including children (ages 5–12) receiving special education services. The correlations between two teacher ratings across these samples ranged from 0.44 to 0.91, with most indexes being above 0.70.

Validity

The dimensionality of the SBS was tested in a way that was somewhat different from that reported for other behavioral rating scales. That is, rather than conducting a factor analysis on the individual items, the item-subscale correlations were compared for each item's correlation with the dimension it is purported to assess and its correlations with other dimensions. While this method led to rationally derived scales that were fairly homogeneous in content, the decision as to whether an item is "more strongly" associated with the dimension it is purported to measure is somewhat subjective in the absence of factor analysis. For example, "Blames others for own problems" is correlated 0.79 with the Behavior Problems subscale on which it is included, but it is also correlated 0.76 with the Verbal Aggression subscale, 0.61 with the Physical Aggression subscale, and 0.54 with the Social Problems subscale. The most problematic in this regard are the three Disruptive Behavior Scales, on which many items load equally high on all three dimensions, although this is likely due to the nature of the criteria they were designed to assess, which tend to be substantially overlapping (Frick et al., 1994).

The manual of the SBS (Lachar et al., 2000) provides (1) the correlations of the SBS subscales with clinician ratings of adjustment problems in 129 primarily clinic-referred children, (2) the correlations among SBS scores and parent- and self-report ratings using the PIC-2 and PIY, and (3) the correlations between the SBS and an early version of the Conners Rating Scale for teachers (see also Pisecco et al., 1999; Wingenfeld, Lachar, Gruber, & Kline, 1998).

In general, these correlations support the convergent validity of the SBS scales, but like most rating scales, the divergent validity was less clear. That is, the SBS subscales were often correlated with the other scales designed to measure similar constructs (i.e., convergent validity), but they were also correlated with other dimensions

of maladjustment as well. For example, the Emotional Distress subscale was significantly correlated with clinician ratings of psychological discomfort ($r = 0.55$), but this subscale was also highly correlated with the ratings of disruptive behavior ($r = 0.44$). Again, this pattern is common for ratings of children's adjustment because children with problems in one area often have problems in many other areas of adjustment as well, and raters may also demonstrate response sets in that a child rating negatively in one area is rated similarly in other areas. One notable weakness uncovered in these validity analyses was for the Unusual Behavior subscale, which seemed to be more strongly associated with measures of disruptive behaviors and ADHD than with more severe psychopathology or reality distortion. For example, it was correlated 0.40 with clinician ratings of disruptive behavior but 0.25 with clinician ratings of serious psychopathology. Similarly, the Unusual Behavior subscale was correlated at 0.41 with parent ratings of impulsivity and distractibility on the PIC-2, but at 0.27 with the Reality Distortion subscale of the PIC-2.

One additional set of validity analyses provided in the manual were comparisons between groups of children either diagnosed with Disruptive Behavior Disorders by clinicians or children elevated on the Hyperactivity Index in an earlier version of the Conners Rating Scale compared to control children. As would be expected, the Social Problem subscale, the three behavior problem subscales, and the disruptive behavior disorder subscales all differentiated children with behavior problems from control children. Also as expected, the academic resources subscales tended to be lower in groups of children with behavioral problems, with the exception of the Parent Participation subscale.

Interpretation

Within the tradition of the PIC-2, which, in turn, was based on the MMPI tradition, the manual of the SBS provides a very detailed step-by-step interpretative guide (Lachar et al., 2000). First, the manual recommends examining items for response adequacy, including ensuring that there are only a few missing responses. The one exception noted in the manual is that many teachers above the early elementary school grades may have difficulty completing the Parent Participation scale because they are less likely to converse with parents on a regular basis (Lachar et al., 2000). Also, it is important to note that, unlike the PIC-2 and PIY, there are no validity indexes on the SBS designed to help in detecting potential threats to the quality of the teacher ratings. Second, and the main focus of the interpretative approach in the manual, is a description of the characteristics that are often associated with children who score in a given range on each subscale.

These interpretive guidelines were developed by correlating the T-scores on the SBS subscales with descriptors provided by clinicians ($n = 379$), parents ($n = 425$), and students ($n = 218$). Descriptors are provided for T-scores below 40 for the academic resources subscales and for (1) T-scores between 60 and 70, and (2) T-scores above 70 for the adjustment problems scales. The authors note that the descriptors for the higher elevations (above 70) should be considered more definitive than those between 60 and 70. The authors clearly note, however, that all interpretations, even those above 70, should be considered only as "interpretative hypotheses," and additional information (e.g., from parent report, child self-report, and clinical observations) should be used to better determine if these hypothetical descriptors are appropriate for a given case.

Strengths and Weaknesses

The strengths of the SBS include:

1. Content that includes a number of adaptive dimensions of classroom adjustment and a rather comprehensive assessment of conduct problems, including separate subscales for verbal and physical aggression, and a general Behavior Problems subscale.

2. Fairly homogeneous subscale content, which greatly enhances the interpretation of scale elevations, as does the very easy-to-use, step-by-step interpretive guidelines, which provide the most common characteristics for children with specific scale scores.

3. A large and representative norm sample.

4. The evidence for subscale reliability using both community and clinic-referred samples is exemplary.

All of these characteristics make the SBS a very useful tool for obtaining teacher ratings of classroom adjustment.

Weaknesses of the SBS include:

1. Limited research on the validity of the SBS scales and subscales

2. A lack of cross-validation in other samples of the interpretative descriptors provided for children who score in a specific range on each subscale need to be cross-validated.

3. The heterogeneous content of the Unusual Behavior subscale includes some items related to inattention (e.g., "Daydreams") and some vague behaviors (e.g., "Behavior is strange and peculiar"). Early evidence suggested that it is more associated with

---

Box 7.3

Sample Case Using the PIC-2 and SBS

Ricky is a 7-year-old boy who was referred for an evaluation by his teacher because of concerns about inattention, overactivity, and poor peer interactions. Ricky was reportedly born at home without the benefit of any medical attention

Most of Ricky's developmental milestones were delayed, especially language. According to his mother, he did not speak his first words until the age of 2 years. He has reportedly demonstrated some improvements in language since beginning speech therapy at age 4.

According to his mother, as a toddler, Ricky began demonstrating significant behavioral problems including frequent temper tantrums, overactivity, and oppositional behavior. Such behavioral concerns continued when Ricky began school. He has a history of getting in trouble at school due tobecause of overactivity and defiance toward his teachers.

Throughout the evaluation, Ricky demonstrated a short attention span and a high

level of motor activity. He did not sit still and was easily distracted by other objects in the room. When he was unable to testing objects and toys in the waiting room with him, Ricky displayed tantrum behavior (e.g., kicking and crying).

Ricky's cognitive assessment results indicated overall functioning in the Low Average range, with his verbal skills being in the Borderline range. Tests of his achievement in reading and math indicated slightly below average achievement relative to his same-aged peers.

PIC-2 Scale elevations were as follows:

Impulsivity/Distractibility	$T = 80$
Delinquency	$T = 82$
Family Dysfunction	$T = 75$
Social Skill Deficits	$T = 68$

(Continues)

Box 7.3   (Continued)

SBS Scale elevations were as follows:

Academic Habits (adaptive)	$T = 33$
Social Skills (adaptive)	$T = 32$
Emotional Distress	$T = 71$
Social Problems	$T = 67$
Unusual Behavior	$T = 63$
Verbal Aggression	$T = 80$
Physical Aggression	$T = 72$
Behavior Problems	$T = 82$
Attention-Deficit/Hyperactivity	$T = 73$
Oppositional Defiant	$T = 73$

On rating scales, both Ricky's mother and teacher reported significant concerns regarding inattention, overactivity, impulsivity, and externalizing behaviors. These behavioral concerns have apparently been present for some time. Furthermore, based on background information as well as ratings by Ricky's mother and teacher, Ricky's behavioral problems are causing impairment in his relationships with others and his ability to perform required tasks in the classroom.

Therefore, diagnoses of Attention-Deficit/ Hyperactivity Disorder (ADHD)-Combined Type and Oppositional Defiant Disorder are warranted. The elevation on the SBS Emotional Distress scale is consistent with reports that Ricky gets upset easily and throws tantrums in the classroom. The slight elevation the Unusual Behavior scale of the SBS completed by Ricky's teacher appears consistent with attention problems in that she reported that he daydreams and seems disoriented. In addition, Ricky's mother indicated concerns regarding a high level of conflict in the home, particularly between Ricky and his parents. This conflict appears to also be related to Ricky's behavioral problems, particularly his defiance.

Recommendations for Ricky included parental consultation with a mental health specialist to address his behavioral problems, consultation with a physician regarding a possible medication trial for his ADHD symptoms, and classroom accommodations to help minimize the impact of Ricky's behavioral concerns and inattention on his academic and social functioning. The results of parent and teacher rating scales in this case highlight the dissimilar structure of the PIC-2 and SBS. However, similar information can still be gleaned from these rating scales that can aid in case conceptualization and recommendations.

disruptive behavior dimensions than with indexes of more severe psychopathology and thought disturbances.

4. The lack of direct correspondence between the three disruptive behavior disorder subscales and *DSM* criteria. This is especially true for the Conduct Problems scale, which appears quite divergent from the criteria for Conduct Disorder. In addition, there is no evidence for how well the specific SBS subscales (e.g.,

Attention-Deficit Hyperactivity) correspond to specific *DSM-IV* diagnoses (e.g., ADHD). As a result, the usefulness of SBS as a screener for specific *DSM* disorders has not been established.

In addition to these issues, it is worth noting that while SBS was developed to be part of the assessment system that includes the PIC-2 and PIY (reviewed previously) the item content and scale structure of the SBS is substantially different from these

other scales. The result is a tool that is very relevant for assessing children's classroom functioning. However, it also makes it more difficult to integrate information from the different informants. A case example with PIC-2 and SBS data follows (Box 7.3).

Sample Impairment-Oriented Scales

As can be determined from the previous review, omnibus rating scales can provide invaluable information about a variety of domains of child functioning. This information, however, tends to describe functioning in terms of severity of problems and/or frequency of problems. Rating scales typically stop short of providing an indication as to what extent the problems interfere with the child's functioning. Information on impairment is often left to the clinician to infer based on interview or other information. However, this information is no less important for case conceptualization and treatment planning. In addition to assessing for impairment via structured or unstructured interviews, one may employ an inventory to gather such information in a time-efficient manner and then follow-up accordingly. A brief discussion of some such inventories follows.

Home Situations Questionnaire (HSQ) and School Situations Questionnaire (SSQ)

The content of the Home Situations Questionnaire (HSQ; Barkley & Edelbrock, 1987) and the School Situations Questionnaire (SSQ; Barkley & Edelbrock, 1987) is markedly different from the other rating scales reviewed in this chapter. Rather than having items that describe different types of child behaviors, these measures include situations (e.g., while playing alone, when visitors are in the home, during individual

desk work, at recess, on the bus) in which a child may have problems. That is, the HSQ and SSQ were not designed to assess specific behaviors but to assess specific situations in which problem behaviors can occur. Therefore, these measures provide an indication of the specific situations in which the child may demonstrate particular difficulty or impairment.

Both measures were designed to be completed in the same manner. The respondent (parent or teacher) rates whether or not the child has any problem in a given situation and then rates the severity of the problem on a 1–9 scale. These measures may be used with a variety of clinical problems, as the respondent can be directed to respond as to whether or not the child "has problems" in the situations provided.

The psychometric development of both measures is limited. Normative information is available from Altepeter and Breen (1989) as well as Barkley and Edelbrock (1987). However, norm-based comparisons may not represent the best use of these tools. Factor analyses have revealed four factors for the HSQ (i.e., Non-Family Transactions, Custodial Transactions, Task-Performance Transactions, and Isolate Play) and three factors for the SSQ (i.e., Unsupervised Settings, Task Performance, and Special Events; Altepeter & Breen, 1989).

The HSQ has demonstrated good test-retest reliability and internal consistency (Altepeter & Breen, 1989). The number of problems and mean severity rating of the HSQ have been found to be related to ratings of impulsivity and hyperactivity (Altepeter & Breen, 1989). Test-retest reliability of the SSQ in a sample of 119 regular education children was estimated at 0.68 for the number of problem situations and 0.78 for the mean severity score (Barkley & Edelbrock, 1987). Also, inter-rater agreement for the SSQ was tested in a sample of 46 students ages 8–17. The correlation between teachers was 0.68 for the number of problem situations and 0.72 for the mean severity score (Danforth & DuPaul, 1996). Barkley and

Edelbrock (1987) reported numerous significant correlations between the SSQ and rating scale measures of externalizing behavior problems and evidence that the SSQ differentiates children with ADHD from children without ADHD. However, for both the HSQ and SSQ criterion-related validity evidence is more difficult to operationalize, as these measures have a different focus than ratings of symptoms or problems. Still, situations in which the child has difficulties, as indicated on the HSQ and SSQ, can assist the clinician in appropriately designing and prioritizing intervention strategies.

Child Global Assessment Scale (CGAS)

Another example of an assessment of impairment takes a different approach. The Child Global Assessment Scale (CGAS; Shaffer et al., 1983) is an adaptation of an adult scale designed to assess overall level of impairment at home, in school, or with friends. The scale extends from a low of 1 (extremely impaired) to a high of 100 (no impairment). A parent, teacher, or interviewer is asked to rate the child on this scale where deciles are accompanied by a descriptor (e.g., 51–60, "some noticeable problems"). Previous studies have demonstrated some evidence of reliability and validity. A cut score is commonly used in studies of child psychopathology (e.g., CGAS 70 or below identifies a clinical case).

The CGAS was used as one of the criteria for validating the *DSM-IV* criteria for the diagnosis of ADHD (Lahey et al., 1994). Lahey and colleagues used a CGAS score of 60 or less as an indication of significant impairment associated with symptoms of ADHD. A noteworthy finding of this study was the differential results for the parent and teacher CGAS scores. The parent CGAS scores were significantly related to symptoms of hyperactivity/impulsivity but not to inattention. Teacher CGAS scores were not significantly related to hyperactivity/impulsivity problems.

These same teacher scores were, however, related to ratings of academic problems. The Lahey et al. investigation then used the relation between teacher and parent CGAS scores and inattention symptoms to shape the DSM-IV criteria for inattention problems associated with ADHD.

The psychometric properties of the CGAS have been well-studied (see review by Schorre & Vandvik, 2004). Of course, the accuracy of CGAS ratings (as is the case for all ratings) depends heavily on the rater's knowledge of the child's functioning in a variety of spheres (Weissman, Warner, & Fendrich, 1990). Can parents, for example, validly rate school and peer functioning as is required by the CGAS? Schorre and Vandvik (2004) call for increased consistency in how clinicians assess and then rate impairment. Certainly consistency in conceptualizing constructs such as attention problems or depression aid in communication and treatment planning. Such could also be the case for assessing impairment caused by these problems.

An additional consideration is whether the best approaches to assessing impairment are already embedded in rating scales such as those reviewed in this chapter. For instance, a study by Bird et al. (1990) found a strong association between CGAS scores and the Total T-score of the CBCL. The most impaired group had a mean Total T of 70, the next most impaired group had a mean of 67, the next group produced a mean of 59, and the no-diagnosis group mean was 53 (Bird et al.). Clinical elevations on standard rating scales may, then, provide an indicator of impairment. However, Mash and Hunsley (2005) concluded that "Assessments of children and adolescents need to focus not only on specific disorders and problems but also on specific impairments that may occur in the absence of a diagnosable disorder" (p. 368). Therefore, it is quite likely and important that measures of impairment will see increasing use in clinical assessment practice (Bird, 1999).

Box 7.4

Assessing Change

Assessment of change is important in clinical practice, particularly in light of increasing calls for accountability for the delivery of health care. Throughout this text, our focus is primarily on assessment as a mechanism to provide an answer to an initial referral question and treatment recommendations. Quite obviously, however, assessment is an ongoing process throughout treatment as well. If for no other reason, assessment of change should be routine because research has shown that such evaluations lead to better fidelity to evidence-based treatments, and ultimately, to better treatment outcomes (Lambert et al., 2003).

Some considerations of assessment tools as useful for answering referral questions may not apply to assessments of change. Demonstration of utility for intervention planning and assessment is a tricky endeavor since psychometric evidence of validity is not clearly applicable to questions of treatment progress. A measure designed for evaluating change, for example, may not need norms. Consequently, the quality of the norming sample is not relevant. Moreover, a rating scale that assesses a child's tendency toward problems in various areas may not be sensitive to changes over a short period of time. Furthermore, it may not be necessary to interpret total scores if one is interested only in change in the individual behaviors. Perhaps ironically, good test-retest reliability may be *undesirable* in the assessment of session-to-session change. Therefore, ratings scales of the nature discussed in this chapter may not be suitable for assessing change – at least not in the short-term.

Clinicians may wish to assess change through instruments that they develop for use with a specific client with a specific treatment plan. Such strategies, if true to the treatment targets, would have strong validity and utility. However, to the extent that standardized measures can be used, a larger evidence base on meaningful ways to assess change will develop. The Eyberg Child Behavior Inventory (ECBI; Eyberg & Ross, 1978) is an instrument that has a long history of use for evaluating treatment results. It should be noted, however, that its

usefulness for assessing particular constructs is not well supported.

The ECBI is a 36-item parent rating scale designed to assess behavior problems for children of age 2 through 17 years (Eyberg & Ross, 1978). Each item is rated in two ways: (1) a Likert-type scale that is used for marking frequency, and (2) a dichotomous scale that the parent uses to identify if the issue is in fact a problem.

The ECBI has been used as a measure of conduct-problem behaviors (Burns & Patterson, 1990), and it does possess some advantages, including the fact that it provides an indication of both frequency and severity for individual behavior problems, which is not common for parent rating scales. This characteristic may make the scale particularly useful for planning and evaluating treatment. The ECBI also produces total scores for both the frequency (Intensity) and severity (Problem) measures.

The available research suggests that the ECBI is an example of an instrument that *may* be useful for identifying treatment objectives for children referred for disruptive behavior problems and for evaluating response to treatment. Certainly, it was ahead of its time in its amenability to evaluations of change. When the ECBI is used for other purposes, such as norm-referenced assessment of constructs, research to date does not reveal significant evidence of construct validity (using the term as outlined by Anastasi, 1988). The ECBI is an excellent example of a scale that has some value only in the hands of the well-informed clinician who applies the scale only in circumstances where it possesses empirical strengths.

We recognize that the preponderance of our attention is devoted to other forms of assessment rather than assessment of change. Nevertheless, an increased research base on individualized as well as large scale approaches to evaluating treatments is imperative. We have every reason to believe that discussions of evidence-based assessment will also include assessments of treatment progress (e.g., Mash & Hunsley, 2005).

Conclusions

Parent and teacher rating scales are now common methods for assessing child problems. The quality of parent and teacher rating scales has improved considerably in recent years. Routinely, scales have national normative samples and provide expansive information about their reliability and validity. In essence, rating scales provide a time-efficient and reliable method for obtaining assessment information from parents and teachers.

We focused primarily on global scales that assess multiple domains of functioning because the nature of childhood problems is such that dysfunction in one domain is often associated with problems in other areas of functioning. Our review of rating scales was not intended to be exhaustive but was designed instead to focus on some of the most commonly used scales and to illustrate what we feel are some crucial areas to consider in evaluating scales for use in a clinical assessment. Also, our overview was not intended to replace a careful reading of the technical manuals of these scales but to highlight some of the important features of the scales that might influence their use in clinical assessments.

Furthermore, the ECBI (Box 7.4) is an example of a parent rating scale that could be used to evaluate change. The Outcomes Questionnaire-45 (OQ-45) is a questionnaire that has been used as a means to provide therapists with feedback from adult clients as often as after every session (Okiishi et al., 2006). The suitability and feasibility of such an approach with parents/child clients and in many clinical settings is uncertain. Therefore, it is likely the case that the clinician is routinely left to evaluate change, whether formally or informally. This strategy has the advantage of being executed by someone trained to define and detect the problems of focus. It has the disadvantage of being utilized by the very person or persons trying to implement and demonstrate the effectiveness of their therapeutic strategies. Research has increasingly addressed the implications of this approach (e.g., Lambert et al., 2003), but relatively little is known. Far less is known about the teacher assessment of changes in behavioral and emotional functioning during and following interventions. An exception would be single-case designs tracking behavioral changes resulting from classroom interventions; many times, these interventions are evaluated by school psychologists or other mental health professionals.

Chapter Summary

1. Concerns about child self-reports and practicality have made parent and teacher rating scales commonplace in modern child assessment practice. These tools tend to be a very efficient means of gathering clinically relevant information.

2. Research has indicated that the construct being evaluated and the child's developmental level influence ratings provided by parents and teachers and even the usefulness of such ratings.

3. The Behavior Assessment System for Children (BASC-2) Parent Rating Scales (PRS) and Teacher Rating Scales (TRS) have three forms of similar items that span the preschool (2–5), child (6–11), and adolescent (12–21) age ranges. The PRS takes a broad sampling of a child's behavior in home and community settings, whereas the TRS does the same for the school setting.

4. The PRS and TRS were developed using both rational/theoretical and empirical means in combination to construct the individual scales.

5. The BASC-2 measures include a relatively comprehensive assessment of adaptive functioning.
6. The Achenbach CBCL and TRF and their predecessors have long been considered one of the premier rating scale measures of child psychopathology.
7. The CBCL and TRF continue to be a preferred choice of many child clinicians because of its history of successful use and popularity with researchers.
8. The CBCL and TRF now include DSM-Oriented scales that are more closely aligned to *DSM* criteria than the Syndrome scales of both measures.
9. The CSI-4 is unique in its content being explicitly tied to the diagnostic criteria in *DSM-IV*. Thus, it provides a screening of severe forms of childhood