Assessment with Brief Behavior Rating Scales

ROBERT J. VOLPE
GEORGE J. DUPAUL
Lehigh University

INTRODUCTION

Behavior rating scales have become an essential part of the psychoeducational assessment of children and adolescents. A wide variety of instruments is now available for obtaining information on how the behavior of youngsters is viewed by parents, teachers, and the children themselves. These measures can be categorized by their bandwidth, or the number of constructs or syndromes they purport to measure. Specifically, broad-band scales measure multiple syndromes or behavioral domains (e.g., internalizing and externalizing), and narrow-band scales contain items focused upon one or two behavioral domains (e.g., attention-deficit/hyperactivity disorder, or ADHD). In a typical behavioral assessment procedure, broad-band rating scales such as the Child Behavior Checklist and/or the Teacher Report Form (CBCL, TRF; Achenbach, 1991a, 1991b) are administered first, to cast a wide net of measurement across several behavioral dimensions. Then, narrow-band scales may be utilized to gather more domain-specific information. The ADHD Rating Scale—IV (DuPaul, Power, Anastopoulos, & Reid, 1998) is an example of a narrow-band scale because it measures a single syndrome (e.g., ADHD).

This chapter focuses on narrow-band rating scales completed by children and the adults who have contact with them. These instruments offer several advantages to practitioners. Several commercially available narrow-band
rating instruments are well validated and normed. In many instances, such instruments allow the acquisition of detailed information concerning child emotions and behaviors of interest across several informants and settings, making them useful sources of diagnostic information. These instruments are also cost effective in that they are easily administered and scored. Other advantages of narrow-band scales include their utility in monitoring child behavior during treatment and their ability to assess low-frequency behaviors that are difficult to evaluate with other methods (e.g., direct observations).

The rating scales reviewed in this chapter can be used at several assessment stages. First, due to their relative brevity, they may be useful as screening instruments for specific behavioral or emotional problems. Next, once children meet screening criteria, these measures may be administered to additional informants to gather information on the behaviors of interest across settings. Finally, the instruments may also be useful for the monitoring of treatment effects.

This chapter reviews and compares three measures that have been designed to measure symptoms of child externalizing behavior: the ADHD-IV, the ADHD Symptom Checklist—4, and the Short Forms of the Conners’ Revised Parent and Teacher Rating Scales. We also review and compare the following child self-reported measures of internalizing symptoms: the Child Depression Inventory (CDI), the Revised Children’s Manifest Anxiety Scale (RCMAS), the Reynolds Child Depression Scale (RCDS), and the Reynolds Adolescent Depression Scale (RADS). It is our goal to provide a quick reference source to practitioners interested in using rating scales in their clinical practices. In addition to summarizing each measure’s purpose, we provide a synopsis of each measure’s reliability and validity as well as a description of its standardization, administration, scoring, usability, and utility. Finally, we compare and contrast rating scales that are focused on the same class of behavior. Due to space restrictions, we limit our review to the most useful and widely studied narrow-band scales relevant to those childhood emotional and behavior problems with the highest prevalence: ADHD and related disorders, depression, and anxiety. It is important to note, however, that narrow-band rating scales are available to assess other childhood disorders and behavioral concerns (e.g., pervasive developmental disorders) that are less common (see Table 12.1).

MEASURES TO ASSESS EXTERNALIZING BEHAVIOR

Occurring in 3–6% of the school-aged population (American Psychiatric Association, 1994), ADHD is the most common childhood psychiatric disorder (Taylor, 1995). Children with ADHD are principally affected by developmentally inappropriate levels of inattention and/or overactivity-impulsivity that
TABLE 12.1
Behavior Rating Scales to Assess Childhood Disorders

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Rating Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduct Disorder</td>
<td>Eyberg Child Behavior Inventory (Eyberg, 1992)</td>
</tr>
<tr>
<td></td>
<td>Sutter-Eyberg Student Behavior Inventory (Eyberg, 1992)</td>
</tr>
<tr>
<td></td>
<td>Children's Hostility Inventory (Kazdin, Rodgers, Colbus, &amp; Siegel, 1987)</td>
</tr>
<tr>
<td></td>
<td>Interview for Antisocial Behavior (Kazdin &amp; Esveldt-Dawson, 1986)</td>
</tr>
<tr>
<td>Autistic Disorder</td>
<td>Childhood Autism Rating Scale (Schopler, Reichler, &amp; Renner, 1986)</td>
</tr>
<tr>
<td></td>
<td>Autism Behavior Checklist (Krug, Arick, &amp; Almond, 1980)</td>
</tr>
<tr>
<td></td>
<td>Checklist for Autism in Toddlers (Baron-Cohen, Allen, &amp; Gillberg, 1992)</td>
</tr>
<tr>
<td>Eating Disorders</td>
<td>Psychiatric Rating Scale for Anorexia Nervosa (Goldberg, Halmi, Casper, Eckert, &amp; Davis, 1977)</td>
</tr>
<tr>
<td></td>
<td>Binge Eating Scale (Gormally, Black, Daston, &amp; Rardin, 1982)</td>
</tr>
<tr>
<td></td>
<td>Eating Attitudes Test (Garner &amp; Garfinkel, 1979)</td>
</tr>
<tr>
<td></td>
<td>Eating Disorders Inventory (Garner, Olmsted, &amp; Polivy, 1983)</td>
</tr>
</tbody>
</table>

must be present in two or more settings (see Table 12.2). More severe externalizing problems (e.g., Oppositional Defiant Disorder and Conduct Disorder) as well as internalizing problems commonly have been found to co-occur with ADHD (August, Realmuto, MacDonald, Nugent, & Crosby, 1996).

TABLE 12.2
Examples of Attention-Deficit/Hyperactivity Disorder Symptoms

<table>
<thead>
<tr>
<th>Inattention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Often does not pay attention to details or makes careless errors</td>
</tr>
<tr>
<td>Often has difficulty paying attention</td>
</tr>
<tr>
<td>Often does not seem to listen</td>
</tr>
<tr>
<td>Hyperactivity</td>
</tr>
<tr>
<td>Often fidgets or squirms around</td>
</tr>
<tr>
<td>Often gets out of seat when he or she is not supposed to</td>
</tr>
<tr>
<td>Often talks too much</td>
</tr>
<tr>
<td>Impulsivity</td>
</tr>
<tr>
<td>Often blurts out answers to questions</td>
</tr>
<tr>
<td>Often cannot wait turn</td>
</tr>
<tr>
<td>Often butts in</td>
</tr>
</tbody>
</table>
Hence, practitioners assessing youngsters at-risk for ADHD may find themselves using narrow-band rating scales that assess both externalizing and internalizing domains.

The three measures reviewed in this section (see Table 12.3) are all psychometrically sound instruments that could aid a clinician in determining whether the frequency of ADHD symptoms displayed by a specific child is developmentally deviant for that child's age and gender (i.e., by comparing a child's scores with normative data). It should be noted, however, that narrow-band rating scales are only one component of a comprehensive multimethod evaluation of ADHD that typically includes diagnostic interviews with the parent and teacher, behavior rating scales, and direct observations of behavior (Barkley, 1998). A case study at the end of this section illustrates the use of narrow-band measures as part of a comprehensive evaluation of a child suspected of having ADHD.

<table>
<thead>
<tr>
<th>TABLE 12.3</th>
<th>Summary of Measures to Assess ADHD and Related Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD-IV</td>
<td>ADHD Symptom Checklist — 4</td>
</tr>
<tr>
<td>Authors</td>
<td>Kenneth D. Gadow &amp; Joyce Sprafkin (1997)</td>
</tr>
<tr>
<td>Publisher</td>
<td>C. Keith Conners (1997)</td>
</tr>
<tr>
<td>Copyrighted</td>
<td>Yes</td>
</tr>
<tr>
<td>Time to Complete</td>
<td>&lt; 15 minutes</td>
</tr>
<tr>
<td>Scoring Software</td>
<td>No</td>
</tr>
<tr>
<td>Ages</td>
<td>Parent — 3–18</td>
</tr>
<tr>
<td>Items</td>
<td>Teacher — 27</td>
</tr>
<tr>
<td>Subscales</td>
<td>3 ADHD subtypes</td>
</tr>
<tr>
<td>Scaling</td>
<td>0–3 for Symptom Severity — 0–3</td>
</tr>
<tr>
<td></td>
<td>0–1 for Symptom Cutoff — 0–1</td>
</tr>
</tbody>
</table>
### ADHD Rating Scale—IV

**General Overview and Psychometric Characteristics**

The home and school versions of the ADHD Rating Scale—IV (ADHD-IV; DuPaul, Power, Anastopoulos, & Reid, 1998) are 18-item questionnaires designed to gather information from parents and teachers, respectively, about child symptoms of ADHD. The items mirror the diagnostic criteria of the Diagnostic and Statistical Manual—Fourth Edition (DSM-IV; American Psychiatric Association, 1994), and are administered in a Likert format (e.g., “never or rarely” = 0, “sometimes” = 1, “often” = 2, and “very often” = 3). Respondents are asked to endorse each item as it best describes the target child in the past six months (or if the child is new to the teacher, since the beginning of the school year). Items on the home and school versions are identical.

The reliability and internal consistency of the ADHD-IV are generally good. Specifically, test-retest coefficients (for administrations four weeks apart) were between .88 and .90 for the three subscales of the school version and between .78 and .86 for the home version. Coefficient alphas were consistently high for both versions of the checklist (alphas ≥ .86; DuPaul, Power, McGoey, Ikeda, & Anastopoulos, 1998).

The results of exploratory and confirmatory factor analyses performed on the home and school versions of the ADHD-IV suggest that both a two-factor (e.g., Inattention and Hyperactivity/Impulsivity) and a single-factor solution (ADHD) fit the data well (DuPaul et al., 1997, 1998). Some analyses favored the two-factor solution in correspondence with the DSM-IV model.

The discriminant and concurrent validities of the ADHD-IV have been investigated and appear adequate. Scores on the ADHD-IV subscales of
Inattention and Hyperactivity/Impulsivity were found to significantly discriminate between children diagnosed with ADHD-combined type (ADHD-COM), ADHD-inattentive type (ADHD-I), and psychiatric controls in a sample of 92 consecutive referrals to a child psychiatric outpatient clinic (DuPaul, Power, McGoey et al., 1998). Significant correlations have been obtained between subscales of the home version of the ADHD-IV and the Conners' Parent Rating Scale—48 (CPRS) as well as between subscales of the school version of the ADHD-IV and the Conners' Teacher Rating Scale—39 (CTRS; DuPaul, Power, McGoey et al., 1998).

The predictive validity of the ADHD-IV was investigated by using a logistic regression procedure (Power et al., 1998). The relevant subscales of the ADHD-IV were able to differentiate between ADHD diagnostic groups, as well as between children diagnosed with ADHD and psychiatric controls. Teacher ratings were shown to be superior discriminators of diagnostic status. However, parent ratings were also significant predictors and, when used in conjunction with teacher ratings, added to the strength of prediction. There are no data available concerning the sensitivity of the ADHD-IV to treatment effects.

Adequate normative data are available for the ADHD-IV (DuPaul, Anastopoulos et al., 1998; DuPaul, Power, McGoey et al., 1998). The school and home versions were standardized on samples of 2000 children and adolescents between the ages of 4 and 20 drawn from over 20 school districts across the United States. Each sample approximated the 1990 Census data for ethnic group and geographic region. Scores on both versions of the ADHD-IV were found to vary significantly by age, gender, and ethnic group. As would be expected, boys tended to receive higher ratings than girls, and older children received lower ratings than younger children. African-American children received higher ADHD-related ratings than did Latino-American or Caucasian children, even when the effects of socioeconomic status were statistically controlled. These findings led the authors of the ADHD-IV to provide separate standardization data by age and gender. Due to insufficient cell size, however, they were unable to develop separate scoring data based on ethnic background. Hence, one should exercise caution when interpreting ratings for African-American children to minimize false positives.

**Administration and Scoring**

The ADHD-IV manual includes all of the materials needed to administer and score both the home and school versions and provides permission for the purchaser to reproduce the checklists and scoring sheets for use with their own clients and students. The ADHD-IV is comprised of two 9-item subscales (Inattention and Hyperactivity/Impulsivity) and should take a parent or teacher no more than 5 minutes to complete. Scoring of the ADHD-IV involves summing the values of the odd-numbered items for the Inattention raw score, summing the even-numbered items for the Hyperactivity/Impulsivity...
sivity raw score, and then summing these two scores to obtain the Total raw score. To obtain percentile scores, the practitioner simply selects the appropriate score sheet (e.g., Home Version for Girls), finds the appropriate column based upon subscale and age (e.g., Hyperactivity/Impulsivity, 5–7), and circles the appropriate raw score (e.g., 19). The corresponding percentile score is printed on the far right or left of the score sheet.

Usability and Usefulness

The ADHD-IV is a well-developed and validated instrument that offers several advantages in the assessment of ADHD in children and adolescents. First, the instrument can be used to obtain reliable data on the presence of ADHD-related symptoms across multiple settings, and is useful in discriminating between children with and without ADHD and between children with different subtypes of ADHD. Second, the instrument is brief, affording easy administration and scoring. Because of the brevity of the scale, it is especially well-suited to monitoring treatment effects. Unfortunately, there is heretofore little evidence for the instrument’s sensitivity to treatment effects. However, the authors of the ADHD-IV do provide statistical guidelines and resources for using the instrument to monitor changes in behavior (DuPaul, Power, Anastopoulos et al., 1998). Given its brevity, this scale lends itself to repeated administrations during the course of a stimulant medication trial or to monitor behavior change due to a classroom intervention. In such cases, one would need to modify instructions about the time frame to consider when rating the child’s behavior (i.e., specify a treatment time frame other than over the previous 6 months).

ADHD Symptom Checklist—IV

General Overview and Psychometric Characteristics

The ADHD-Symptom Checklist—IV (SC-4; Gadow & Sprafkin, 1997) is a 50-item parent- and teacher-completed checklist composed of four categories: (a) ADHD, (b) Oppositional Defiant Disorder (ODD), (c) the Peer Conflict Scale, and (d) the Stimulant Side Effects Checklist. Items of the ADHD and ODD categories are highly similar to individual diagnostic criteria for the corresponding disorders set forth in the DSM-IV. The SC-4 was developed for several uses. First, the SC-4 was designed as a screening instrument for the most common causes of referral (e.g., disruptive child behavior) to child psychiatric clinics and to monitor changes in these symptoms during treatment. Second, given the prescription rate of psychostimulants in children with externalizing behavior difficulties, the developers of the SC-4 provided a measure of stimulant side effects that includes three indices (Mood, Attention-Arousal, and Physical Complaints).

No internal consistency data are available for the SC-4. The test-retest
reliability of the SC-4 appears adequate. Reliability coefficients for the symptom severity scores of the ADHD, ODD, and Peer Conflict categories (6-week latency) ranged from .67 to .89.

No factor analytic data are available for the SC-4. The discriminant and concurrent validities of the SC-4 have been investigated and appear adequate. With a few exceptions, scores on the SC-4 have been shown to discriminate between “normal” and clinically referred groups of children and adolescents (Sprafkin & Gadow, 1996; Gadow & Sprafkin, 1997). Supporting the concurrent validity of the SC-4, the manual reports moderate to high correlations between the SC-4 categories and commonly used checklists such as the CBCL, TRF, the Mother’s Objective Measure for Subgrouping (MOMS; Loney, 1984), and the IOWA Conners’ (Loney & Milich, 1982).

The predictive validity of the SC-4 was assessed by investigating the degree to which cutoff scores on various SC-4 categories agreed with relevant clinical diagnoses. The statistics of sensitivity (the degree to which a measure minimizes false negatives) and specificity (the degree to which a measure minimizes false positives) are commonly used for assessing predictive validity. Generally, the predictive validity of the parent- and teacher-completed SC-4 was moderate to high (i.e., sensitivity between .58 and .89; specificity between .57 and .94).

The treatment sensitivity of the SC-4 has been investigated in several double-blind placebo-controlled studies of stimulant medication (Gadow, Nolan, Sverd, Sprafkin, & Paolicelli, 1990; Gadow, Sverd, Sprafkin, Nolan, & Ezor, 1995). Differences in scores between doses indicate that the SC-4 is a good measure of response to stimulant medication. Furthermore, the instrument appears sensitive to several stimulant side effects.

The normative data for the SC-4 were recently expanded (Gadow & Sprafkin, 1999). According to the authors, T-scores between old and new samples are very similar; however, some differences may be noted between the manual (Gadow & Sprafkin, 1997) and the revised Score Sheets. Normative data are available on 4559 children and adolescents between the ages of 3 and 18. It should be noted that normative data for the SC-4 categories of ADHD and ODD were, with few exceptions, generated with other checklists developed by the same authors (e.g., Early Childhood Inventories and Child Symptom Inventories). Items are identical except for eight ADHD items that were shortened for the SC-4. With the exception of the preschool samples, the normative samples are smaller for the Peer Conflict scale. In general, data were gathered across a number of geographic regions; however, minorities were somewhat underrepresented for some age groups.

**Administration and Scoring**

Checklists, Score Summary Records, and Score Sheets for the SC-4 may be obtained with the manual as a kit, and purchased separately thereafter. Iden-
tical checklists may be used for parents and teachers, and both parent and teacher scores can be recorded on the same Score Summary Record. There are also separate Score Sheets for parent- and teacher-completed checklists, which present male and female scoring information on either side of the form.

The SC-4 should take no more than 10–15 minutes for informants to complete, which they do by recording raw category scores in the cells on the form. In using the symptom severity method of scoring, individual items are scored as follows: “Never” = 0, “Sometimes” = 1, “Often” = 2, and “Very often” = 3. Item scores from each category are then summed to obtain raw category scores. The Inattentive and Hyperactive/Impulsive scores are summed to obtain an ADHD Combined Type raw score. Separate Score Sheets are available that include tabulated T-scores. The SC-4 may also be scored using the Symptom Criterion method. Here, items are scored as follows: “Never” and “Sometimes” = 0, “Often” and “Very often” = 1. The DSM-IV specifies the number of symptoms required to meet criteria for various diagnoses, and this serves as the basis for meeting criteria for the disorders represented in the SC-4. The authors of the SC-4 provide the DSM-IV symptom count criteria on the Score Summary Sheet.

**Usability and Usefulness**

The SC-4 appears to be a useful measure of childhood externalizing behavior difficulties (ADHD, ODD, and Interpersonal Aggression). Given the high degree of comorbidity among childhood disruptive behavior disorders, an instrument like the SC-4 that is sensitive to both ADHD and more severe behavior problems is highly desirable. The Peer Conflict Scale and the Stimulant Side Effects Checklist appear to be useful indices of interpersonal aggression and stimulant side effects, respectively. Furthermore, data supporting the sensitivity of this instrument to treatment conditions suggest that the SC-4 is a useful instrument for monitoring child behavior. Finally, the inclusion of the Stimulant Side Effects Checklist makes the SC-4 a valuable tool in the titration of stimulant medication.

**The Conners’ Rating Scales—Revised (Short Form)**

**General Overview and Psychometric Characteristics**

The parent and teacher short forms of the Conners’ Rating Scales—Revised (Conners, 1997) were designed for repeated and/or brief assessment of symptoms relevant to ADHD and related disorders. These instruments are available in long or short versions for parent, teacher, and adolescent completion. The long versions will not be reviewed here because they represent broadband measures. It should be noted that only the long forms of the revised
Conners' scales contain the DSM-IV symptom subscales that may be preferred for diagnostic purposes. These subscales should not be confused with the ADHD Index that is included in the short form. The 12-item ADHD Index is not factor-derived; rather, it represents the best items for distinguishing ADHD from nonclinical children. Due to space limitations, we will restrict our review to the 27-item parent short form (CPRS-R:S) and the 28-item teacher short form (CTRS-R:S). The parent and teacher short forms contain the same subscales/indices (e.g., Oppositional, Cognitive Problems, Hyperactivity, and the ADHD Index), thus allowing easy comparisons across informants. Respondents are asked to rate behavior that has been problematic over the preceding month using a four-point Likert scale labeled with both levels of appropriateness (e.g., "Not true at all" = 0), and frequency (e.g., "Very frequent" = 3).

The test-retest reliability and internal consistency of the CPRS-R:S and CTRS-R:S have both been investigated (Conners, 1997). The internal consistencies of these instruments are good. Coefficient alphas range between .86 to .94 for the parent version and between .88 to .95 for the teacher version. The test-retest reliability (over a 6–8-week period) is somewhat variable across scales and informants, with coefficients of stability ranging between .62 to .85 for parents and between .72 and .92 for teachers. For example, the coefficient of stability for the parent-rated Oppositional subscale is low (.62), whereas the stability for the teacher ratings on the same subscale is good (.84). Conversely, the stability of the parent-rated Hyperactivity subscale is good (.85), whereas teacher-rated Hyperactivity is moderate (.72).

The Oppositional, Cognitive Problems, and Hyperactivity subscales of the CPRS-R:S and the CTRS-R:S were drawn from exploratory factor analyses of the long forms of the Conners' scales (see Conners, Sitarenios, Parker, & Epstein, 1998a, 1998b). Those items with the highest factor loadings were used in the construction of the shortened subscales. These items were then subjected to a set of confirmatory maximum likelihood analyses. Goodness of fit for both the parent and teacher versions was adequate as assessed across multiple indices (e.g., AGFI, GFI, RMS). Intercorrelations between subscales were highly similar across child gender.

Correlations between the long and short versions of the three factor-derived subscales approached 1.0. Hence, Conners (1997) performed concurrent and discriminant analyses on the long forms only. The concurrent and discriminant validity of the Conners' scales appear adequate. The Oppositional, Cognitive Problems, and Hyperactivity subscales and the ADHD Index all differed significantly in comparisons of nonclinic and ADHD groups. Correlations between the revised Conners' scales and the subscales of the CPRS-93 and the CTRS-39 indicate significant overlap across relevant constructs. For example, correlations between pertinent subscales such as CPRS-R Hyperactivity and CPRS-93 Hyperactive Immature range between .63 to .89 for parent-completed instruments, and between .71 and .88 for teacher-completed instruments.
Adequate normative data (Conners, 1997) are available for the CPRS-R:S (N = 2426) and CTRS-R:S (N = 1897). It should be noted that the majority of these data were derived from rescored long forms. Caucasians in these samples appear to be overrepresented (over 80%). Males received higher scores on the three subscales and the ADHD Index on both parent and teacher versions, and in general, younger children received higher ratings on Hyperactivity and lower ratings on Cognitive Problems. It appears, however, that the gap between boys and girls on the Hyperactivity subscale narrows as children age. Several ethnic differences were also found. Specifically, the Native-American group differed from the African-American, Hispanic, Asian, and Caucasian groups on the Oppositional subscale of the CPRS-R:S. Furthermore, teacher ratings of Hyperactivity for the African-American group were found to be significantly higher than those from the Asian and Caucasian groups, and the Asian group was found to be significantly lower than the Hispanic group. These ethnic differences should be considered in interpreting scores from these instruments due to the lack of separate normative data for various ethnic groups.

Administration and Scoring
The CPRS-R:S and the CTRS-R:S each contain the Oppositional, Cognitive Problems, and Hyperactivity subscales and the ADHD Index. These instruments should take informants between 5 and 10 minutes to complete, and just a few minutes are needed to score them using the QuickScore forms. The CPRS-R:S and CTRS-R:S are assembled in three layers. As informants endorse items on the protocol layer, their responses are also recorded on the QuickScore layer. Adjacent to the recorded responses on the QuickScore form is a scoring grid where item scores should be recorded as indicated by shading and column heading. Appropriate item scores are then summed, and total scores may be recorded at the bottom of the scoring grid. Tables of T-scores are available on the bottom layer of the QuickScore forms. T-scores of 65 or greater are usually indicative of a clinically significant problem. Simple descriptors of each scale are presented on the back of the QuickScore form.

Usability and Usefulness
The CPRS-R:S and CTRS-R:S are abbreviated versions of arguably the most commonly used rating scales for assessing child externalizing behavior. These scales offer a time-efficient alternative to the longer instruments while maintaining a substantial degree of coverage across symptoms of externalizing behavior. Clinicians can use these scales to assess the frequency that a child is reported (by parents and teachers) to exhibit symptoms of externalizing disorders relative to the normative population. This normative comparison is important because all children exhibit behaviors associated with externalizing disturbance on occasion. Therefore, it is critical to determine the
degree to which a child differs quantitatively from his or her peers. Given the choice between the long and short forms of the Conners’ scales, one should choose the short form when time-efficiency is important or when one expects to administer the measure repeatedly, as for treatment evaluation (Conners, 1997). Although the CPRS-R:S and the CTRS-R:S appear to be useful in the monitoring of change, no data are currently available in this regard.

Comparison of Measures for Child Externalizing Behavior

There are a number of factors to consider when comparing the three narrow-band rating scales for the assessment of externalizing behaviors. First, the degree to which each questionnaire represents the diagnostic criteria for disruptive behavior disorders as set forth in the DSM-IV is important to consider given the use of these criteria in clinical practice. Both the ADHD-IV and the SC-4 are explicitly tied to these criteria whereas the short form of the Conners’ Rating Scales is not. It should be noted, however, that the Conners’ ADHD index has been empirically derived to discriminate children with ADHD from normal control children.

A second factor to consider is the balance between brevity (to facilitate practical administration) and breadth of coverage. Specifically, the ADHD-IV can be completed in a very time-efficient fashion because it contains only 18 items. However, this measure does not include items related to ODD or CD, and therefore its use is limited to the assessment of ADHD symptomatology. The SC-4 and Conners’ Scales are longer but do allow evaluation of ODD and CD symptoms. Thus, when choosing among these questionnaires one must consider the nature of the referral and the degree to which broad assessment of symptoms is necessary. Moreover, when one is selecting questionnaires to assess treatment response on an ongoing basis, it is important to use a measure that is quick to complete while tapping into the most critical dimensions of potential behavior change.

Another important factor is the degree to which the questionnaire has been normed on the population of interest. For example, if one is assessing preschool-aged children, then the SC-4 or Conners’ scales should be used because these have normative data for children aged 3 years and up. Alternatively, when assessing ADHD in school-aged children, the ADHD-IV may be the best choice because its normative data are most representative of the U.S. population.

Finally, it should be noted that all three of these questionnaires are convenient and easy to administer and score. Furthermore, they all have sound psychometric properties when used for their intended purposes. Given that all three measures are very practical for clinicians to use, decisions about which measure is appropriate for a given situation should be based on the factors discussed above.
Case Study: Child Exhibiting Externalizing Behavior Difficulties

Tom is an 8-year-old third grader in a general education classroom in a public elementary school. According to his parents, Tom experienced normal cognitive and physical development until he reached kindergarten. At that time, Tom's kindergarten teacher reported that he had difficulty playing quietly, had poor attention, and had a difficult time obeying classroom rules. These problems were also reported to be problematic at home. Tom's parents reported that he often seemed to ignore their requests, was difficult to control, and was easily frustrated. His current teacher reported that Tom had been inattentive and often had a difficult time staying in his seat when required to do so. The teacher also reported that he was almost always doing things other than what he was supposed to be doing, and this was disruptive to classroom activities. He frequently appeared to have a reason to avoid work (e.g., sharpening a pencil or throwing away a scrap of paper), but appeared able to do just enough work to "get by." Currently, Tom is performing at grade level in all academic areas, but he seldom completes homework assignments, which significantly impacts his grades. Also, he is beginning to experience some difficulty in mathematics. Tom's teacher is concerned with his disruption of the classroom and its impact on other students. Furthermore, he is concerned that Tom will have more difficulty in later grades if these problems persist.

After briefly discussing the case with Tom's teacher, the school psychologist asked him to complete the Teacher Report Form (TRF; Achenbach, 1991b) as a screening instrument. On the TRF, elevated scores were obtained on both the Attention Problems and Aggressive Behaviors subscales. T scores on these scales were greater than 70 (above the 98th percentile). All other scores were within the normal range (below the 93rd percentile). Based on these scores and the teacher concerns, a multimethod assessment of ADHD appeared justified.

The next step in the assessment procedure was to interview Tom's teacher. In the interview, the teacher reported that Tom often experienced problems with attention, overactivity, and impulsivity in most school settings. Problems with oppositional or aggressive behavior were not reported to occur. Tom's teacher was also asked to complete the ADHD-IV. Tom's ratings were above the 93rd percentile on the total score as well as on the Inattention and Hyperactivity-Impulsivity scores. Additionally, 14 of the 18 DSM-IV symptoms of ADHD were endorsed as occurring at least "often." Observations conducted in Tom's classroom were consistent with the reports from Tom's teacher. Tom was significantly off-task compared to randomly observed classroom peers.

When Tom's mother was interviewed, she corroborated the teacher's reports of problems with inattention, overactivity, and impulsive behavior in
the absence of significant internalizing problems, and reported that most of the symptoms of ADHD had occurred frequently since kindergarten. Tom's mother was then asked to complete the Child Behavior Checklist (CBCL; Achenbach, 1991a). Her ratings generated a significantly elevated score on the Attention Problems subscale (i.e., greater than the 95th percentile). All other scores were in the normal range and below the 93rd percentile. Scores obtained on the home version of the ADHD-IV were consistent with the teacher ratings, suggesting that Tom was exhibiting significant symptoms of ADHD across school and home settings.

Based upon these assessment results, Tom appeared to be exhibiting significant symptoms of ADHD at home and school. A treatment plan was designed by the school psychologist to include behavior modification programming across settings and a referral to Tom's physician in order to explore the utility of stimulant medication. After discussing the medication option with the physician, Tom's parents opted for a trial of stimulant medication. They asked that the school psychologist be involved in the titration process so that the effects of the medication could be properly evaluated.

Tom's response to three doses (5, 10, and 15 mg) of methylphenidate (MPH) and placebo was evaluated using multiple assessment measures, including ratings on the ADHD-IV and Stimulant Side Effects Checklist of the SC-4 as well as direct observations of Tom's classroom behavior and academic performance. Each dose condition lasted for one week. Tom and his parents, teacher, and school psychologist were blind to what dose he was taking in a given week.

Analysis of the resultant data using the Reliability Change Index for ADHD-IV ratings indicated that Tom's behavior improved from placebo at each of the three doses of MPH. However, two pieces of data led the school psychologist to recommend the 5 mg dose. First, both the observational and ADHD-IV data indicated that there was not a significant change in behavior from the 5mg to 10mg dose of MPH. Second, maternal ratings from the SC-4 indicated that there was an increase in stimulant side effects from the 5mg to 10mg dose. After the titration procedure, Tom's progress was monitored by use of observation and rating scale data each month.

MEASURES TO ASsess ANXIOUS AND DEPRESSIVE SYMPTOMS

Children and adolescents can exhibit significant symptoms of a number of internalizing disorders, such as Separation Anxiety Disorder, Generalized Anxiety Disorder, Major Depression, and Dysthymic Disorder (American Psychiatric Association, 1994). Although the specific nature of symptomatology may differ between children and adults (e.g., children who are depressed may
exhibit a chronic irritable mood rather than a depressed mood), these disorders lead to impairment in functioning as they do in adults. Specifically, children with anxiety disorders and/or depression can experience significant difficulties with academic performance and social functioning (Callahan, Panichelli-Mindel, & Kendall, 1996). Due to the overlap of symptomatology between anxiety and depression as well as the possibility that children with externalizing disorders could also experience internalizing disturbance, it is important for clinicians to assess internalizing symptoms in many cases. In particular, self-report ratings are critical because many of the symptoms of these disorders can only be detected by the children themselves (Kazdin, 1990).

The measures reviewed in this section can be used for a number of purposes, including screening for further assessment, being part of a multimodal evaluation battery, and assessing treatment outcome (see Tables 12.4 and 12.5). A case study illustrates the use of these scales as part of a multimodal assessment of internalizing behavior.

**TABLE 12.4**

**Summary Information for the Revised Children's Manifest Anxiety Scale**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Cecil R. Reynolds and Bert O. Richmond (1985)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher</td>
<td>Western Psychological Services Los Angeles, CA</td>
</tr>
<tr>
<td>Copyrighted</td>
<td>Yes</td>
</tr>
<tr>
<td>Time to Complete</td>
<td>&lt; 15 minutes</td>
</tr>
<tr>
<td>Scoring Software</td>
<td>No</td>
</tr>
<tr>
<td>Ages</td>
<td>6–19</td>
</tr>
<tr>
<td>Items</td>
<td>37</td>
</tr>
<tr>
<td>Scaling</td>
<td>Yes-No</td>
</tr>
</tbody>
</table>

**Available Reliability Information**

| Test-Retest        | Yes                                           |
| Interrater         | No                                            |
| Internal Consistency | Yes                                        |
| Normative Data     | N = 4,972                                      |

**Available Validity Information**

<p>| Construct         | Yes                                           |
| Discriminant      | Yes                                           |
| Concurrent        | Yes                                           |
| Predictive        | No                                            |
| Treatment-Sensitive | No                                        |</p>
<table>
<thead>
<tr>
<th>Author</th>
<th>Publisher</th>
<th>Publisher</th>
<th>Time to Complete</th>
<th>Scoring Software</th>
<th>Ages</th>
<th>Items</th>
<th>Subscales</th>
<th>Author</th>
<th>Publisher</th>
<th>Publisher</th>
<th>Time to Complete</th>
<th>Scoring Software</th>
<th>Ages</th>
<th>Items</th>
</tr>
</thead>
</table>

| Test-Retest            | Yes                                   | Yes                                   | Yes               | Yes               | Yes      | Yes   | Yes                                           | Yes                                   | Yes                                   | Yes               | Yes               | Yes      | Yes   |
| Interrater             | No                                    | No                                    | No                | No                | No       | No    | No                                            | No                                    | No                                    | No                | No                | No       | No    |
| Internal Consistency   | Yes                                   | Yes                                   | Yes               | Yes               | Yes      | Yes   | Yes                                           | Yes                                   | Yes                                   | Yes               | Yes               | Yes      | Yes   |
| Normative Data         | N = 1,266                             | N = 2,460                             | N = 1,620         | N = 1,620         | N = 1,620 | N = 1,620 | N = 1,620                                     | N = 1,620                             | N = 1,620                             | N = 1,620         | N = 1,620         | N = 1,620 | N = 1,620 |

**The Revised Children's Manifest Anxiety Scale**

**General Overview and Psychometric Characteristics**

The Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1985) is a revision of the Children's Manifest Anxiety Scale (Casteneda, McCandless, & Palermo, 1956), which was derived from an adult rating scale, the Taylor Manifest Anxiety Scale (Taylor, 1951). The RCMAS is a 37-item self-
report questionnaire designed to obtain information from children concerning their feelings of anxiety across several dimensions. Items on the RCMAS are worded as short statements (e.g., "I am afraid of a lot of things") and children are asked to circle "Yes" if the statement is true of them and "No" if the item is not true of them. The checklist is headed by the nonthreatening title, "What I Think and Feel."

The test-retest reliability and internal consistency of the RCMAS appear adequate. Specifically, a Kuder-Richardson 20 alpha (an index of internal consistency for dichotomously scored items) of .85 has been obtained for the total scale (Reynolds & Paget, 1983). Coefficients of stability have ranged from .68 for elementary school children tested 9 months apart to .90 for administrations 3 weeks apart (Reynolds & Paget, 1983).

In their exploratory factor analytic study, Reynolds and Paget (1981) found a five-factor solution to be the most lucid reduction of the RCMAS items. The five factors include the 11-item Worry/Oversensitivity factor, which accounted for the most variance (42%); Physiological Anxiety (10 items); Social Concerns/Concentration (7 items); Lie Scale 1 (6 items); and Lie Scale 2 (3 items). The two lie scales can be useful in determining the degree to which children respond to items in a socially desirable manner (e.g., "I tell the truth every single time").

Although the RCMAS has been found to discriminate between groups of children with anxiety disorders and normal control groups, its ability to discriminate children with anxiety disorders from other psychiatric groups has been mixed (cf., Mattison, Bagnato, & Brubaker, 1988; Perrin & Last, 1992). The RCMAS has been found to correlate highly with the Trait anxiety scale of the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973), yet correlations between the RCMAS and the State scale of the same STAIC were found to be nonsignificant. Finally, little information exists on the utility of the RCMAS in the monitoring of treatment regimens.

The standardization sample for the RCMAS (N = 4972) consisted of students between 6 and 19 years old who attended 80 school districts in 13 states across all major geographic regions in the United States (Reynolds & Paget, 1983). Although efforts were made to obtain ethnic diversity in the sample, children of color appear to be underrepresented (11.8% of the total sample). Approximately 600 children in the standardization sample were in classes for children who were educably mentally retarded, learning disabled, or intellectually gifted. RCMAS scores have been found to differ by age, race, and gender. Hence, normative data are organized by these variables.

**Administration and Scoring**

The RCMAS was designed to be administered to children and adolescents between 6 and 19 years old. As the readability of the RCMAS is estimated by its authors to be approximately third-grade level, items should be read to
young children and children who are not fluent at the third-grade level. The instrument should take no more than 15 minutes to complete. If a child is observed marking more than one response on an item (e.g., “Yes” and “No”), which often happens when children are not sure how to respond, the child should be instructed to select the choice that best describes him or her.

Scoring the RCMAS is relatively straightforward. A transparent scoring key can be placed over the child’s responses to derive 5 scores: Physiological Anxiety, Worry/Oversensitivity, Social Concerns/Concentration, Lie Scale (Lie 1 and Lie 2 combined), and Total Anxiety. Each subscale has a partially opaque column that only displays the item responses for that subscale when placed over the response area of the protocol. Items can then be summed (“Yes” = 1, “No” = 0) to obtain subscale scores.

Usability and Usefulness

The RCMAS is arguably the best known measure of general childhood distress. It is easily administered and scored, and it possesses adequate psychometric properties. However, due to mixed results in the measure’s ability to discriminate between psychiatric groups, it appears to be most useful as a screening instrument. A structured psychiatric interview may be preferred for diagnostic purposes. The Lie scales of the RCMAS make it a useful tool for practitioners who wish to measure a child’s tendency to respond in a socially desirable manner. Finally, normative data for the RCMAS was collected over 20 years ago and might be considered outdated. Furthermore, the normative sample is composed almost entirely of Caucasian children. Thus, clinicians should be cautious when using this instrument with children from minority backgrounds.

The Children’s Depression Inventory

General Overview and Psychometric Properties

The Children’s Depression Inventory (CDI; Kovacs, 1981, 1992) is a downward extension of the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) for children and adolescents between 7 and 17 years old. The CDI is available in both a 27-item version, which was designed to measure cognitive, affective, and behavioral indicators of depression, and a 10-item version (CDI-S; Kovacs, 1992), which may be preferred for a quick screening instrument. The administration of the CDI and CDI-S involves the presentation of three alternative statements (e.g., “I am sad once in a while,” “I am sad many times,” and “I am sad all the time”). Children are asked to select the statement that best describes them within the past two weeks.

The psychometric properties of the CDI have been well studied (e.g., Finch,
Saylor, & Edwards, 1985; Kazdin, French, Unis, Esveldt-Dawson, & Sherick, 1983; Kovacs, 1981). The internal consistency and item-total score correlations appear adequate (coefficient alphas ranged between .71 to .89). Several studies have found adequate coefficients of stability for the CDI. For example, Finch et al. (1987) obtained adequate test-retest reliability coefficients for 2-week (.82), 4-week (.66), and 6-week (.67) latencies in a normative sample. However, for shorter latencies (e.g., 1 week), low stability has been found (.38) in a group of normal school children, compared to high stability (.87) for a group of psychiatric inpatients (Saylor, Finch, Baskin, Furey, & Kelly, 1984).

Several studies have investigated the factor structure of the 27-item CDI (e.g., Helsel & Matson, 1984; Saylor, Finch, Spirito, & Bennett, 1984; Weiss & Weisz, 1988; Weiss et al., 1991). At least partially due to disparate factor analytic strategies, investigators have found different numbers of factors for the CDI items (between three and eight). Based upon a set of maximum likelihood factor analyses on a normative sample of 1266 school-aged children, prior factor analytic studies, and the interpretability of factors, Kovacs (1992) chose a five-factor solution: Negative Mood, Interpersonal Problems, Inefficacy, Anhedonia, and Negative Self-Esteem. All of these factors have been found to be significantly intercorrelated (Kovacs, 1992).

The concurrent validity of the CDI has been studied by several investigators (e.g., Kazdin et al., 1983). The CDI has been found to correlate with other measures of depression (e.g., Weissman, Orvaschel, & Padain, 1980) and measures of other related constructs such as self-concept and hopelessness (e.g., Kazdin, French, Unis, & Esveldt-Dawson, 1983; Saylor, Finch, Baskin, Furey, & Kelly, 1984). Information concerning the discriminant validity of the CDI is mixed. Although the CDI has demonstrated the ability to discriminate depressed children from non-depressed children as identified by child interviews (Hodges, 1990), its ability to discriminate between psychiatrically diagnosed groups has been inadequate. For example, Wendel, Nelson, Politano, Mayhall, and Finch (1988), investigating the ability of the CDI to discriminate between inpatient clinic samples and a large sample of normal children, found the CDI to correctly classify 83% of all subjects. However, 100% of the inpatient children were incorrectly classified as normal. The treatment sensitivity of the CDI has been demonstrated in a study of group therapy for children of divorced parents (Garvin, Leber, & Kalter, 1991).

The standardization sample of the CDI and CDI-S consisted of 1266 children and adolescents between 7 and 16 years old from public schools in Florida (Kovacs, 1992). Although demographic data were not collected, it has been estimated that 77% of the sample was Caucasian and a relatively large percentage of the sample was middle class. No formal comparisons were made to U.S. Census data. Based upon the homogeneity of this normative sample (e.g., geographic region, race, and class), one should exercise caution in making generalizations from these data to underrepresented populations.
Administration and Scoring

The CDI and CDI-S were designed to be administered to children and adolescents between 7 and 17 years old, although normative data are limited to boys between 7 and 15 years old and girls between 7 and 16 years old. The 27-item CDI should take no more than 15 minutes for children to complete and the 10-item CDI-S no more than 5–10 minutes. Although the specific readability of the CDI has not been assessed, it may be advisable to read the items to younger children or those with reading difficulties while allowing them to follow along on their own form. One should ensure that the child is aware that he or she should mark only a single box out of the three alternatives. Occasionally, children express concern that some items do not pertain to them. In this event, one should encourage the child to mark the choice that best fits him or her. Like the revised Conners’, the CDI and CDI-S use QuickScore forms, which automatically transfer child responses to the scoring layer of the protocol. The CDI/CDI-S items are scored on a 3-point scale. Next to each box on the scoring layer is a number from 0 to 2. Each item also is labeled with a letter indicating which factor it is associated with (e.g., A = Negative Mood, B = Interpersonal Problems). The scores for each item contributing to the factor in question (e.g., all the items labeled A) are summed to obtain the factor raw score. T-scores 65 or greater are considered clinically significant.

Usability and Usefulness

The CDI is the most widely used self-report rating scale to assess childhood depressive symptoms (Kazdin, 1990). Items are reflective of affective, cognitive, and behavioral symptoms associated with depression, and the CDI has demonstrated reasonable psychometric properties. The CDI is relatively easy to administer and score, and the CDI-S may be used when an especially brief measure is desired. Finally, the CDI may be a useful tool in assessing treatment efficacy, although this has not been extensively studied.

The Reynolds Child and Adolescent Depression Scales

General Overview

The Reynolds Adolescent Depression Scale (RADS; Reynolds, 1987) is a 30-item self-report questionnaire developed to measure the severity of depressive symptoms in adolescents between 13 and 18 years old. The protocol itself is labeled in a nonthreatening manner (e.g., “About Myself”). Items are presented as short statements (e.g., “I feel lonely”), and adolescents are asked to respond on a 4-point Likert scale based on frequency (“Almost never,” “Hardly ever,” “Sometimes,” and “Most of the time”). The Reynolds Child Depression Scale (RCDS; Reynolds, 1989) is a similar 30-item self-
report questionnaire designed for children between 8 and 12 years old. The first 29 items are presented as short statements; most of which are identical to the statements in the RADS. Children are asked to respond on a similar scale to the RADS ("Almost never," "Sometimes," "A lot of the time," and "All the time"). The final item presents children with 5 line-drawn faces that appear in the range from very happy to very sad; children are asked to "Fill in the circle over the face that shows how [they] feel."

Psychometric Properties of the RADS

The RADS has demonstrated good internal consistency, including alphas of .90 (Reynolds, 1987). The stability of the RADS was measured at 6-week, 3-month, and 1-year intervals in heterogeneous groups of depressed and nondepressed high-school students. Coefficients of stability ranged between .80 and .79 for 6-week and 3-month intervals, respectively, and .63 for a 1-year latency (Reynolds, 1987).

A series of factor analytic studies (Reynolds, 1987) indicate a five-factor solution for the RADS, with the majority of items loading on the first factor (Generalized Demoralization), which accounted for approximately 32% of the overall variance. The other four factors accounting for an additional 28% of the overall variance are Despondency and Worry, Somatic-Vegetative, Anhedonia, and Self-Worth. However, norms have not been developed for these individual factors.

The validity of the RADS has been supported by several methods. Consistent with the factor analytic data, item-total correlations in a sample of 2296 adolescents suggest good content validity (median $r = .53$). The concurrent validity of the RADS appears adequate. Specifically, the RADS has been compared to several other self-report depression rating scales including the Beck Depression Inventory (Beck et al., 1961), the Center for Epidemiological Studies Depression Scale (Radloff, 1977), the Self-Rating Depression Scale (Zung, 1965), and the Children's Depression Inventory (Kovacs, 1992). The RADS has been shown to discriminate between groups of adolescents with and without depression (Evert & Reynolds, 1986). Finally, a wait-list-controlled comparative treatment study of adolescents with severe depression has supported the utility of the RADS as a treatment-monitoring instrument: While the treatment groups evidenced declining RADS scores, the wait-list control group's RADS scores were relatively stable (Reynolds & Coats, 1986).

Normative data are available for 2460 7th- through 12th-graders from one senior- and two junior-high schools in an urban/suburban community in the midwestern United States. Despite the apparent overrepresentation of Caucasians (76%) and the restricted geographic sampling, Reynolds (1987) reports that the sample is well stratified in terms of age, sex, race, and socioeconomic status.
Psychometric Properties of the RCDS

The RCDS has demonstrated good internal consistency and stability (Reynolds, 1989). Specifically, coefficient alphas computed for individual grades and for the entire standardization sample were consistently high (alphas between .85 to .90). The RCDS has demonstrated good 4-week stability ($r = .85$) in a sample of 220 third- through fifth-grade students (Reynolds & Graves, 1989).

Results of a factor analytic study conducted by Reynolds (1989) indicate a five-factor solution to the RCDS items. The first factor, Despondency-Worry, accounted for 27.5% of the total variance. The four remaining factors (Generalized Demoralization-Despondency, Somatic-Vegetative, Dysphoric Mood, and Anhedonia) account for an additional 19.6% of the total variance. The results of this analysis suggest that the RCDS, like the RADS, measures several dimensions of depression.

Supporting the content validity of the RCDS, the vast majority of items (24 out of 30) were found to significantly covary as indicated by a set of item-total scale correlations (Reynolds, 1989). The concurrent validity of the RCDS has been investigated utilizing several other self-report measures of depressive symptomatology (Reynolds, 1989); for example, the RCDS was found to correlate significantly with the Children's Depression Rating Scale—Revised (Poznanski, Freeman, & Mokros, 1985) as well as with clinical interviews. The convergent validity of the RCDS has been found to correlate significantly with measures of constructs related to depression such as self-esteem and anxiety (Reynolds, Anderson, & Bartell, 1985). A study conducted by Stark, Reynolds, and Kaslow (1987) demonstrated the utility of the RCDS as a treatment outcome measure: While the two treatment groups in this study demonstrated significant declines in RCDS scores, the control group's scores were relatively stable.

Normative data for the RCDS are based upon a sample of 1620 second-through seventh-grade students from the midwestern and western United States. Grade-level normative data are only available for grades 3 through 6 due to a limited sample size in grades 2 and 7. According to Reynolds (1989), the standardization sample was heterogeneous in terms of age, sex, gender, and socioeconomic status. Approximately 71% of the sample children were Caucasian.

Administration and Scoring of the RADS and RCDS

The RADS and RCDS are available in both hand-scored and computer-scored versions, and neither instrument should take informants more than 5 to 10 minutes to complete. As with some of the instruments described above, it may be preferable to read items to young children while they follow along on their protocol. Reynolds (1987, 1989) suggested that children and adoles-
cents be informed that there are no right and wrong answers, and that the purpose of the instrument is to assess the individual's general feelings about himself or herself. Responses should be made with a pencil so that children can change them. Both the RADS and RCDS generate a single total score. Transparent scoring keys are available for hand-scoring the RADS and RCDS. It is essential that the scoring keys be used because several items on both instruments are reverse-scored. Protocols with less than 24 items endorsed should not be considered valid. For protocols that have between 24 and 29 items endorsed, multiply the total score by 30 and divide by the number of items completed. Other validity checks are available in the manuals of these measures and should be referred to when interpreting scores. Percentile scores based on grade and gender are available in the appendices of the RADS and RCDS manuals. It is suggested that a raw score greater than 77 on the RADS or greater than 74 on the RCDS is clinically significant; however, certain critical items (e.g., those indicating self-destructive behavior) should also be examined in children and adolescents who do not meet the clinical cutoff (Reynolds, 1987, 1989).

Usability and Usefulness of the RADS and RCDS

The RADS and RCDS appear to be psychometrically sound measures of child and adolescent depressive symptomatology, and both are easily administered and scored. They seem well suited for use as screening tools, as part of a comprehensive diagnostic assessment, and for the monitoring of intervention effects.

Comparison of Measures for Depressive Symptoms

All three of the depression measures reviewed in this chapter have adequate psychometric properties, are of approximately the same length, and are easy to administer and score. The CDI has been in use for a longer period of time and consequently has been subjected to greater empirical scrutiny. Furthermore, the response format of the CDI may be easier for children to understand because it is more concrete than the anchor points used in the Likert scale on the RADS and RCDS. However, the total sum scoring of the RADS and RCDS may be more appropriate for assessing depressive symptoms given that the construct of depression in children appears to be a unitary (i.e., one-factor) phenomenon. When one examines the factor analytic results for the CDI, it appears that it is composed of one strong factor (accounting for 23% of the variance) and possibly a second factor (Negative Mood). Thus, the clinical utility of CDI factors beyond the total score and Negative Mood score may be limited. Finally, it should be noted that minority representation in the normative data of all three measures is very limited. Clinicians should use caution when using these depression measures to assess children from
Robert J. Volpe and George J. DuPaul

non-Caucasian backgrounds. Both the RADS and the RCDS have demonstrated adequate treatment sensitivity. Hence, by administering these scales before the implementation of treatment and by intervals throughout the intervention, these scales may be a source of useful information in monitoring treatment effects.

**Case Study: Child Exhibiting Internalizing Behavior Difficulties**

Michelle is a 10-year-old fifth-grader who was referred by her teacher to the school psychologist for academic difficulty and inattentive behavior. A review of Michelle’s academic records indicated that she was an above-average student, receiving As and Bs since school entry.

An interview with Michelle’s teacher revealed that Michelle’s grades were average in September, but had been dropping steadily for several months. It seemed that Michelle was having great difficulty staying on task and would frequently exhibit escape behaviors (e.g., asking to go to the lavatory or for a drink of water). Over the past several weeks, Michelle began to complain of aches and pains during class tests, and her teacher was becoming concerned that Michelle appeared withdrawn and sad.

The scores on the TRF generated by Michelle’s teacher showed clinically significant elevations (T scores above 70) on both the Withdrawn and Anxious/Depressed subscales.

Michelle’s parents noted similar concerns during a phone interview. They reported that Michelle had become increasingly withdrawn and seemed to change activities frequently. She often appeared to be sad and did not seem to enjoy her usual play activities. Furthermore, she had been having difficulty sleeping through the night. The parental interview revealed that Michelle’s parents had been experiencing marital discord since the previous summer, and Michelle’s father had recently moved out of the house.

Michelle was asked to complete both the RCDS and the RCMAS. On the RCDS, Michelle responded to all 30 items, receiving a score of 89 (98th percentile). Some of the items that she endorsed as happening “a lot of the time” were “I worry about school,” “I feel like crying,” and “I feel my parents don’t like me.” Michelle’s response of “sometimes” to the item “I feel like hurting myself” was considered an important issue to raise during a subsequent interview with her. On the RCMAS, Michelle received a scaled score of 13 (88th percentile) on the Physiological subscale, a 14 (92nd percentile) on both the Worry/Oversensitivity and Social Concerns subscales, and a Total Anxiety T score of 66. Examples of some of the items endorsed by Michelle that were addressed in the subsequent interview with her were “I feel alone even though people are with me,” “I worry about what is going to happen,” and “It is hard for me to keep my mind on my schoolwork.” Michelle was interviewed by the school psychologist and reported that she was getting a lot of homework this year and that she was feeling overwhelmed. Michelle
described the situation at home and said that she was worried about what would happen with her parents. She also reported feeling sad quite frequently. The issue of suicidal ideation was discussed and Michelle said she had thought about what would happen if she were to hurt herself, but these were passing thoughts with no clear plans for following through.

A curriculum-based assessment revealed that Michelle was at or close to grade level in the areas of reading, mathematics, and written expression. Thus, taken together, the assessment data suggested that Michelle was experiencing clinically significant depressive symptoms and some noteworthy anxiety disorder symptoms, all of which appeared to be the result of her home situation. These symptoms appeared to be affecting her ability to concentrate and socialize in school, which had a negative impact on her academic performance. In turn, Michelle’s difficulty in keeping up with task demands in school exacerbated her symptoms of anxiety.

A treatment plan was designed to reduce task demands so that Michelle would not be overwhelmed by her class assignments and homework. It was further recommended that Michelle’s parents help her to understand the marital difficulties they were experiencing and to consider family counseling.

### Evaluating Children from Diverse Backgrounds

The population of children in North America is becoming increasingly diverse in terms of ethnicity and language. Thus, clinicians are now faced with the challenge of evaluating children from diverse backgrounds, and unfortunately the normative samples of many behavior rating scales do not include adequate numbers of children from non-Caucasian, non-English-speaking backgrounds. The measures reviewed in this chapter vary with respect to the degree to which their normative data are representative of the U.S. population. Some recently devised scales, such as the Conners’ and the ADHD-IV, include norms that are based on nationally representative samples. Older questionnaires, like the RCMAS, have been standardized on virtually all-Caucasian samples. Thus, clinicians must be aware of the degree to which normative data are applicable when evaluating a specific child. Furthermore, even when normative data are representative, there may be systematic differences in scores between ethnic groups (e.g., higher ratings for African-Americans than Caucasians on the ADHD-IV). In such cases, clinicians should be cautious in using normative cut points. For example, a higher threshold (e.g., 98th percentile) might be used to determine the clinical significance of ADHD-IV scores for an African-American child.

### Integrating Data from Multiple Informants

It is generally accepted that a thorough assessment of child psychopathology requires reports from one or more adults (e.g., mother and/or teacher) in addition to the report of the child. Efforts of combining information from
multiple informants to increase diagnostic accuracy have been complicated by low correlations between types of informants (e.g., Edelbrock, Costello, Dulcan, Conover, & Kalas, 1986). The discrepancies in scores can be attributed to several factors (e.g., variance across situations and differential perceptions of behavior). How data are combined to arrive at reliable diagnostic decisions depends largely on the type of assessment one is conducting. For example, by summing responses across informants one improves sensitivity (e.g., increasing true positives), but almost always at the cost of specificity (e.g., increasing false positives). When the mission of assessment is to screen for behavior or emotional problems, this trade-off may be acceptable. Alternatively, when using rating scales to derive a clinical diagnosis, false positives are far less acceptable. In such cases, one may choose to take an optimal informant approach (e.g., Loeber, Green, Lahey, & Stouthamer-Loeber, 1989). This methodology involves selecting the most reliable informant based on age and the dimension of behavior being assessed. For example, when assessing a child in elementary school, teacher ratings may be more useful than parent ratings because teachers have more experience with school-aged children and presumably are in a better position to make normative comparisons. Unfortunately, the optimal informant approach does not fully take into consideration differences in behavior across settings. For example, a child may only exhibit behavior problems when he or she is at home. Hence, teacher reports would not be sensitive to the problems experienced by the child’s family. Offord et al. (1996) suggest that child emotional and behavioral problems may best be viewed as “informant specific,” and that combining data from different informants may obscure the true pattern of behavior. Thus, practitioners must decide how to handle informant discrepancies based on the context and purpose of the evaluation.

Evaluating Treatment Outcome

Several of the measures (SC-4, CDI, RADS, and RCDS) reviewed here have been studied for their sensitivity to treatment regimens. Due to their relative brevity, all of the measures reviewed here may be considered for assessing changes in child behavior. Supplied with such measures, practitioners are left with the problem of determining if a measured change is clinically significant. Fortunately, the Reliability Change Index (RCI; Jacobsen & Truax, 1991) is well suited to fulfill this need. The RCI is simply the difference between a child’s pretreatment and posttreatment score, divided by the standard error of difference between the two test scores. When this value is greater than 1.96, it is unlikely that the change between administrations is due to chance ($p < .05$). Although only the ADHD-IV manual contains tables helpful in deriving RCI scores, clinicians can determine this for any of the rating scales reviewed here by using the test-retest reliability coefficient to calculate the standard error of difference (see Jacobsen & Truax, 1991, for details).
SUMMARY AND CONCLUSIONS

A number of narrow-band rating scales are available to assess the internalizing and externalizing behavior difficulties exhibited by children. All of the questionnaires reviewed in this chapter possess adequate psychometric properties and are clinically useful. In particular, these instruments are quite valuable when screening children for further assessment. Specifically, narrow-band rating scales provide clinicians with important information regarding the degree to which children exhibit behavior difficulties at a significantly higher rate or of greater severity than do other children of the same age and gender. This information is helpful because, for many of the disorders assessed by these rating scales (e.g., ADHD), one must establish that behavioral symptoms are quantitatively different from the normative population. Although these instruments also are useful for diagnostic decision making and treatment evaluation, it should be noted that rating scales should never be used in isolation for these purposes; instead, assessment methods that incorporate information from multiple sources and settings should be used to make important clinical decisions. The rating scales reviewed in this chapter can be useful in this context when supplemented with other data to give a more complete picture of a child's functioning.

The empirical underpinnings of child behavior rating scales are relatively strong and support the use of these instruments in most clinical circumstances. Nevertheless, several factors should be taken into account when interpreting scores from behavior questionnaires. The relevance of available normative data to the child population being assessed is extremely important. Specifically, some measures (e.g., RCMAS) may have normative data that are relatively outdated and that do not include adequate numbers of children from minority groups. In general, one must be cautious in using clinical cutoff scores, especially when assessing children from diverse ethnic or linguistic backgrounds. There simply are not enough data to support the strict use of percentile and cut-point scores with minority children, especially when one is concerned about overidentification of children from diverse backgrounds as disordered. When evaluating students from minority backgrounds, clinicians should use more conservative cut points (e.g., 98th percentile) for diagnostic decision making. Another possible solution is to collect data from a local sample that is representative of the population being served. This may not be practical in some circumstances but can serve as a stopgap measure until more representative normative data are available for these instruments.

Another factor to consider when interpreting the results of behavior rating scales is the nature of responses to individual items. Although one would typically place most credence in total or subscale scores, in some instances individual item scores should be examined for clinically relevant information. This is particularly true for measures of internalizing symptoms like certain items on the CDI (e.g., "I want to kill myself") and RCDS (e.g., "I feel like..."
hurting myself”). Individual item responses may indicate problem areas that are not immediately evident when relying solely on summary scores.

Narrow-band rating scales are very useful in the assessment of disruptive and internalizing behaviors exhibited by children and adolescents. This chapter has reviewed some of the more commonly used measures employed by clinicians for screening, diagnostic, and treatment-evaluation purposes. Such scales have become more popular for clinical decision making and fortunately their use is supported by a growing body of empirical investigations. When included as part of a multimethod assessment battery, narrow-band rating scales can provide information that is both time-efficient and clinically relevant.

References


12. Assessment with Brief Behavior Rating Scales


The literature on the assessment of culturally diverse individuals is saturated with statistics and projections regarding the rapid growth of ethnic groups in the United States and Canada. Often in this context there is commentary on the dearth of providers who share cultural backgrounds and language with the individuals who are referred to them, and on the barriers to providing assessment and intervention services. Extreme positions are sometimes taken, such as maintaining that it is unethical to provide services when one does not share the language or culture of the referred individual, a position that ignores the fact that there is a shortage of bilingual and bicultural providers. Unfortunately, few alternatives are offered to practitioners, with the result that clinicians often default to an inadequate solution of refusing services rather than face the challenge of cross-cultural assessment. In this chapter, we hope to encourage practitioners to view the challenge of cross-cultural assessment as surmountable with the proper tools. We present for discussion a number of variables that occur in the context of culture, but
ultimately suggest that empirical knowledge and knowledge of each individual's unique background and experiences will facilitate the assessment process and outcome.

The chapter begins with an overview of the concept of cultural values and the role of acculturation in determining "readiness" for assessment. Education and literacy, and the influence of these factors on cognitive performance, are discussed. Language and language proficiency are addressed in the context of the language used in the assessment, test selection, test adaptation, and the use of translators. Examples of specific language characteristics on cognitive performance are also provided. Finally, two case studies are presented that illustrate the ways in which the process of assessment can be conceptualized and undertaken, with the ultimate goal of improving the ecological validity of assessment results.

**CULTURAL AND COGNITIVE VALUES**

Cultural values and expectations have a significant influence on both the process and the outcome of cognitive and neuropsychological assessment. Culture provides specific models for ways of behaving, feeling, thinking, and communicating. In general, culture dictates what is, and what is not, situationally relevant. According to Berry (1988), cognitive values are "the set of cognitive goals which are collectively shared and toward which children are socialised in a particular society. It is essential to understand these goals, since one cannot assess how far a person has gotten unless one understands where he is going" (p. 12). As Berry has maintained, in order to understand the goals for cognitive competence in children, it is essential to understand the skills and abilities that are valued in that society. This is critically important for the accurate assessment and understanding of both normal and abnormal cognitive functioning and for the accurate interpretation of cognitive performance.

To understand the limitations of interpreting and predicting cognitive performance based on knowledge of one's culture, it is necessary to appreciate that within a given culture there are, in reality, multiple cultures and societies. For this reason, intracultural variability in cognitive performance may be as great or greater than that observed cross-culturally. The diversity found among American Indian and Alaska Native populations is a good example. There are 556 federally recognized American Indian tribes and Alaska Native entities (Bureau of Indian Affairs, 2000). Furthermore, 135 Indian and 20 Alaska Native languages have been identified (Dillard & Manson, 2000). These statistics reflect vast diversity in language, tradition, religious beliefs, and other cultural values among Indian and Alaska Native people. The diversity in tradition and, to a lesser extent, language is also evident among individuals of "Hispanic origin," who can belong to one of many races and may
be identified as belonging to one or a mixture of various subgroups, each with
their own cultural values and Spanish dialects (e.g., Mexican, Puerto Rican,
Cuban, Dominican, Central American, South American, or Spaniard). Similar
diversity exists among individuals of African origin, some of whom have long
cultural and familial histories in the United States dating back to the southern
slaves in the 1700s, while others have more recently immigrated from
Egypt, Ethiopia, Ghana, Nigeria, Haiti, Panama, Jamaica, Trinidad, Barbados,
and other Caribbean nations (U.S. Department of Commerce, 1995). Even
immigrants originating from the same country, such as Vietnamese, Cambodian,
and Laotian immigrants, have differing circumstances and periods of
arrival as well as differing languages, making the Southeast Asian popula-
tion, for example, far from homogeneous, and distinct from other Asian and
Pacific Islander groups (Holzer & Copeland, 2000). Differences in socioeco-
nomic status, education, and geographic location within cultural or ethnic
groups further explain why intracultural differences may be greater than
those identified between groups. Generalities aimed at describing any one
broadly defined cultural group (e.g., Asian Americans) or its cultural and cog-
nitive values may be grossly misleading or inaccurate in the individual case,
and may contribute little that is constructive to the assessment process.

**CULTURE AND ASSESSMENT**

Cultural values and expectations are exemplified in the manner in which in-
dividuals approach a cognitive evaluation and respond to test stimuli. Many
individuals who have been raised and educated in the United States have
been exposed to the expectations for performance in the context of a test sit-
tuation. Testing in many public school systems is introduced early in the for-
mal education process, and the role of an examinee becomes familiar early
on. Standardized testing may occur with regularity for children in the United
States, while for others, such as those from some Latin American cultures,
standardized testing is an unfamiliar and often unwelcome concept.

For individuals familiar with formal testing, the examiner may be readily
accepted as a person of authority who is to be trusted. This may be particu-
larly true if the examinee shares the same cultural heritage as the examiner.
This may not be true, however, for examinees who have immigrated from
other cultures or whose collective histories are characterized by victimization
or oppression. For example, Native American and Native Alaskan populations
suffered tremendous losses of people, leaders, language, and traditions when
they were relocated to reservations and English-only boarding schools upon
the arrival of European settlers and missionaries (Krauss, 1980; Nies, 1996).
The interventions "on behalf of" Indian and Native people resulted in multi-
faceted, multigenerational losses, which have been handed down through the
generations of their respective cultures. Consequently, examinees may have
"valid" fear and distrust of both the assessment process and the examiner who purports to act in the interest of the examinee or family. Under some circumstances, parents may be fearful and suspicious of the purpose of testing, may resist an evaluation, and may directly or indirectly communicate their discomfort to their children, thus compromising performance and the validity of the evaluation. Those residing in the United States without "documented" or legal residency, for example, may be particularly fearful of any scrutiny or evaluation by those in authority. Examiners must be sensitive to these issues in testing and may consider performing an evaluation in a less structured or more informal manner. It may be necessary to meet with the child and parents over extended periods of time in order to build rapport and clarify the nature and purpose of an evaluation, the confidential nature of the evaluation, and the manner in which results will be utilized.

Acculturation

Clinicians and diagnosticians are taught to consider acculturation in determining whether an individual from one culture can be assessed using tests and norms developed with individuals from another "source" culture. Determination of the level of acculturation of the examinee may be based on superficial, informal, and somewhat limited impressions, such as whether the examinee appears to understand the conversational language of the examiner, or the examinee’s length of residence in the United States. Assessing acculturation from this perspective asks the question of whether the individual is "ready" for the test. This perspective does not directly address whether the test is "ready" or appropriate for the examinee and does not reveal the presence of specific cultural values that might influence the assessment process and results. For example, speed and competitiveness are valued in some Western cultures, and it may be implicitly understood by examinees from these cultures that speed is essential to achieving optimal scores on tests. For individuals from some Latin American cultures, speed and quality may be contradictory. In these cultures, there is a social norm that good products are the result of a careful and slow process, and work is generally not to be rushed. Individuals from some tribal and Hispanic cultures may also display in test situations a style of interpersonal relatedness that includes silence, passivity, and deference. These are behaviors intended to convey respect for the person in authority (the examiner) and may occur even in response to direct questioning, particularly when personal information is solicited (Black, 1973). An examiner who is naive to these behaviors may inappropriately assume that the behaviors indicate defensiveness or evasiveness. The individual meanings of these behaviors and their relationship to traditional values are not likely to be revealed through the administration of acculturation surveys or superficial questions designed to gauge acculturation. Rather, these
behaviors will be understood through exploration of the cultural values held by the examinee.

Most literature concerning the assessment of multicultural populations suggests that the first step in the assessment process is a determination of the examinee's level of assimilation or acculturation. More accurately, the first step in the assessment process should be for the clinician or diagnostician to acculturate to the examinee's culture (Malgady et al., 1987). This step includes an understanding of the cognitive abilities that are valued and trained within the culture; developmental expectations for acquisition of various cognitive competencies; cultural expectations for display of cognitive competencies and abilities; knowledge of the collective and individual history of the examinee; and knowledge of cultural values, traditions, and beliefs that will impact both the process and outcome of assessment (Berry, 1988; Dillard & Manson, 2000).

**TEST TRANSLATION AND ADAPTATION**

Psychologists have long been using tests developed for Western, urban, middle-class, literate people to evaluate cognitive abilities of people belonging to other cultural groups (Ardila, Rosselli, & Ostrosky-Solis, 1992). Survey data collected by Echemendía, Harris, Congett, Diaz, & Puente (1997) illustrate that the vast majority of Spanish-speaking individuals who receive neuropsychological evaluations in the United States are tested with the same tests as their English-speaking counterparts. These tests are either formally or informally translated and the results frequently are interpreted using norms from an English-speaking population, such as the population on which the original English-language measure was normed.

The literal translation of a test from the original language for use with a new target population is fraught with methodological, statistical, and theoretical problems. Literal translations are problematic because words and phrases differ in level of difficulty, familiarity, and meaning from one culture to another. Culturally specific nuances cannot be directly translated. Even seemingly "non-linguistically" based items may be problematic. For example, the use of a picture-sequencing item that depicts the changing seasons of the year may be more complex for a Latino born in Costa Rica and living in Miami than it is for a Latino living in New York. Padilla (1979) has emphasized that the adaptation of a test to a culture requires more than simply translating the instrument from one language to another; it is also necessary to adapt the new measure to the specific demands existing in the new target culture. In a comprehensive article, Geisinger (1994) states that "[the] adaptation of assessment instruments for new target populations is generally required when the new target population differs appreciably from the original
population with which the assessment device is used in terms of culture or cultural background, country, and language" (p. 304).

The second step in the assessment process should be an evaluation of whether the assessment techniques, tests, and norms are suitable for the examinee. The International Test Commission has set forth guidelines (Van de Vijver & Hambleton, 1996) for translating and adapting tests, which are summarized as follows: (1) The test translation methods must minimize or avoid construct, method, and item bias; (2) a thorough knowledge and understanding of the language and culture into which a test will be translated must guide the translation process; (3) the materials, methods, and techniques of administration must be familiar to the target cultural group; and (4) the ecological validity of the test must be statistically documented to assure accurate score interpretation. Test translation is also addressed in Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), in which Standard 9.7 states: “When a test is translated from one language to another, the methods used in establishing the adequacy of the translation should be described, and empirical and logical evidence should be provided for score reliability and the validity of the translated test's score inferences for the users intended in the linguistic groups to be tested” (p. 99).

Once translated and adapted, the validity and reliability of the new instrument must be reestablished. It is not appropriate to assume that simply because a set of measurement domains is useful in one culture that those domains will be equally useful (or similarly measured) in a different culture. The test developer must demonstrate that the new measure assesses the same characteristics as the original measure, and does so reliably, in the new culture. Similarly, once adapted, new normative data must be generated that are specific to the new culture. Translating a test and using the normative tables from the original-language version of the test, without documenting the comparability of norms, is unacceptable practice. Norms based on native English speakers either should not be used with those who speak English as a second language or results should be interpreted by in some way reflecting the level of English proficiency (American Educational Research Association et al., 1999). Geisinger (1994) cautions that a very good (and preferably empirically supported) reason must exist in order to justify the use of U.S. norms with linguistically and culturally different populations. When making decisions about test use, “Fairness of testing in any given context must be judged relative to that of feasible test and nontest alternatives” (American Educational Research Association et al., 1999, p. 73).

Table 13.1 is a select list of tests and references that may be useful in the assessment of Spanish-speaking children and adolescents. Neuropsychological batteries in Spanish are also included.
TABLE 13.1
Select List of Tests with References for Use with Spanish-Speaking Children and Adolescents

<table>
<thead>
<tr>
<th>Test Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual Verbal Ability Tests (Munoz-Sandoval, Cummins, Alvarado, &amp; Ruef, 1998)</td>
<td></td>
</tr>
<tr>
<td>Boston Naming Test (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>NEUROPSI (Ostrosky-Solis, Ardila, &amp; Rosselli, 1999)</td>
<td></td>
</tr>
<tr>
<td>NeSBHIS (Ponton et al., 1996)</td>
<td></td>
</tr>
<tr>
<td>Phonologic and Semantic Verbal Fluency (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>Recognition of Overlapped Figures (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>Rey-Osterrieth Complex Figure (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>Sequential Verbal Memory Test (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>Spanish Language Assessment Procedures (Mattes, 1995)</td>
<td></td>
</tr>
<tr>
<td>Token Test (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>Wechsler Intelligence Scale for Children—Revised (Wechsler, 1983, 1993)</td>
<td></td>
</tr>
<tr>
<td>Wechsler Memory Scale (Ardila &amp; Rosselli, 1994)</td>
<td></td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test (Rosselli &amp; Ardila, 1993)</td>
<td></td>
</tr>
<tr>
<td>Woodcock-Johnson—Revised in Spanish (Woodcock &amp; Munoz-Sandoval, 1996a, 1996b, 1996c)</td>
<td></td>
</tr>
</tbody>
</table>

Note: In addition to the measures listed above, there are a number of tests classified as non-verbal intelligence tests available from major test publishers.

EDUCATION AND COGNITIVE PERFORMANCE

Attendance of school is perhaps the most formal way that academic learning takes place in the context of culture. Despite the considerable variability across cultures found regarding both the aims and structure of schooling, the process of formal education trains certain abilities and reinforces certain attitudes in all cultures. Development of vocabulary, grammar, and syntax skills are an example of the way that formal education can facilitate the development of verbal skills and abilities. Even on relatively simple and seemingly straightforward tests, the influence of literacy and reading skills becomes apparent. Lecours et al. (1987) found that illiterate individuals tended to select one of the incorrect foils on a test of simple and complex sentence comprehension. In one example, 35% of the illiterate individuals selected a picture representing a girl running as the correct representation of the statement, “The girl walks.” The authors suggested that the illiterate individuals may have considered the action of running equivalent to walking fast and that such errors might represent incomplete semantic organization related to lack of school education, not to be confused with aphasic errors. Translating verbal items from one language into another may also yield less familiar tenses or
undesirable syntax, which may impact the manner in which children and adults decode and understand logical relationships.

Examiners often make the mistake of assuming that, if an individual has had limited formal education, literacy and linguistic skills understandably will be influenced but that nonverbal skills will be unaffected. Several studies have compared performance on verbal and visuospatial/nonverbal tasks in adults. For example, Ardila, Rosselli, & Rosas (1989) and Rosselli, Ardila, & Rosas (1990) found performance to be worse for illiterate persons compared with literate persons on nearly all tasks, including tests of visuospatial, memory-related, linguistic, and praxic abilities. Manley et al. (1999) found that even when illiterate subjects are matched to literate persons on years of education, illiterate individuals have more difficulty on tasks such as visual matching and recognition. Manley et al. suggested that literacy may facilitate the development of skills necessary to analyze visuospatial information or, alternatively, that linguistic skills associated with literacy may assist in mediating performance on nonverbal tasks.

The development of writing skills associated with formal education has been suggested to facilitate not only performance on visuospatial, graphomotor, and fine motor tasks, but to influence other cognitive skills as well (Deregowski, 1980; Pontius, 1989; Ardila, Rosselli, & Rosas, 1989). Harris & Cullum (2001) investigated the performance of lower- versus higher-educated subjects in a sample of Mexico-origin individuals who completed both the Digit Symbol (Wechsler, 1981) and the Symbol Digit (Smith, 1982) tests. The lower-educated individuals performed worse than the higher-educated subjects on Digit Symbol but not on Symbol Digit. One possible way of interpreting this finding is that the subjects with limited education were compromised in their graphomotor ability to make the less familiar symbol responses. The writing of digits may be less influenced by education than the drawing of figures, provided some degree of literacy is attained.

Variables such as education and cultural values, however, should not be confused with the quality of one's environment. Certainly, some ethnic minority groups have high rates of poverty and may consequently have higher rates of exposure to prenatal and postnatal complications, malnutrition, exposure to environmental toxins, and increased risk of neurocognitive trauma related to occupational hazards (see also Ostrosky-Solis et al., 1985; Amante, Vanhouten, Grieve, Bader, & Margules, 1977). All of these factors represent additional variables that may be operating in a culture or subculture and may impact cognitive performance, but they should not be confused with culture per se. It is also important to bear in mind that, in the assessment of individuals who have received little or no formal education by Western standards, the examinee may simply lack familiarity with the task demands and may in fact be able to display the skill or intellectual ability under other, more culturally relevant test conditions. Goodenough (1936) cautioned that test items and concepts should be "representative and valid samples of the ability in
question, as it is displayed within a particular culture" (p. 5). It is dangerous to define knowledge or skills as formal "schooled" knowledge, because this may disregard the knowledge base otherwise defined by one's culture.

**LANGUAGE AND PROFICIENCY**

Language may impact assessment from a number of perspectives. For a variety of reasons, children and adults may overestimate their ability to speak English. There may be shame associated with a lack of mastery of the English language, and a risk of not being accepted by peers or by the examiner if one's true level of proficiency were discovered. Some individuals have mastered conversational English and may appear to be fluent when engaged in superficial conversation, and this may prompt the clinician to assume that the individual can be adequately tested in English. Another common mistake is to assume that, because an individual appears to possess spoken language proficiency in English, all language related skills (e.g., reading and writing) are at the same level of proficiency. Such an assumption may lead to the erroneous interpretation of variations in performance, such as erroneously finding evidence of a learning disability.

Even for bilingual individuals who appear to have mastery of both languages, fluency in the second language may be diminished when placed in a stressful situation such as an interview or evaluation (Peck, 1974). With or without a stressful situation, concepts may be less available in the second language, and words to express internal experiences or symptoms may be less accessible in the second language. Schooling language can indeed be a highly significant and decisive variable in developing a language preference as well as in determining the facility with which complex material can be discussed. Limiting an individual to one language in the assessment may limit the completeness of the communication, which has the potential of introducing error into the history-taking and testing. Monolingual assessment in either language may fail to address the need of a bilingual person to utilize both languages for effective communication. This is a complex issue to address in testing because, although it may be inappropriate to use norms based on a monolingual sample, norms for bilingual subjects are largely unavailable.

**THE USE OF TRANSLATORS**

The Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations (American Psychological Association, 1990) state that psychologists should interact in the language requested by the client. When
an examiner does not speak the language of an examinee, translators often are included in the assessment. However, translators must be used cautiously because error can be introduced into the evaluation at many points. First, while proficient in the language of the examinee, a translator may not be educated in the particulars of that individual's culture, and may fail to communicate the nonverbal content of the communication or the emotional content that is shared between the translator and the examinee. McShane & Plas (1984) noted that in some Indian cultures there may be more reliance on nonverbal language systems than on verbal language to communicate complex "language." A lack of response to a question may reflect a cultural norm regarding the permissibility of discussing the topic and may have little to do with resistance or aggression. A skilled translator should not merely be a translator of language, but of the examinee's culture as well.

Translators also may be burdened with greater clinical responsibilities than they are trained or prepared to manage. Echemendía, Harris, Congett, Diaz, and Puente (1997) found that translators were often used in the assessment of monolingual Spanish speakers in the United States, but that the translators possessed little or no formal training in neuropsychology. When using a translator, the clinician responsible for the assessment has no direct knowledge as to the accuracy of the translation, or whether the translated information maintains the intended meaning of the original communication. If the translator has little or no training in the theories and methods of cognitive assessment, the psychologist at best may be losing valuable qualitative data, and at worst may be receiving information that has been embellished by the translator in an attempt to "help" the patient. The translator should be viewed as an extension of the clinician. It is the responsibility of the clinician to thoroughly train the translator, particularly regarding the goals of assessment, and to educate himself or herself on the complexities of cross-cultural assessment and test selection.

**LINGUISTIC CONSEQUENCES IN PERFORMANCE**

Specific linguistic idiosyncrasies also may directly influence performance in cognitive or neuropsychological testing. For example, formal spelling does not receive the same emphasis in Spanish as it does in English. Also, the use of spelled abbreviations is common in English, but not in Spanish. A native Spanish speaker would never read UCLA as "U-C-L-A" (or "u" "se" "ele" "a" in Spanish) but rather "ukla." Furthermore, English speakers may know the alphabet sequence better because it is essential in spelling. These factors may explain some differences between Spanish and English speakers in phonemic word fluency (Taussig, Henderson, & Mack, 1992; Rosselli et al., 2000). In tests of phonemic fluency, subjects are instructed to generate words that
begin with a specific letter (Rosen, 1980). Typically, subjects are given three
different letters (e.g., F, A, and S) and are given one minute for each letter
to generate words. Spanish speakers tend to generate words according to
phonologic similarities, disregarding spelling (e.g., saI, cebolla, and cespIed).
In most cases it is inappropriate to strictly adhere to scoring criteria when evalu-
ating a Spanish speaker because this penalizes the examinee for choosing
phonologically correct ("ce" or "se") but orthographically incorrect ("C"). It is also
inappropriate to apply English-language fluency norms when evaluating
performance.

The strong tendency to "spell" in English is observed even when reciting
numbers. For example, in English the numbers "3, 4, 5" are typically read as
"three, four, five." Spanish speakers, on the other hand, prefer to cluster num-
bers when presenting numbers to others (e.g., a phone number), such that "3,
4, 5" becomes "three hundred forty-five". The single-digit presentation of
digit-span stimuli is less familiar to Spanish speakers and this may interfere
with encoding and subsequent chunking for recall. Indeed, reduced digit-
span performance has been identified in Spanish speakers (Naveh-Benjamin &
Ayres, 1986; Loewenstein, Arguelles, Barker, & Duara, 1993). It also is pos-
sible that children and adults who have not achieved proficiency in English
may utilize internal translation strategies to convert test stimuli into more fam-
iliar terms in an effort to improve encoding. For example, a child who hears
a string of numbers in English may attempt to translate the numbers into the
native language for rehearsal and then translate them back into English when
providing the examiner with the response. This kind of effort reduces the
availability of processing resources for other aspects of the task, such as re-
ordering digits in reverse when performing Digit Span Backwards.

A second linguistic-related explanation for reduced digit span concerns
the reduction of immediate memory span due to the longer amount of time
it takes to articulate words in some languages. Baddeley, Thomson, and Bu-
chanan (1975) first demonstrated that memory span for short words was
greater than for long words. Although the number of syllables or phonemes
in the stimulus may explain the effect, the effect persists when syllables and
phonemes are held constant (Baddeley, Thomson, and Buchanan, 1975; Ellis &
Henneley, 1980; Olazaran, Jacobs, & Stern, 1996) The more critical variables
that affect the number of words that can be rehearsed in the phonological
loop appear to be spoken word length or articulation time. When the tech-
nique called articulatory suppression (e.g., repeatedly saying "A-B-C-D") is
used to prevent rehearsal, the word length effect is no longer evident.

It is known that languages differ in their spoken word lengths. Languages
with short articulation time for words therefore are expected to make fewer
demands on phonological loop resources. Studies of word or digit span
in non-English languages illustrate the potential differences in span. For
example, Cantonese and Chinese spans are greater than English on simple span tasks (Hoosain, 1979), and it appears there are differences between English and Spanish digit spans, with English having the greater span (Loewenstein et al., 1993; Naveh-Benjamin & Ayres, 1986).

Our own work in a recent pilot study (Harris, Adler, & Cullum, 2000) has suggested that it is important to consider test stimuli and the interaction with spoken word length for a given language. We compared performance of Spanish and English speakers on four tests of working memory, one of which we designed to be equivalent in the Spanish and English languages on spoken word length. Preliminary data analyses comparing 10 normal Spanish and 5 normal English speakers revealed significantly worse performance for Spanish speakers on the three tasks for which word length was not controlled. However, there was no difference between English and Spanish speakers on the equated task.

These very preliminary findings suggest that processing resources may be differentially affected in English versus Spanish on "simple" working memory tests and that proficiency in English also can be expected to influence available processing resources. These findings also have implications for performance on other cognitive tasks (e.g., those that rely on working memory and organizational strategies to aid encoding or learning of new information), and suggest that care must be taken to equate tasks on word length and difficulty in the assessment of cross-linguistic samples.

Word-length effects also may account for reduced semantic fluency performance found in adults who speak Spanish, compared with other languages. For example, Kempler, Teng, Dick, Taussig, and Davis (1998) compared performances of Chinese, Hispanic, and Vietnamese immigrants in their native languages on a semantic verbal fluency task (animal naming) and found that the Spanish-speaking Hispanics produced the fewest responses, even after statistically controlling for age and education. The average word length of animal names was longest in Spanish and shortest in Vietnamese. Jacobs et al. (1997) also reported fewer semantic category exemplars for Spanish compared with English speakers. While these issues have not been directly investigated in children, there is reason to expect that semantic fluency tasks given in the child's native language would yield similar findings.

CASE STUDIES

Having now reviewed some of the important considerations in cognitive and neuropsychological assessment with regard to culture, education, and language, we present two case studies as illustrations of these considerations. In the first case, relevant history and specific test results are presented to assist the clinician in formulating hypotheses and arriving at a working diagnosis. In the second, a description of specific factors likely to influence the ex-
aminee's and family's participation in the assessment process are addressed, with an example of reconceptualizing the assessment.

Case One

Reason for Referral

Maria was an 8-year, 8-month-old white Hispanic female of Mexican origin. She was referred for a neuropsychological evaluation because of learning difficulties at school. She was involved in a motor vehicle accident approximately 2.5 years prior to the evaluation. There were no medical records available concerning the accident. However, according to her parents, Maria had been unrestrained and was thrown out of the vehicle. She was initially non-responsive but responded immediately to CPR initiated by her father. She had a right frontal laceration and numerous other abrasions on her shoulders, arms, and leg. She was taken to a local hospital and then a regional trauma center, where she was discharged after one day. Her parents described her as withdrawn and apathetic upon discharge from the trauma center but believed she had returned to her baseline personality functioning within two to three days. Her parents were unaware of any difficulties regarding language or other cognitive functions following the accident. Her neurological exam at the time of her neuropsychological evaluation was completely normal, and no sequelae from the accident were identified. Vision and hearing were also reported as normal.

Sociodemographic History

Maria was born in a large urban city in Mexico, where her parents worked in manufacturing. Maria's prenatal and birth histories were unremarkable, and she was reported to be a healthy child, attaining normal developmental milestones. There was no history of loss of consciousness or head injury aside from the car accident. Spanish was the only language spoken in the home, and consequently Maria only began to learn English when she entered preschool at the age of four. She and her parents immigrated to the United States when Maria was two, in search of a better economic life and future for Maria and her sister. Both parents completed 11 years of education in Mexico, equivalent to a high-school education in the United States.

Observations

Maria's parents reported that she had difficulty expressing herself in both English and Spanish. In kindergarten, Maria had difficulty learning to write her name and recognizing letters of the alphabet. In first grade, additional problems were identified in learning to read and understanding and following directions in English. Her third-grade teacher reported that Maria had
difficulties comprehending spoken language and needed frequent repetition in order to understand what she was learning. Maria was described as frequently forgetting what she had learned. Additionally, she was described as having difficulties with spelling, the mechanics of math computations, problem solving, following simple directions, and telling time. She had problems understanding stories that were read aloud and discussed in the classroom, and she took an excessive amount of time to respond to questions, sometimes giving no response at all. She was functioning below grade level in all academic areas at the time of the evaluation.

Because her language difficulties initially appeared to be worse in English than in Spanish, she was placed in an English as a Second Language (ESL) program during first grade and remained in the program through the third grade. Maria complained of being embarrassed about going to school and did not like leaving the regular classroom to receive ESL instruction in English, math, and reading. She was aware that she had difficulties with learning compared with peers, and she at times became tearful in school. She reportedly enjoyed doing homework and reading in English at home.

Maria was evaluated by a bilingual examiner whose first language was Spanish. Maria presented for the evaluation as a pleasant, well-groomed child who was rather passive and quiet. She did not initiate any conversation with the examiner and frequently would stare quietly after being asked a question in either Spanish or English. When given extended time to respond, she would often attempt to answer a question, but at other times she required many cues and follow-up prompts in order to formulate a response. Test instructions were presented to Maria in both English and Spanish. Departure from standardized test administration method was done to give Maria the best chance of understanding test requirements. She remained motivated and persistent throughout the evaluation, but often required extra assistance, such as repeated explanation in both languages. She also tended to exhibit a trial-and-error approach to novel tasks.

Results

Maria's test results are presented in Table 13.2. Although both raw and standard scores are provided for illustrative purposes, it is the responsibility of the test user or examiner in any testing situation to know the applicability of each test to specific individuals and to employ norms appropriately (American Educational Research Association et al., 1999).

Maria's performance on the WISC-III resulted in a Verbal IQ score of 72, a Performance IQ score of 87, and a Full Scale IQ score of 77. This pattern of better Performance compared to Verbal IQ score is common among Hispanic children, especially those whose first language is Spanish. With the exception of the Verbal Comprehension index score, which was below average, Maria's index scores were in the average to low-average range.
### Table 13.2
Case One Results

- **Wechsler Intelligence Scale for Children—Third Edition**

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Scale IQ</td>
<td>77</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>72</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>87</td>
</tr>
<tr>
<td>Verbal Comprehension Index</td>
<td>68</td>
</tr>
<tr>
<td>Perceptual Organization Index</td>
<td>90</td>
</tr>
<tr>
<td>Freedom from Distractibility Index</td>
<td>84</td>
</tr>
<tr>
<td>Processing Speed Index</td>
<td>88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>6</td>
</tr>
<tr>
<td>Similarities</td>
<td>5</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>9</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>2</td>
</tr>
<tr>
<td>Comprehension</td>
<td>3</td>
</tr>
<tr>
<td>Digit Span</td>
<td>5</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>9</td>
</tr>
<tr>
<td>Picture Arrangement</td>
<td>6</td>
</tr>
<tr>
<td>Block Design</td>
<td>9</td>
</tr>
<tr>
<td>Object Assembly</td>
<td>9</td>
</tr>
<tr>
<td>Coding</td>
<td>7</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>8</td>
</tr>
</tbody>
</table>

- **Wechsler Individual Achievement Test**

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Reading</td>
<td>81</td>
</tr>
<tr>
<td>Mathematical Reasoning</td>
<td>76</td>
</tr>
<tr>
<td>Spelling</td>
<td>77</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>80</td>
</tr>
<tr>
<td>Numerical Operations</td>
<td>87</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>NA</td>
</tr>
<tr>
<td>Oral Expression</td>
<td>NA</td>
</tr>
<tr>
<td>Written Expression</td>
<td>NA</td>
</tr>
</tbody>
</table>

- **California Verbal Learning Test**

| Trial 1                          | R = 4  | z = -1.0 |
| Trial 5                          | R = 9  | z = -0.5 |
| Trials 1–5                       | R = 30 | T = 40   |
| List B                           | R = 6  | z = 0.5  |
| Short Delay Free Recall          | R = 2  | z = -2.0 |
| Short Delay Cued Recall          | R = 2  | z = -3.0 |
| Long Delay Free Recall           | R = 0  | z = -3.5 |
| Long Delay Cued Recall           | R = 3  | z = -2.5 |
| Recognition Hits                 | R = 11 | z = -1.0 |
| False Positives                  | R = 3  | z = 0.0  |
| Discriminability                 | R = 84%| z = -0.5 |
| Perseverations                   | R = 0  | z = -1.0 |
| Intrusions                       | R = 3  | z = -0.5 |

*(continues)*
Table 13.2—Continued

- Rey-Osterrieth Complex Figure
  Copy R = 24 \( z = 0.05 \)
  Recall R = 16 \( z = -0.05 \)
- Boston Naming Test R = 21 \( z = -3.92 \)
- Animal Naming R = 6 \( z = -2.34 \)
- Token Test R = 48 \( z = -1.17 \)
- Clinical Evaluation of Language Fundamentals—Third Edition
  Scaled Scores
  Sentence structure 8
  Concepts and directions 7
- Gordon Diagnostic System
  Delay Task
  Efficiency ratio R = 0.83 WNL
  Vigilance
  Correct R = 40 WNL
  Commissions R = 3 WNL
  Distractibility
  Correct R = 35 WNL
  Commissions R = 6 WNL
- Sensory-Perceptual Examination
  Double simultaneous visual stimulation
  Right errors R = 0 WNL
  Left errors R = 0 WNL
  Tactile finger identification
  Right errors R = 0 WNL
  Left errors L = 0 WNL
  Finger Tapping Test
  Right R = 32 WNL
  Left L = 29 WNL

Note: R = raw score, T = T score (M = 50, SD = 10), z = z score (M = 0, SD = 1), WNL = within normal limits. Qualitative levels are provided when published norms do not permit conversion.

Maria was administered two receptive language subtests from the Clinical Evaluation of Language Fundamentals—Third Edition (CELF-III) in Spanish and scored in the low-average to average range. She refused to converse in Spanish with the examiner, precluding administration of any of the expressive language subtests. Qualitative observations of subtest performance revealed that she had particular difficulty with lengthy and complex directions. She was unable to follow three-step directions and had difficulty with right-left orientation. Her ability to follow directions on the Token Test was consistent with the CELF-III. She was able to follow some multistep instructions in English but had difficulty with more complex syntax. She was administered the Boston Naming Test and asked for the names of objects in Spanish whenever she made an error in English. This did not appreciably improve her performance. There was only one item (comb) that she was able to name in
Spanish but not in English, and she was unable to name a broom in either language. Her verbal fluency skills, as assessed on the Animal Naming Test, were below average in English, and she was unwilling to attempt the task in Spanish. There was limited opportunity to observe her spontaneous language in Spanish with her parents. She exhibited some grammatical and syntactical errors, although to a lesser extent in Spanish than when she spoke in English.

Using the Gordon Diagnostic System, attention was in the normal range. Maria's verbal learning was adequate, particularly considering her verbal cognitive abilities. On the California Verbal Learning Test, she displayed an acceptable learning slope. However, following interference, her spontaneous recall was impaired, and she did not benefit from cueing. Her recognition of words presented earlier in the evaluation was in the low-average range. This pattern suggested that she was adequately encoding words but had difficulty with retrieval. Maria refused to participate in a Spanish list-learning test. Nonverbal learning and memory were evaluated as adequate. Maria's motor performance and sensory perceptual examination were within normal limits.

On academic achievement testing using the Wechsler Individual Achievement Test (WIAT), Maria's word recognition and reading comprehension skills were in the low-average range. On the Reading Comprehension subtest, she exhibited problems drawing conclusions, comparing and contrasting types of material, recognizing implied cause and effect, and remembering or recognizing details from narrative information she had just read. Spelling skills were impaired and errors were predominantly phonetic in nature. Maria's mathematical reasoning skills were also quite poor, although she was able to perform simple calculations somewhat better than word problems.

**Summary**

In summary, Maria demonstrated deficits in verbal and linguistic skills consistent with a mixed receptive-expressive language disorder, which did not appear to be attributable to her status as a child learning English as a second language. Her results were not consistent with a traumatic brain injury. In fact, a careful exploration of her history revealed that the language problems predated the car accident. Maria's reluctance to speak with the examiner in Spanish suggested that she was aware she would gain little advantage in Spanish, although it could also suggest that Maria was struggling with identity and acculturation issues regarding her ethnic and linguistic background.

Although the assessment was hampered by Maria's reluctance to participate in testing in Spanish, test data were supplemented by the clinician's careful history-taking and interviews with the parents and previous and current teachers. This approach supported hypothesis-testing and arrival at a working diagnosis for further evaluation. Some of Maria's deficits, such as
Josette G. Harris et al.

her retrieval difficulties on the list-learning task, could be accounted for by her language dysfunction. On the other hand, Maria displayed a number of strengths that could go unrecognized in school because of the emphasis on verbal academic instruction. These strengths included intact attention, nonverbal learning and memory, and visuospatial and visuomotor skills.

Specific recommendations included a thorough speech and language evaluation in both languages to confirm the diagnosis of a language disorder. The examiner also recommended that the assessment process continue in other respects. Maria’s level of intellectual functioning may have been underestimated because of a possible language disorder. Use of norms based upon English-speaking children may not be appropriate in this case, and may have been another source of variability in estimating Maria’s intellectual and cognitive skills.

Case Two

Reason for Referral

Carlos was a 7-year, 8-month-old white Hispanic male of Mexican origin living in a small rural farming community. He was referred for educational testing by his homeroom teacher because of poor compliance with classroom assignments, inattentiveness, and academic underachievement, particularly in reading and spelling. According to Carlos’ teacher, Carlos frequently did not respond to her questions, although he appeared to hear and comprehend them. Although he lagged behind his peers in reading and spelling, he was making some progress in his oral reading skills. His teacher was concerned, however, that when Carlos was asked questions to ascertain his understanding of reading material, he often responded that he did not know the answer. Carlos attended ESL programs for two hours per day and remained in the regular classroom for the remainder. The description of his behavior by the ESL tutor was consistent with that of his homeroom teacher.

Sociodemographic History

Little was known about Carlos’ history, other than that he was born in Mexico. He attended only one grade in Mexico prior to coming to the United States with his family, who initially migrated as seasonal farm workers.

Assessment

The psychologist first recommended that the teacher, with the assistance of a translator, call Carlos’ parents in order to arrange a meeting to review Carlos’ school progress, prior to planning any formal assessment. This approach was recommended because of the teacher’s more direct contact with Carlos and his family, compared with the psychologist, who had never met the child
or his parents. With the assistance of a skilled translator who was knowledgeable about the Mexican immigrant culture and of migrant families in a neighboring community, the teacher arranged for a meeting. The meeting was arranged and the family was encouraged to invite a friend or relative from the community to the meeting if that would be helpful to them. The teacher told the family that an individual from the school who helps evaluate the progress of students and identifies ways to help students with their learning would also be present, and provided the psychologist's name. The family was asked if they would like to have a translator present, and they agreed that this would be helpful.

Prior to meeting with the family, the psychologist consulted with individuals of Mexican origin and individuals actively involved in providing services to Hispanic individuals in the community, such as spiritual guidance and assistance services. The psychologist learned that formal education could be limited for migrant and first-generation immigrant parents in the community and that literacy could be in question. This called for careful attention to the informed consent procedure. He also learned that newly immigrated individuals might have undocumented residency status, which could influence the comfort level of the parents in response to working with school personnel and responding to questions regarding Carlos or the family. It would be necessary to reiterate the educational rights of children and to be mindful of the potential anxiety that any formal evaluation might initiate.

The parents, Carlos, and two younger siblings arrived at the meeting unaccompanied. The parents previously had had very limited contact with the school and appeared both concerned and anxious, as did Carlos. The family also had very limited contact with the teacher, but her presence appeared to be reassuring. Care was taken to explain the purpose of the meeting and to explain that it would be helpful to understand more of Carlos' culture and educational experience in order to understand both his educational weaknesses as well as strengths. The psychologist apologized in advance for asking any questions that might touch on private matters of the family and explained the relevance of such questions to understanding Carlos' skills and abilities and to clarifying the source of educational difficulties and planning interventions. The psychologist also invited questions from the parents.

The psychologist, with the assistance of the translator, learned that the family had lived in an isolated and very poor rural community in Mexico. Carlos attended a school in which books, paper, and pencils were quite limited and shared among students, who ranged in age from 6 to 10 and were all educated in the same informal classroom together. Most children in the community did not advance beyond four years of school and in fact there was no option available in their community to advance beyond six years. Carlos' mother herself had attained four years of education and his father six years. At the urging of a relative, Carlos' parents left their home in Mexico to seek a better life for their children in the United States. They first worked as
seasonal migrant workers. Although Carlos’ parents were very interested in his education, they were unsure of their own ability to contribute to his learning process.

The psychologist obtained a thorough history of Carlos’ behavior in the home and learned that Carlos did not exhibit in the home any of the behaviors in the classroom as described by his teacher. In fact, he was quite respectful of his parents, always responded to their questions and directions in Spanish, and was quite attentive. His parents reported that Carlos liked to watch English-language television in the afternoons and would do so for extended periods of time. The parents supported this as a means for their son to gain additional exposure to the English language. They also described other age-appropriate play and interests for their child but indicated that it was necessary for Carlos to assist his mother with the care of his younger siblings in the evenings due to the long hours his father worked. Carlos’ parents appeared to appreciate the focus of attention on their son and equally appreciated the opportunity to ask questions of the clinician. This provided the clinician an opportunity to explain that there was some question concerning Carlos’ skill level given that his native language was Spanish, but that the clinician and Carlos’ teacher hoped to better understand his strengths and weaknesses in the weeks and months to come. The meeting closed with a plan to meet again in one month to evaluate Carlos’ progress.

Summary

In this example, assessment was limited to initiating a process of information gathering and acculturation of the psychologist and teacher to the background and experience of Carlos and his family, both in their native country and in their new place of residence. The psychologist determined that Carlos’ formal education in Mexico had been qualitatively much different than his relatively new experience in an educational setting in his new country, and that he was still early in the process of learning expectations in his new school culture. Following the meeting with Carlos and his parents, the psychologist felt that it was too early in the educational process to evaluate Carlos with formal measures in English and that even baseline testing in Spanish would be of limited value. A plan of intervention for Carlos included assisting him to acculturate to the educational setting, assisting him with peer relationships, and reassessing his ESL plan to consider more intense tutoring and close monitoring of his progress.

CONCLUSION

The challenge of providing assessment and intervention services to culturally and linguistically diverse individuals, particularly children, clearly will continue. It is estimated that, by the year 2050, some ethnic minority groups (e.g.,
Hispanic and Asian/Pacific Islander) will increase by more than 200% from the year 2000 estimates, while others will also see sizable growth (U.S. Census Bureau, 2000). Immigration patterns for 1995–1998 indicate that approximately 20% of the immigrants admitted each year were under the age of 15, and another 33% were between the ages of 15 and 29 (U.S. Department of Justice, 1999). This trend likely will continue. Children currently constitute 31% of the Hispanic population, 27% of the American Indian/Aleut/Eskimo, 27% of the Asian/Pacific Islander, and 19% of the Black population (U.S. Census Bureau, 2000). It is unrealistic to assume that the challenge of cross-cultural assessment of children and adults can be managed by simply referring ethnic minority individuals to bicultural, bilingual clinicians and providers. A practical point is that there simply are not enough providers to match the cultural and linguistic backgrounds of examinees. An additional point, however, is that bilingual, bicultural psychologists are not immune from the cultural issues addressed in this chapter. The tools for meeting the challenge of cross-cultural assessment lie in the achievement of competence and expertise (American Psychological Association, 1992; Canadian Psychological Association, 1991), the acculturation of providers to the individual cultures of those they serve, and the expansion of the empirical knowledge base regarding cross-cultural and cross-linguistic issues in assessment.

References


Bureau of Indian Affairs (2000). Answers to frequently asked questions. (On line). Available

tario: Canadian Psychological Association.

Academic Press.

Dillard, D. A., & Manson, S. M. (2000). Assessing and treating American Indians and Alaska Na-
tives. In I. Cuéllar and F. A. Paniagua (Eds.), Handbook of multicultural mental health. San Diego:
Academic Press.


testing and the relative ease for mental calculations in Welsh and English. British Journal of Psy-
chology, 71, 43–51

influencing the normative interpretation of assessment instruments. Psychological Assessment,
6(4), 304–312.

Goodenough, F. (1936). The measurement of mental functions in primitive groups. American Anthro-
pologist, 38, 1–11.

Harris, J. G., Adler, L. E., & Cullum, C. M. (2000). Spoken word length effects on working memory

Harris, J. G., & Cullum, C. M. (2001). Symbol vs. digit substitution task performance in culturally
and linguistically diverse populations.

Holzer, C. E., III, & Copeland, S. (2000). Race, ethnicity, and the epidemiology of mental disor-
ders in adults. In I. Cuéllar and F. A. Paniagua (Eds.), Handbook of multicultural mental health. San

of Genetic Psychology, 135, 263–268.

neuropsychological assessment: A comparison of randomly selected, demographically
matched cohorts of English-and Spanish-speaking older adults. Journal of Clinical and Experi-
mental Neuropsychology, 19(3), 331–339.

tion, and ethnicity on verbal fluency. Journal of the International Neuropsychological Society, 4(6),
531–538.


testing in culturally contrasted populations (control subjects). Neuropsychologia, 25(1B),
231–245.

neuropsychological test performance of Spanish-speaking and English-speaking patients


fect of literacy on neuropsychological test performance in nondemented, education-matched

Services.

ing from the WISC to the WISC-R. School Psychology Review, 13(1), 61–70.
13. Cross-Cultural Cognitive and Neuropsychological Assessment


