Utah State University

# DigitalCommons@USU

5-1997

# A Comparative Analysis of Seven Published Self-Report Measures for Assessing Internalizing-Type Symptoms in Children and Adolescents

Kathryn E. Anderson
*Utah State University*

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Psychology Commons

Running Head:  INTERNALIZING TYPE SYMPTOMS

A Comparative Analysis of Seven

Published Self-Report Measures for Assessing Internalizing-Type Symptoms in

Children and Adolescents.

Kathryn E. Anderson

Utah State University

Plan B Research Paper

# A Comparative Analysis of Seven
# Published Self-Report Measures for Assessing Internalizing-Type Symptoms
# in Children and Adolescents.

The broad domain of internalizing disorders encompasses a variety of symptoms that are specific to child and adolescent populations and generalizable to adult populations. Internalizing disorders, commonly referred to as "emotional problems," include such problems as depression, anxiety, social withdrawal, somatic complaints, and low self-esteem. The other side of this classification dichotomy is that of externalizing disorders (e.g., conduct disorder, attention-deficit hyperactivity disorder), which involve overt behaviors considered as "undercontrolled". In contrast, internalizing problems involve behaviors that possess an "overcontrolled" quality. Such a covert nature leads to difficulty in identification and diagnosis, as they often go unnoticed by the child's teachers, parents, and peers, thus resulting in prolonged distress of the individual. In addition, the broad construct of internalizing disorders often involves an element of comorbidity among the various internalizing disorders. This element of co-existence is further evidenced in the recent interest of clinicians regarding the question as to whether anxiety and depression represent two distinct states or rather a broad-band construct termed "negative affectivity" (Hodges, 1990). Negative affect is a "broad and pervasive predisposition to experience negative emotions that have further influences on cognition, self-concept, and world view" (Carey, Clark, & Watson, 1988, p. 347).

Within the past decade, the clinical importance and magnitude of internalizing disorders (i.e., emotional problems) in children and adolescents has increasingly permeated the concerns of educators and psychologists. This increasing level of attention arises from several sources, one of which is the strong negative effect that internalizing disorders have on academic achievement. One of the most prevalent negative outcomes of internalizing disorders is low self-esteem. Low self-esteem not only effects how children feel about themselves in general, but it is also strongly associated with problems in academic achievement (Merrell, 1994). It has been suggested that

academic self-concept is the single most powerful affective predictor of academic success (Bloom, 1976). Prevalence estimates for the most common internalizing disorders (depression and anxiety) for children and adolescents have ranged from 2% to 6.4% (Bartels, et.al., 1995) and have been evidenced to become increasingly prevalent with age. Despite the prevalence and negative effects on academic achievement, the identification and treatment of children experiencing emotional distress are hindered by the covert nature of internalizing disorders. In general, early symptom recognition and effective diagnosis of child/adolescent internalizing disorders are critical elements in prompt, appropriate interventions to reduce prolonged suffering and the associated social and academic problems.

School professionals are facing increasing demands to provide appropriate psychological services for emotionally disturbed students. The necessity of effective and comprehensive assessment instruments thus is brought to the forefront of psychological evaluation of children and adolescents. Although the majority of existing self-report measures effectively assess specific syndromes (i.e., anxiety, depression) within the domain of internalizing disorders, few instruments are designed to measure the broad band domain of internalizing symptomatology. Consequently, self-report measures are consistently utilized in combination as part of a multiaxial assessment system in evaluating all domains within the breadth of internalizing disorders. The use of these diagnostic instruments in a comprehensive battery for the general assessment of psychopathology further emphasizes the necessity for existing self-report instruments to be reliable and valid measures of internalizing symptoms among children/adolescents.

Psychometrically sound tests are essential for valid identification of children/adolescents who may be experiencing symptoms of internalizing disorders. Measures must be both valid and reliable to be considered as technically adequate. Through this review, the psychometric properties and characteristics of seven self-report instruments will be critically evaluated to provide counselors, psychologists, and other professionals working with school-aged populations with pertinent information regarding the clinical efficacy of the measures and their intended purposes. Self-report measures of internalizing type symptoms can facilitate various assessment decisions and can be used as: (a) a screening device in school-based or clinical settings; (b) an evaluative measure of treatment progress; (c) a research instrument for studying internalizing disorders; and (d) a part of the referral and identification process in assessing general

psychopathology. An important advantage of using self-report measures over other assessment methods is their ability to lend themselves well to recording what goes on in the inner world of the child/adolescent (i.e.,thoughts, feelings, attitudes, internal reactions to people and events).

Although the majority of researchers and clinicians view self-report measures as an integral part in the assessment of child/adolescent psychopathology, limitations in self-report instruments' efficacy in tapping the subjective world of the individual may arise from several underlying factors. The extent to which children are capable or willing to report on internalizing symptoms has the potential to affect the acquisition of a truly representative sample of the child's psychological health or pathology. Additionally, developmental factors may play a role in restricting a child's ability to validly quantify and qualify their emotional, behavioral, and cognitive status (Piaget, 1983). However, various clinicians and researchers purport self-report measures as being sensitive to such issues as a child's capabilities and developmental status, as well as, the necessity to minimize the demands for verbal expression in the obtainment of systematic input from children and adolescents (La Greca, 1990). Another factor which may limit the validity of the reported internalizing symptoms is the incident in which an youngster recognizes the intended purpose of the particular scale. This knowledge of the scale's objectives may elicit acquiescent responding on the part of the individual; such as, answering in a socially desirable manner or faking of his/her responses. Some self-report instruments do incorporate "validity checkpoints" (i.e., Lie scale or reverse-scored items) to help control for acquiescent responding, but it is unclear whether such methods result in a more valid assessment with children.

Self-report measures have the potential to be useful components in the general assessment of psychopathology among children and adolescents. Thus, consumers should be cognizant of the merits associated with the available self-report instruments. In raising the awareness of consumers to such beneficial components for the assessment process, this paper provides the user with critical reviews and comparative analyses of seven self-report measures for internalizing type symptoms. Prior to discussing the reviews and comparisons, a brief presentation is provided of the guidelines for evaluating self-report measures of internalizing type symptoms. Subsequent to the review and comparisons is a discussion of the procedures employed for inclusion and for critical review of the instruments.

**Guidelines for Analyzing Self-Report Measures for Internalizing Type Symptoms**

1. *Content and use of self-report measure.* The important aspects to be considered with regard to this dimension include the instrument's comprehensive nature and utility of manuals and material, and the appropriateness of scoring procedures and administration format.

2. *Standardization sample and norms.* Norm-referenced self-report measures must be based on a representative standardization sample for appropriate interpretation since such measures are employed throughout the United States. Procedures for norming should be clearly delineated with regard to the year norming ensued, descriptive statistics, and the sampling procedures.

3. *Scores/interpretation.* A detailed description of the scores and the appropriateness of the scores for the instrument's purposes should be one of the principal aspects of this dimension. Interpretation of scores should also be clearly delineated and remain within the intended scope of the instrument.

4. *Psychometric properties.* This area addresses both the reliability and validity of the self-report measure. Internal-consistency and test-retest are pertinent reliability considerations for the majority of self-report instruments. Analysis of a measure's validity includes the assessment of the measure's internal validity (i.e., content validity, construct validity), as well as, the external validity (i.e., criterion-related validity, predictive validity, convergent/divergent validity) of the self-report measure.

**Procedural Guidelines**

The seven self-report measures selected for review are all norm-referenced. These instruments include (1) *Children's Depression Inventory* (CDI; Kovacs, 1992); (2) *Reynolds Adolescent Depression Scale* (RADS; Reynolds, 1986); (3) *Reynolds Child Depression Scale* (RCDS; Reynolds, 1989); (4) *Revised Children's Manifest Anxiety Scale* (RCMAS; Reynolds & Richmond, 1985); (5) *State-Trait Anxiety Inventory for Children* (STAIC; Speilberger, 1973); (6) *Youth Self-Report: Internalizing scale* (YSR; Achenbach & Edelbrock, 1987); and (7) *Behavioral Assessment System for Children: Self-Report of Personality form, Internalizing scale* (BASC; Kamphaus & Reynolds, 1992). Summaries of each instrument are presented in Table 1.

Selection criteria for inclusion of the self-report instruments were: (a) a specific focus on the assessment of internalizing-type symptoms among children/adolescents; (b) applicability in school settings; (c) status of self-report instrument as published at time of this review. The focus on published self-report instruments was deemed as being important in presenting only readily available instruments that possessed adequate research and clinical data. The seven self-report measures are grouped in relation to the particular internalizing syndrome being assessed. The scales are arbitrarily presented with the three depression measures being first, followed by the two anxiety measures, and then the two measures which evaluate the broad domain of internalizing symptoms. The analysis of each measure begins with a brief description, followed by evaluations based on the four previously stated dimensions as well as the author's critical review and judgment of the quantity and quality of available research data and summary information. Psychometric properties of each instrument are characterized according to the criteria of *limited*, *adequate/good*, and *excellent* (see Table 2 and 3 for reliability and validity summaries, respectively). The criteria and evaluations employed in this analysis were guided by the author's judgments of converging evidence for each instrument due to the fact that no ruling guide is available for the determination of technical acceptability of a given test or scale (APA, 1985). In general, the criterion of *excellent* indicates a strong converging evidence within a given dimension (e.g., internal consistency $\geq$ .90, 2-week test-retest $\geq$ .80). The criterion of *adequate/good* indicates the evidence is acceptable for use of the measure, whereas *limited* is used to denote a marginal or unacceptable level of evidence.

## CHILDREN'S DEPRESSION INVENTORY

The *Children's Depression Inventory* (CDI, Kovacs, 1992) is a 27-item self-report instrument designed to assess the severity of depressive symptomology among school-aged children and adolescents, aged 6 - 17 years. The CDI item format consists of 3 statements about a particular depressive symptom of which the individual selects the statement that best summarizes his/her feelings for the past two weeks. The descriptive statements are rated on a 3-point scale (0=absence of symptom; 1=mild symptom; 2=definite symptom), with higher scores indicating increasing severity in depressive symptoms. The CDI items provide information on five subscales (i.e., negative mood, interpersonal problems, ineffectiveness, anhedonia, and negative

self-esteem) and a total score of depressive symptomology. The CDI item construction was modeled after the Beck Depression Inventory, a 21-item, self-rated symptom scale for adults. Research literature supports the use of an adult measure as a model in view of the overlap between the salient manifestations of depressive disorders in youths and adults (Kovacs & Beck, 1977). This measure of depression is purported as requiring the lowest reading level (1st grade level) among the available measures of depression for children (Berndt, Schwartz, Kaiser, 1983; Kazdin & Petti, 1982). The CDI was developed as: (a) a routine screening device in school-based and clinical settings; (b) an evaluative measure of treatment progress; (c) one component of a total diagnostic system of psychopathology; and (d) a research instrument for studying depression and related constructs.

**Content and Use**

The CDI provides an efficient and cost effective screening procedure for assessing depressive symptomology among children and adolescents. The brief item format enables the individual to complete the self-report measure in average time period of 15 minutes. The CDI can be orally administered to individuals with poor reading skills and may be group administered, although the author does not recommend CDI use in a group setting. The CDI provides two comparable self-report formats, long (27 items) and short (10 items), enabling the user to access either the form that will provide a more robust description of depressive symptoms or the form that will provide a quick screening in the periodic monitoring of a child's emotional status. Another beneficial aspect of the rating profiles is the direct incorporation of the norms into the profile forms, thus alleviating the need for tedious conversion of raw scores to standard scores.

**Standardization Sample and Norms**

The CDI standardization sample consisted of 1,463 Florida public school children in grades 2 through 8. The normative data is broken down by age and gender. Although the standardization sample consists of a large group of subjects, it demonstrates moderate limitations in the use of one geographic region for establishing normative data. The sample is adequately stratified by gender, age, socioeconomic status, race/ethnicity, and supplementary data on single parent versus intact households. The apparent lack of a truly representative sample is counterbalanced in the extensive research literature that utilized the CDI as the principle instrument. This strength of the CDI will be further discussed in the following section. Previous

versions of the unpublished CDI manuscript (Kovacs, 1983) provided additional score norms based on 75 psychiatric referrals and a sample of 860 Toronto public school children aged 8-15. In addition to the normative sample, the author of the instrument conducted additional research to investigate the psychometric properties of the CDI. One such study consisted of 134 adolescents diagnosed with one of the following: 44.8% major depressive disorder (mdd), 23.9% mdd on dysthymic, 17.2% dysthymia, and 14.2% adjustment disorder. This research further demonstrated the CDI as a valid and reliable measure of the severity of depressive symptomology among children.

**Scores and Interpretation**

The CDI administration results in a total raw score and five symptom-specific subscale scores, which summarize the child's tendencies with regard to depressive symptoms. The interpretative framework resides in the raw score conversion to percentile scores for each age/gender combinations (i.e., Boys: 7-12; 13-17: Girls: 7-12; 13-17) and to T-scores (mean=50; sd=10). Rationale for the use of different age level groupings is derived from two major developmental trends between the age of 12 and 13 years. Weiss et. al. (1991) argued for the appropriateness of this age split based on developmental theory which purports the two major developmental changes as being: (1) the cognitive shift from concrete operations to formal operations (Piaget, 1983); and (2) the common physical and emotional changes associated with puberty. Another component of the interpretative process is the use of critical cutoff scores based on the CDI total raw score. The manual suggests that a cutoff score of 11 be considered significant if the CDI was administered as a screening measure for depression, resulting in few false-negative errors. If the CDI was administered as an assessment of depression among children with emotional and behavioral problems, a cutoff score of 13 is suggested as an appropriate point for interpreting a significant level of reported depressive symptomology. In further assuring that the CDI be used and interpreted in a valid manner, the manual provides a "step-by-step interpretative sequence," developmental trends by gender, and interpretative cautions with respect to cultural effects.

**Psychometric Properties**

The psychometric properties of the CDI have been extensively researched and well documented by an impressive number of published research articles. The findings of broad

research efforts are clearly presented in the CDI manual and demonstrate the internal consistency and test-retest reliability as being excellent. Kovacs (1992) does allege that the CDI is a measure of state depression rather than trait depression. However, the relative stability of the CDI ratings over time causes one to question if the CDI is not a measure of trait depression as well.

The validity of the CDI was presented through the various methods of construct validity (predictive, divergent/convergent types) and content validity (i.e., concurrent and face validity). The CDI has been found to have significant relationships with a variety of self-report instruments (e.g., State-Trait Anxiety Inventory for Children, Reynolds Adolescent Depression Scale, Revised Children's Manifest Anxiety Scale, Coopersmith Self-Esteem Inventory, Piers-Harris Children's Self-Concept Scale) which measure depression or related constructs (e.g., anxiety, self-esteem). Children who score high on the CDI tend to report high levels of anxiety and low self-esteem. The latter constructs are theoretically and clinically related to depression, thus they support the CDI's validity (Kovacs, 1992). The CDI demonstrated an adequate level of discriminant validity in its ability to differentiate between normal and clinical samples. In contrast, some research maintains inconclusive evidence regarding the instrument's ability to discriminate among various diagnostic categories (Wendel, Nelson, Politano, Mayhall, & Finch, 1988). Content validity of the instrument was evidenced by several sources. Initially, the scale construction involved a careful selection of items as modeled after the Beck Depression Inventory (1967), a published adult self-report measure of depression. Review of the item selection was accomplished through various scale construction trials. The content validity of the CDI was further supported through adequate levels of item-total correlations.

**Summary**

The CDI is the most widely-researched child self-report measure of the internalizing domain (Merrell, 1994). The well-documented psychometric properties of the CDI evidence this measure to have adequate to excellent reliability and ample evidence of the scale's validity as a measure of depressive symptoms among children. The weight of evidence demonstrated through extensive research literature increases the utilitarian value of the CDI as a measure of the severity of depressive symptoms among children. Yet, the diagnostic precision in diagnosing depressive conditions is not perfect. This follows from theoretical and empirical evidence demonstrating that depressive symptoms are not unique to depressive disorders.

An impressive number of psychometric demonstrations exist in the literature that the CDI has adequate temporal stability, internal consistency, consistent correlations with other scales and various syndromes, and utility as a predictive measure.

A weakness of the CDI resides in the need for further research on its discriminant validity. The difficulty in conducting such research is the lack of standardized and replicable methods for the derivation of psychological diagnosis for child and adolescent disorders. The manual also states a need for longitudinal studies on the children identified as experiencing a depressive syndrome by the CDI and evaluating the natural progression through varying treatment methods. Further research is also needed to examine whether the CDI can reliably differentiate the construct of depression from other types of psychopathology (i.e., anxiety, adjustment disorder).

## REYNOLDS ADOLESCENT DEPRESSION SCALE

The *Reynolds Adolescent Depression Scale* (RADS, Reynolds, 1986) is a 30-item self-report instrument designed to assess the severity of depressive symptomology among adolescents aged 13-18. The RADS item selection was delineated from the DSM-III symptomology on depression and dysthymic disorders. Additional symptoms specified by the RDC (Research Diagnostic Criteria; Spitzer, Endicott, & Robins, 1978), as assessed by the SADS (Schedule for Affective Disorders & Schizophrenia; Endicott & Spitzer, 1978), were also utilized in the item construction of the RADS. The RADS consists of 30 symptom-related items and employs a 4-point Likert-type response format ("almost never" to "all the time") to assess the frequency of symptoms that are positive psychopathological signs of depressive disorder. The items are phrased to reflect either the presence of depressive symptoms (e.g., "I feel that no one cares about me.") or the absence of depressive symptoms (e.g., "I feel loved." ). The items are formulated in the present tense in order to elicit current symptom status among adolescents. The RADS was developed as: (a) a screening measure for the identification of depressive symptoms in school-based and clinical populations; (b) a research instrument for studying depression and related constructs; (c) an evaluative measure of treatment outcomes, because it has been shown to be sensitive to treatment outcomes (Reynolds & Coates, 1986); and (d) one part of a comprehensive battery of diagnostic instruments for general assessment of psychopathology.

**Content and Use**

The RADS provides a brief item format which enables the adolescent to complete the self-report measure with considerable ease in an average time period of 10 minutes. The RADS manual presents clear guidelines and instructions, for teachers and coordinators, regarding the various applications of the RADS. The manual also provides descriptive case studies to illustrate the clinical application and interpretation of the RADS in school and clinical settings. The RADS can be administered individually, in small groups, or in larger groups. Reynolds (1987) noted that the primary design purpose of the RADS involves its' use in large-scale screening of adolescents for depression in school settings. In addition, the RADS format allows for the items to be orally administered to individuals with poor reading skills. There are 2 forms of the RADS. A hand-scored version which is facilitated by the use of a scoring template. And an optical character recognition (OCR) answer sheet designed to be scored by an optical scanner machine. The publisher of the RADS provides a mail-in service for this purpose. The manual provides the user with valuable foundation knowledge on the assessment of depression, utilizing available statistical information, and constructs related to depression.

A strong feature of the RADS is that it employs several checkpoints in assuring the validity of the adolescent's response set on the RADS. The first check in considering the RADS protocol as a valid measure of depression involves the necessity for at least 24 of the 30 items to be completed. If one to six items are left blank, a rough estimate of the total RADS score can be obtained through a prorating technique described in the manual. In addition, the blank items should be examined for a possible reflection of other problems. Another validity check rests on the examination of unusual patterns of responding (e.g., same response to all items.). The RADS does contain some reverse-scored items which highlight such unusual response patterns as being inconsistent with valid responding. The inclusion of reverse-scored items requires that the individual be more attentive to each item and its associated response set. Finally, the RADS protocol includes two pairs of items that are opposites but logically very similar. Therefore, the adolescent's endorsement of the same score at extreme ends of the scale for the items within each pair would suggest an invalid protocol due to an inconsistency in responding.

The RADS also contains a number of items, "critical items", that have been designated as possessing the ability to discriminate between clinically depressed and nondepressed adolescents.

The occurrence of four or more critical items being endorsed by the adolescent should be cautiously viewed as significant, regardless of the RADS total score.

## Standardization Sample and Norms

The standardization sample consisted of 2,460 adolescents from grades 8 through 12. The large normative sample of the RADS was formed by one high school (grades 10-12) and two junior high schools (grades 7-9) in an urban/suburban community in a midwestern region of the United States. Although the standardization sample represents only one geographic region, it is well-stratified with regard to gender, age, race, and socioeconomic status. In addition, independent investigations conducted in other U.S. regions demonstrated similar results in regards to standard distributions and means, thus a geographically-stratified sample was not indicated. Although the standardization sample did not include special education students, a study by Reynolds and Miller (1985) demonstrated higher endorsement of depressive symptoms among educably mentally retarded students than among regular education students. In regards to the absence of normative data broken down by race, the lack of significant score differences between black and white subjects occurred during the initial formulation of the standardization sample, thus normative data stratified by race was not indicated.

## Scores and Interpretation

The interpretative framework of the RADS principally lies with raw score conversion to percentile ranks and T-scores. A cutoff score on the RADS (i.e., raw score at or above 77) can be utilized in judging the severity of depressive symptom endorsement and identifying the adolescent for further evaluation aimed at diagnosing significant psychopathology. The manual provides normative tables based on the total standardization sample and the standardization sample broken down by grade and sex. The comparisons with these normative data allow for more specific examinations of the significance of scores.

## Psychometric Properties

The manual presents internal consistency and test-retest reliability evidence for the RADS. Internal consistency is evidenced by the efforts of individual researchers across the U.S. as being excellent. Adequate levels of test-retest reliability, in addition to the excellent level of internal consistency, indicate the RADS as a reliable self-report measure of depressive symptomology among adolescents. The RADS stability was further evidenced in its' maintenance of an adequate

level of test-retest reliability at an administration at a one-year interval. Reynolds (1987) did note in the RADS manual that any assessment of test-retest reliability of depression measures must be cautiously interpreted due to the fluctuation tendency of some symptoms of depression, particularly mood-related symptoms. Some symptoms of depression may be influenced by external factors; such as, weather, situation specific interaction, or mild illnesses.

The validity of the RADS was presented through the following methods; content validity (i.e., concurrent and face validity) and construct validity (i.e., convergent and discriminant types). Content validity of the RADS was evidenced by several sources. These sources include item-total scale correlations which demonstrated the item consistency with the sum of all other items, and the concurrence of item content with clinically specified diagnostic and research symptoms of depression. In demonstrating the concurrent validity of the RADS, a clinical interview, stated as the "most sensitive methodology" (Hamilton, 1982, Puig-Antich & Gittelman, 1982) in the assessment of depression, was utilized in determination of presenting symptoms as specific to depression. This initial concurrent validity procedure used the Hamilton Depression Rating Scale on the basis of it as a similar measure of the severity of depressive symptoms and its frequent use with adolescent population in psychiatric studies of depression. The correlation between the scores on the RADS and the Hamilton Rating Scale provide a strong support for the concurrent validity of the RADS. Construct validity of the RADS was also evidenced in several ways. The RADS as a purported measure of depressive symptomology among adolescent populations was evidenced through studies of convergent validity, discriminant validity, factorial validity, and clinical efficacy (e.g., accurate categorization of adolescents into correct groupings, sensitivity to treatment manipulation).

**Summary**

The RADS is a brief and easy to use measure appropriate for screening adolescents for the occurrence and/or severity of depressive symptomology. The RADS has been evidenced through vast research efforts as being a highly reliable and valid instrument in the assessment of depressive symptomology among adolescents. The strong psychometric properties of the RADS demonstrate it as being a helpful tool for evaluation and appropriate intervention planning with adolescents who are experiencing depressive symptomology. Reynolds provides an extensive research and analyses, over a 6 year period, in support of the psychometric and clinical

characteristics of the RADS. The manual is clearly written and provides professionals with much more than simply a description and use of a self-report measure for depressive symptomology. The manual serves several purposes: to provide the psychometric properties and normative information; to provide foundation knowledge on depression (i.e., nature & evaluation); and to provide an overall contribution to the provision of mental health services to youth. In sum, an interdependent relationship is purported by Reynolds' RADS manual in that "the utility of a test is in part a function of the adequacy of the manual".

## REYNOLDS CHILD DEPRESSION SCALE

The *Reynolds Child Depression Scale* (RCDS, Reynolds, 1989) is a 30-item self-report instrument designed to assess the severity of depressive symptomology among children in grades 3-6 (ages 8 - 12). The RCDS item selection was principally based on the depressive symptomology delineated in the DSM-III and by the RDC (Research Diagnostic Criteria; Spitzer, Endicott, & Robins, 1978). The RCDS consists of 30 symptom-related items and employs a 4-point Likert-type response format ("almost never" to "all the time") to assess the frequency of symptoms that are positive psychopathological signs of depressive disorder. Similar to the RADS, the items on the RCDS are phrased to reflect either the presence of depressive symptoms (e.g., "I feel lonely.") or the absence of depressive symptoms (e.g., " I feel important."). In addition, the items were written to be consistent with a child's developmental experiences in the 8- to 12-year age range. Thus, the RCDS incorporates five pictorial representations of basic emotions (i.e., 5 faces depicting graduated emotional states from very sad to very happy). The child is asked to select the facial expression that best shows how they feel. The items are similar in format to the RADS, in that they are in present tense in order to elicit symptom status among children. The RCDS was developed as: (a) a screening measure for the identification of depressive symptoms in school-based and clinical populations; (b) a research instrument for studying depression and related constructs; (c) a reliable measure in evaluating treatment outcomes (Stark, Reynolds, & Kaslow, 1987); and (d) a component of a comprehensive battery of diagnostic instruments for general assessment of psychopathology.

## Content and Use

The RCDS brief item format allows for the child to complete the self-report measure in an average time period of 10-15 minutes. The RCDS construction is similar to the RADS in that it to can be administered individually, in group settings, and also orally administered to individuals with poor reading skills. Items were written at a second-grade reading level. It is advised that the group administrations do not exceed 20-30 children at a time. There are 2 forms of the RCDS. A hand-scored version which is facilitated by use of a scoring template, and an optical character recognition (OCR) answer sheet designed to be scored by an optical scanner machine. The publisher of the RCDS provides a mail-in service for this purpose. The manual presents clear guidelines and instructions for the administration, scoring and interpretation of the RCDS to be followed for reliable and valid administration and interpretation of the scale. The manual also provides descriptive case studies to illustrate the clinical application and interpretation of the RCDS in school and clinical settings. Foundation knowledge on the nature and assessment of depression is also presented in the manual.

The validity checks previously cited in the RADS section were also used in the development of the RCDS, with adaptations to children 12-years and younger. The RCDS incorporates reverse-scored items to identify invalid response patterns, and also includes logically similar but opposite item pairs, and the necessity for at least 24 out of 30 items to be completed. See the previous RADS "Contents & Use" section for in depth explanation of the validity checkpoints.

The RCDS also contains six items tentatively identified as "critical items". These items are similar to those on the RADS in that they are purported to demonstrate an ability to discriminate between clinically depressed and nondepressed children. Instances in which four or more of the "critical items" are endorsed at a critical level should be tentatively viewed with concern, regardless of the RCDS total score.

## Standardization Sample and Norms

The standardization sample consisted of 1,620 children from elementary schools in western and midwestern regions of the United States. The normative sample is well-stratified with regard to grade, gender, ethnicity, urbanism, and socioeconomic status. The sample can be considered to be racially heterogeneous and heterogeneous with respect to socioeconomic status.

Although the primary target grades of the RCDS are 3rd through 6th, several 2nd and 7th grades were included in the total standardization sample. The majority of the sample (95%) were in grades 3 through 6 and no separate norms were provided for the 2nd and 7th graders in the normative data tables. The limited geographic stratification of the RCDS standardization sample has been supported by various research efforts which suggest that a nationally representative norm group is not necessary. In general, available data has not evidenced depression as a characteristic which demonstrates geographic differences. In addition to the standardization sample, the author and several independent researchers collected descriptive data from several additional samples (e.g., separated/divorced families, nuclear families, depressed families). Data from such studies indicate that higher RCDS total scores are associated with increased risk status.

**Scores and Interpretation**

Interpretation of the RCDS total score is based on raw score conversions to percentile and critical raw score values. The manual presents of the raw score value of 74 as being the critical value for clinical significance, which has been demonstrated as possessing an exceptional accuracy in identifying children who meet additional criteria for depressive symptomology. The manual provides descriptive statistics for total standardization sample, and separately for male and female subgroups.

**Psychometric Properties**

The RCDS has been evaluated with regards to both internal consistency and test-retest reliability. Internal consistency of this scale has been evidenced at excellent levels. Adequate to excellent levels of test-retest reliability, in addition to the excellent level of internal consistency, indicate the RCDS is a reliable self-report instrument for assessing depressive symptomology among children. As stated by Reynolds (1987) in the RADS manual, any assessment of test-retest reliability of depression measures must be cautiously interpreted due to the fluctuation tendency of some symptoms of depression, particularly mood-related symptoms. Some symptoms of depression may be influenced by external factors; such as, weather, situation specific interaction, or mild illnesses. The standard error of measurement is also presented in the manual and is purported to lend support to the clinical use of this scale.

Validity studies of the RCDS have been ongoing since the initial field test in 1981. The validity evidence is presented through various content and construct validity methods. Content

validity of the RCDS was indicated through various sources. Item-total scale correlations were utilized as such a source and ultimately demonstrated the item consistency with the sum of all other items, and the concurrence of item content with clinically specified diagnostic and research symptoms of depression. In demonstrating the concurrent validity of the RCDS, the Children's Depression Rating Scale-Revised (CDRS-R) interviews, a semi-structured interview schedule modeled after the Hamilton Depression Rating Scale was utilized in identifying the symptoms as specific to depression. The results indicated a strong relationship between scores on the RCDS and the CDRS-R, thus supporting the concurrent validity of the RCDS as a measure of depression. Construct validity of the RCDS was demonstrated through studies of convergent and discriminant validity, factorial validity, and clinical efficacy (e.g., sensitivity to treatment manipulation). In examining the convergent validity, the RCDS demonstrated strong correlations with another depression measure. Further evidence for the convergent validity of the RCDS has been indicated through positive correlations with measures of related constructs, including self-esteem and anxiety. Discriminant validity of the RCDS was evidenced through extensive research efforts demonstrating low correlations with measures of academic achievement and cognitive ability. These results support additional research efforts which concluded that only a weak correlation exists between children's self-reported depression and academic achievement (McGee & Williams, 1988). However, such research findings do not preclude depression as having an impact on a child's school performance. Factor analytic investigations suggest the RCDS measures underlying dimensions of depression (i.e., cognitive, demoralization-despondency, somatic-vegetative, mood-anhedonia) and thus constitutes evidence for the validity of the RCDS as a measure of depression. The RCDS was also evidenced to be sensitive to treatment and experimental manipulation and accurately place individual's in correct categories (e.g., depressed vs. nondepressed) which are important predictors of clinical validity.

**Summary**

The RCDS has been utilized in several years of extensive research studies regarding its psychometric properties and characteristics as a measure of depressive symptomology among children. The vast research efforts have evidenced the psychometric properties of the RCDS at adequate to excellent levels, thus supporting this scale as a reliable and valid instrument in the assessment of depressive symptomology among children. The manual is clearly written and

provides a comprehensive overview of childhood depression, its' identification and treatment practices, and the normative information and psychometric characteristics for use in the interpretation of RCDS scores. Several researchers cite the RCDS as possessing the extremely desirable characteristic of being an appropriate measure of treatment outcome. In addition, the scale has demonstrated a beneficial level of accuracy in categorizing children into correct groupings (nonclinical vs. clinical depression). In general, the RCDS can be considered a valuable tool in the screening process of children who may be experiencing depressive symptoms.

## REVISED-CHILDREN'S MANIFEST ANXIETY SCALE

The *Revised-Children's Manifest Anxiety Scale* (RCMAS, Reynolds & Richmond, 1985), subtitled "What I Think and Feel", is a 37-item self-report instrument specifically designed to assess the level and nature of anxiety in children and adolescents, aged 6-19. The nature of anxiety assessed by the RCMAS is that of trait anxiety, the tendency to be anxious over settings and time. The 37 statements are responded to in a Yes/No manner and yield five scale scores (Total anxiety, Physiological anxiety, Worry/Oversensitivity, Social Concerns/Concentration, and Lie scale). The lie scale consists of 9-items that are socially desirable but almost never true (e.g., "I am always nice to everyone."). The RCMAS was developed as; (a) a screening measure for the identification of anxiety symptomology (i.e., trait anxiety) among school-aged children; (b) one component of a multiaxial assessment system of child/adolescent psychopathology; and (c) a research instrument for analyzing anxiety and related constructs.

**Content and Use**

The RCMAS brief item format allows for the child to complete the form with ease in an average time of 10-15 minutes. The self-report instrument can be administered individually or in a group setting. It is suggested that the group administration be used in the fundamental screening for anxiety; whereas the individual administration is recommended for use with young children and those children with reading problems and other learning disabilities. Items were written at a 1st grade reading level. The RCMAS manual is clearly written and extensive in its presentation of the psychometric properties and characteristics. Included in the comprehensive scale manual is additional preparatory information on the nature of anxiety and its measurement, as well as the reasoning behind the authors' stated position on the necessity for using objective measures of

anxiety with children and adolescents. The RCMAS scale provides a hand-scoring key to assist in scoring the individual's responses. The RCMAS provides the user with several appendices, which address the various normative populations (e.g., ethnic/sex combinations for blacks and whites) to further aid in the ease of interpretation.

The four subscales of the RCMAS further provide insight into the individual's responses. These scales provide information on several distinct areas of anxiety (physiological, cognitive, & behavioral manifestations): (1) the physiological manifestations of anxiety (i.e., Physiological Anxiety), (2) obsessive worrying (i.e., Worry/Oversensitivity), (3) distracting thoughts and certain fears that lead to difficulty in concentration and attention (i.e., Social Concerns & Concentration), and an additional scale which assesses the level of acquiescence or "faking" of responses on the part of the individual (i.e., Lie scale). A limitation of the RCMAS resides in the possibility of some children understanding the purpose of the scale, thus extreme low Total Anxiety scores (i.e., more than 2 standard deviations below the mean) may place some question on the child's accuracy of responding.

**Standardization Sample and Norms**

The standardization sample consisted of 4,972 cases that were comprised from 80 school districts representative of each geographical region in the United States. The standardization sample is well-stratified with regard to age, gender, socioeconomic status, urbanism, and ethnicity (i.e., black/white cultures). These data were initially documented in a study by Reynolds and Paget (1983). The developers of this instrument included a greater number of students attending special education classes (i.e., educably mentally retarded, learning disabled, intellectually gifted) than typically utilized in such projects, in their efforts to increase the probability of obtaining an adequately representative sample. Normative data is clearly delineated and presented in the RCMAS manual.

**Scores and Interpretation**

Interpretation of the RCMAS is based on raw score conversions to percentile ranks and scaled scores. Norms are presented for the total normative sample at one-year intervals and also for each ethnic/sex combination (i.e., black/white cultures). The broad norm presentation is used in this scale as no clear consensus is available regarding which norms are most appropriate. If one subsample was demonstrated as manifesting more anxious symptoms, the total sample norms

would be most appropriate. The three RCMAS subscales are provided to assist the clinician in developing a hypothesis about the child and their behavior and represent distinct areas in which anxiety is manifested. The brief structure of the scales should direct one to only view them as an aid in hypothesis building and cautiously interpreted. In addition to the anxiety scales, the RCMAS also contains a lie subscale which was designed to detect acquiescence, social desirability, or the deliberate faking of responses. Clinical significance can be interpreted when scores fall in the range outside of T-score>60. In general, anxiety-based interpretations of a child's scores should be cautiously used when both the Lie and Total Anxiety scores exceed the test mean by one standard deviation or more (i.e., Lie subscale scaled score>13; Total Anxiety T-score>60).

**Psychometric Properties**

The normative sample of 4,972 cases was utilized in establishing the reliability and validity of the RCMAS. The manual presents internal consistency and test-retest reliability evidence for the RCMAS. Internal consistency for the total anxiety score has been evidenced at an adequate level. The three subscale scores have been evaluated as possessing a somewhat weaker level of internal consistency as compared to that of the total score. Yet, the reliability scores are still surprisingly good for such brief scales. The RCMAS stability was further evidenced by adequate levels of test-retest reliability at intervals of 1 week, 5 weeks, and 9 months. The RCMAS manual also provides a discussion of the standard error of measurement, sometimes viewed as a more practical statistic than the reliability coefficient when interpreting an individual's test score.

The validity of the RCMAS was examined through the methods of construct validity (i.e., convergent and discriminant validity) and content validity (i.e., concurrent and face validity). The RCMAS has been utilized as a principal instrument in a number of correlational studies with other self-report instruments. The RCMAS has demonstrated strong correlations with the Trait scale of the State-Trait Anxiety Inventory for Children (STAIC, Speilberger, 1973), and low correlations with the State scale of the STAIC. Thus indicating that the RCMAS is best considered a measure of chronic manifest anxiety, independent of state or situational anxiety. In providing an evaluation of convergent and divergent validity of the RCMAS, a number of variables were selected as follows, traditional intelligence (i.e., Goodenough Harris Drawing Test IQ, Goodenough &

Harris, 1963) and personality measures, and behavioral assessment in the class setting. The correlation between the RCMAS scales and teachers' observations of behavior were indicated to be almost all positive. Discriminant validity of the RCMAS was evidenced through lower total scores of gifted vs. average children, and higher total scores of learning disabled vs. average children. These findings further support several researchers who have stated that internalizing symptoms have a strong negative effect on academic self-concept, which is the single most powerful predictor of academic success (Bloom, 1976; Merrell, 1994; Quay & La Greca,1989).

**Summary**

The psychometric properties of the RCMAS indicate adequate levels of validity and reliability across gender and racial groups (Reynolds & Paget, 1981). The three RCMAS subscales have evidenced questionable stability thus is recommended that the total score be used rather than the subscale scores for the majority of purposes (Merrell, 1994). Although somewhat limited in the amount of research available on the psychometric properties of the RCMAS, the existing literature does support the RCMAS as a measure of chronic anxiety among children and adolescents. Several aspects of the RCMAS enable the user to employ this scale with a certain level of confidence. The RCMAS consists of a large, well-stratified normative group, and a broad number of published studies available in support of the various uses and properties of this instrument. With the exception of the STAIC, few children's anxiety scales have as much reliability and validity evidence available for consideration. Given these strengths of the RCMAS, its face validity, and its ease of use, this instrument can be recommended as a useful part of a multiaxial assessment battery in the assessment of childhood internalizing disorders. In spite of the many beneficial characteristics of this scale, there is a need for continuing research.

In sum, the RCMAS possesses a utilitarian value in the personality assessment of elementary school children. This scale has the potential to facilitate further discussions regarding a child's endorsed/nonendorsed items and possible avenues for discussing the identified problems on the individual's RCMAS profile.

## STATE-TRAIT ANXIETY INVENTORY FOR CHILDREN

The *State-Trait Anxiety Inventory for Children* (STAIC, Speilberger, 1973) is a 20-item self-report instrument designed to assess individual differences in anxiety-proneness among

children ages 9-12. This inventory was developed as a downward extension the State-Trait Anxiety Inventory (Speilberger, Gorsuch, & Lushene, 1970), a self-report measure for adolescents and adults. Items are scored on a 3-point Likert-type scale with the higher scores reflecting stronger symptoms of anxiety. The STAIC consists of two scales, each consisting of 20 items, which assess the separate constructs of state anxiety (i.e., how anxious the child feels at the time the inventory is being completed) and trait anxiety (i.e., how anxious the child feels in general). On the Trait scale, the child is asked to decide whether statements are hardly ever, sometimes, or often true for them. On the State scale, the child is asked to report on how he/she feels at a particular moment in time. The instructions for the State scale may be modified to permit the evaluation of the level of state anxiety for any situation or time frame that is of particular interest to the clinician. The choices are presented along a continuum indicating the degree of anxiety, on both the State and Trait scales. The differentiation between state and trait anxiety is based on theoretical evidence as purported by Speilberger (1966,1972). Trait anxiety in children tends to be displayed through a stable response style in their perception of a variety of situations as threatening, while state anxiety in children is manifested by feelings of anxiousness due to specific situations or events. The State-Trait Anxiety Inventory for Children was primarily developed as a research tool in studying anxiety among elementary school children and in clinical settings. The STAIC can also be utilized as: (a) a screening instrument for both situational anxiety (i.e., state anxiety) and chronic manifest anxiety (i.e., trait anxiety); (b) one part of a comprehensive psychological assessment battery for general analysis of psychopathology.

**Content and Use**

The STAIC may be administered individually or in a group setting. The brief format allows for the STAIC to be readily completed by most children in an average time of 10 minutes for each scale. Both Trait and State scales are conveniently presented on opposite sides of a single hand-scored response form. A computer-scored answer sheet is also available for ease of administration and scoring in group settings. It is advised that such multiple choice answer sheets be employed with 5th and 6th graders to possess a level of confidence in the comprehension of its' purpose. The STAIC is presented in a brief 12 page manual which a clearly formatted for presenting the psychometric properties and characteristics of the scale. In addition, the STAIC

provides a scoring template for use in the quick scoring process (1-2 minutes) of the hand-scored version.

Validity checkpoints incorporated into the STAIC consist of some reverse-scored items which help to control for acquiescent responding. The instance of a respondent omitting one or two items, on either the State or Trait scale, does not invalidate the protocol and is compensated for through prorating procedures described in the user manual. To extend the valid application of this measure with non-English speaking children a Spanish version of the STAIC is also available.

**Standardization Sample and Norms**

Normative data was based on two large samples:(1) 913 children; (2) 638 children; consisting of elementary school children in grades 4 through 6. The combined samples were formed by six different schools from two districts within the Florida state school system. The total normative sample of 1,551 children is largely representative of African-American children and is provided in the manual with gender and grade level breakdowns as well as total sample norms. However, the STAIC was limited in regards to the extent of available data presenting specifics of the scale's stratification (e.g., socioeconomic status, urbanism, other ethnic cultures). The manual does present the scale's diversity with regards to black/white cultures and gender, yet the practice of deriving a sample from only one state establishes the utility of generalizing the normative data as being questionable. A limitation in the normative data is the fact that it is over twenty years old and thus an outdated normative sample.

**Scores and Interpretation**

The manual provides tables for which the STAIC raw scores may be converted to normalized standard scores (mean = 50; SD = 10) and percentile ranks. Normative data is presented separately by sex and grade. The manual provides two tables to efficiently ascertain the normalized T-scores and percentile ranks for the respective grade, gender, and state or trait scales for each respondent.

**Psychometric Properties**

The normative sample of 1,551 children was utilized in establishing the reliability and validity of the STAIC. The Trait-anxiety scale has been reported to have higher test-retest reliability in relation to the State-anxiety scale at a six week interval. These findings are consistent with the theoretical orientation of the STAIC. Internal consistency of the state and trait scales of

the STAIC have been evidenced at adequate levels. In general, the STAIC State and Trait scales demonstrate adequate levels of reliability. According to various researchers, the STAIC is still not evidenced as reliable in comparison to the adult version (STAI).

Validity for the STAIC has been provided in the manual with respect to construct (discriminant validity) and content validity (concurrent and face validity). Concurrent validity of the STAIC has been evidenced through a number of studies correlating the STAIC with other self-report instruments. In particular, the Trait scale has demonstrated adequate correlations with three anxiety scales, the Children's Manifest Anxiety Scale (CMAS), the RCMAS, and the General Anxiety Scale for Children (GASC). In contrast, the State scale has demonstrated a limited correlation with the RCMAS thus supporting the RCMAS as a measure of trait anxiety. STAIC correlations with non-anxiety scales have indicated this scale as lacking good predictive ability of either aptitude or achievement. Yet, recent experimental studies present a level of optimism for its' predictive value with respect to the Trait scale. The STAIC has demonstrated the ability to differentiate between anxiety-disordered and non-anxiety-disordered children (Hodges, 1990). In addition, this inventory has also demonstrated the ability to differentiate between anxiety levels among students with varying degrees of academic achievement problems (Rhone, 1986). In sum, these findings support the discriminant validity of the STAIC. Further evidence of the construct validity of the State scale has been demonstrated through the higher scores when the child is under stress inducing (test conditions) situations versus norm (standard instructional) conditions.

**Summary**

Psychometric properties as stated by the test manual and extensive research efforts have generally been adequate to good (Merrell, 1994). The theoretical foundation of the inventory has been considered as a definite advantage in previous reviews of the STAIC.

Limitations in the clinical utility of the STAIC reside in the limited availability of psychometric validation studies and a geographically limited normative population. There has been a shift in the severity of the first weakness as over 40 studies have been published between 1987 and 1992 with the STAIC as a primary measure. The secondary problem area for the STAIC continues to deflate the utility of this measure. Despite the fact that the norm sample is limited to one state, the norms are now approximately 20 years old. Increasing the confidence in

obtained test scores and ensuring the continued use of this test can be achieved through the establishing a current and nationally representative standardization of the STAIC.

Despite the limitations previously stated, various researchers and clinicians have purported this scale as being one of the best measures, with the exception of the RCMAS, in the assessment of anxiety among children. In general, strong aspects of the STAIC principally arise from its' careful development and the sound theoretical writings on which it is based. The STAIC also demonstrates adequate reliability and validity, and adequate normative data.

## YOUTH SELF-REPORT
### (Internalizing broad-band)

The *Youth Self-Report* (YSR, Achenbach, 1991) is a 119-item norm-referenced self-report instrument designed to assess the perceptions of 11- to 18-year old individuals regarding their own competencies and problems across situations. The 119-items are descriptive statements rated on a 3-point scale ("not true" to "very true"), of which the youth selects how much an item describes his/her feelings, behaviors, and interests for the past 6 months. This measure is comprised of three major scales: Internalizing, Externalizing, and Competence. In this review, only the common items from the YSR internalizing broad-band domain were used, to maintain constancy with the other self-report instruments under present review. The internalizing broad-band scale consists of three narrow-band scales (withdrawn, somatic complaints, anxious/depressed). These three types of syndrome scales have been purported by psychology professionals as the main symptoms within the broad domain of internalizing disorders. The YSR, 1991 Profile, was developed as: (a) a screening tool for the identification of children and youth who are emotionally/behaviorally at-risk; (b) one component of a multiaxial empirically-based, multi-source assessment battery for the designing of appropriate intervention strategies; and (c) research instrument for the examination of particular disorders that correspond to any of the eight symptom scales.

**Content and Use**

The YSR can be readily completed and done by most youths in a 15-minute period. The YSR format does allow for the items to be orally read to youths with poor reading skills (i.e., below 5th grade reading level), which would cause the administration time to be slightly

lengthened. The components of this self-report instrument include a clearly written comprehensive technical manual, an accompanying one page hand-scoring profile, and a supplemental computerized scoring system. The availability of a computer scoring system significantly increases the ease of the overall interpretation of rated items. The computer system provides a clear graphic display of all scales and how they relate to the clinical range of problem behaviors. The YSR is an empirically sound component in the Achenbach CBCL multiaxial assessment system. The components of the Achenbach assessment system [i.e., teacher report form (TRF), child behavior checklist (CBCL) completed by parents, and youth self-report (YSR)] facilitate a further understanding of the nature of internalizing problems through the comparison between different informant reports about the youth and the youth's behavior.

**Standardization Sample and Norms**

The standardization sample consisted of 1,315 youths from grades 6 through 12. The large normative sample of the YSR-1991 Profile was nationally representative. The standardization sample was well-stratified with regard to socioeconomic status, race, and urbanism. Normative data is broken down by gender and age.

**Scores and Interpretation**

The manual presents raw scores, standard T-scores, percentiles, and defined ranges of clinical significance of internalizing and externalizing symptoms (i.e., normal range, borderline, clinically significant). The internalizing score for each profile is the sum of the items on the three Internalizing scales of that profile. Each item is scored once for either the Internalizing or the Externalizing score. The scoring process does provide separate normative data for girls and boys. The author suggests the following guidelines for the interpretation of the clinical significance of syndrome scales, total problem score and internalizing grouping of syndromes, respectively. Clinical cutpoints of syndrome scales are reflected by a T-score of 70 or above and a borderline range assigned to T-scores values between 67 to 70. Borderline range is denoted by two broken lines on the profile which are not so clearly in the normal range. Clinical cutpoints of the total problem and internalizing grouping of syndromes are reflected by a T-score of 60 or greater and a borderline range assigned to T-score values between 60 to 63. The rationale for the use of lower clinical cutpoints in the latter two areas was presented in the manual as a way to minimize the

percent of nonreferred youths who scored above the cutpoint ("false positive") in addition to the percent of referred youths who scored below the cutpoint ("false negative").

**Psychometric Properties**

The manual presents internal consistency and test-retest reliability evidence for the two broad-band scales and respective narrow-band scales. As noted previously, this review will focus on the internalizing broad-band scale and thus the presented statistics will reflect this fact. Internal consistency of the internalizing scale is evidenced by vast research efforts as being at an excellent level. Adequate levels of test-retest reliability also indicate the YSR-internalizing broad band as a reliable self-report measure of internalizing symptoms among youth. These reliability coefficients possess some limitations, in that the cited 7-month interval results did not include the 15- to 18-year old youths in the general population utilized for that study. Another study utilizing a clinical sample did employ the entire population addressed by the YSR in a 6-month interval study. This 6-month interval analysis ultimately demonstrated an adequate level of test-retest reliability.

The validity of the YSR was investigated in several ways. Content validity was supported by the measure's ability to significantly differentiate between demographically matched referred and nonreferred youths. With the demographic effects partialled out, criterion-related validity of the YSR scale scores are evidenced through discrimination between referred and nonreferred youths. The evaluation of the YSR construct validity is limited by the current lack of similar instruments. At this time, correlations between the YSR, CBCL, and TRF are evidenced through a variety of analyses. Yet, this method of testing construct validity is limited by the different perspectives of adolescents, their parents, and teachers.

**Summary**

The YSR is a unique measure as it is one of only two self-report measures, the other being the recently developed Behavioral Assessment System for Children (BASC, Kamphaus & Reynolds, 1992), addressing two broad constructs of symptoms. While the other self-reports in this review assess only one disorder (i.e., depression, anxiety). The YSR does possess some limitations in its efficacy as a self-report measure in assessing internalizing and externalizing problems among children and youth. One of those limitations resides in the fact that the YSR includes a set of socially desirable items in order to provide a balance to the problem items.

However, these socially desirable items are not included in the overall interpretation, thus further highlighting the fact that the YSR does not possess any validity scales. Such validity scales may enable the user to detect a deviant response set or manipulation on the part of the youth. Another disadvantage of the YSR resides in the necessity for all items to be completed in order to achieve a valid interpretation of the data.

Although some problems do exist, the YSR demonstrates a number of positive reasons for selecting this measure for use in an assessment of emotional and behavioral problems among youth. Some of its strengths include a comprehensive manual with regard to broad technical information and practical applications, an easy-to-use response format, and a rationally and empirically defined relation between the YSR and the components of the CBCL assessment system.

## BEHAVIORAL ASSESSMENT SYSTEM FOR CHILDREN
### (Internalizing self-report form, scales)

The *Behavioral Assessment System for Children* (BASC, Kamphaus & Reynolds, 1992) provides a comprehensive assessment system in evaluating behavioral and self-perceptions among children, aged 4-18 years. The system involves a multimethod, multidimensional approach to achieving an integrated understanding of the child's behavior in a variety of settings and with a number of different viewpoints. The integrated approach is achieved through the various components of the BASC: teacher and parent rating scales, structured developmental history form, student observation system, and child and adolescent self-reports of personality. In this review, the self-report of personality for children and adolescents will be the principal focus as it is the scale that correlates with the other self-report instruments under present review. The self-report of personality measures the feelings and self-perceptions of children and adolescents rather than asking them to describe their own behavior. For this analysis, only the internalizing scales comprising the Emotional Symptoms Index (ESI) will be used as it is purported to be the SRP's global indicator of serious emotional disturbance, specifically internalizing disorders. The internalizing scales included in the ESI are anxiety, social stress, depression, interpersonal relations, self-esteem, and a sense of inadequacy, which connote internal feelings of emotional upset. The ESI evaluates both negative or clinical scales and positive or adaptive scales which

involves reverse-scoring. The child self-report form (SRP-C) is a 152-item, 12-scale, inventory of behavior and self-perceptions among children aged 8-11 years. The adolescent self-report form (SRP-A) is a 186-item, 14-scale, inventory of behavior and self-perceptions among individuals aged 12-18 years. Both the SRP-C and the SRP-A are descriptive statements rated on a true/false format. The true/false response format differs from the teacher and parent rating scales, which employ a 4-point rating system. The authors' choice to use a simpler response format reflects their prime consideration for readability and comprehensibility of SRP items for those children with poor reading skills, limited attention span, and less developed intellectual abilities than adults. The BASC was primarily designed to facilitate differential diagnosis and educational classification of a variety of emotional and behavioral disorders among children to aid in the design of appropriate treatment programs. The BASC was also developed as: (a) an evaluative measure of treatment outcomes; and (b) a research system for studying childhood psychopathology and behavior disorders.

**Content and Use**

The BASC can be readily completed by most children in an average time of 30-minutes. The components of this self-report instrument include a comprehensive manual, a one page hand-scoring profile, and a computer form which allows the respondent to key in item responses in about 5-minutes. The manual provides practical guidelines for the administration and scoring of all BASC forms, a multitude of norm tables broken down by gender, age, clinical and general samples, and a discussion of the appropriate applications of the BASC. The convenient scoring system of the BASC allows for rapid scoring without cumbersome scoring templates or keys. Each hand-scored form is comprised of two parts which only need to be separated to reveal an inner page with the items already scored. The inner page also includes a summary table and graphical profile. In addition, to the materials for the "normal" population, the BASC also provides supplementary materials for children with poor reading ability (i.e., audiotaped items for dictation purposes) or Spanish-speaking children (i.e., translated versions of the BASC rating scales and self-report forms).

**Standardization Sample and Norms**

The standardization sample consisted of 9,861 children from preschool through 12th grade. The notably large normative sample of the BASC was nationally representative and

stratified with respect to geographic region, socioeconomic status, race/ethnicity, gender, clinical status of the child, educational placement (i.e., special or regular education), and parental education (parent rating scale norms only).

**Scores and Interpretation**

The interpretative framework of the BASC principally lies with raw score conversions to T scores (mean=50; SD=10) and percentile ranks. T scores are utilized in judging the severity of internalizing symptom endorsement and identifying the child for further evaluation aimed at specific diagnosis of psychopathology. T scores of 70 or above on the clinical scales and 30 or more on the adaptive scales are considered as clinically significant. On the Emotional Symptoms Index (ESI), T scores of 65 or above demonstrate clear and pervasive distress, while a T score of 70 or higher clearly suggest a serious emotional disturbance of some form, if the response pattern is valid. On the Emotional Symptoms Index, a T score below 40 in a referred case is likely to represent denial or "faking good" (i.e., faking responses in a socially desirable manner). In gathering a further understanding of the presenting emotional disturbance, the other SRP scales that do not contribute to the ESI should be utilized in examining the nature of the emotional disturbance. The available normative data allows for further comparison of scores in assessing their significance as related to the individual's demographic status.

**Psychometric Properties**

The BASC manual provides internal consistency and test-retest reliability evidence for the composite and scales. Excellent levels of internal consistency and test-retest reliability indicate the BASC as a reliable assessment system of behavior and self-perceptions among children and adolescents. The BASC stability was further evidenced in its' maintenance of an adequate level of test-retest reliability at a 7-month interval, with regards to the Emotional Symptoms Index (ESI), the adolescent (SRP-A) and child (SRP-C) forms.

The validity of the BASC was presented through the following methods; content validity (i.e., concurrent and face), and construct validity (i.e., factor analysis, discriminant, and convergent types). Content validity of the BASC was evidenced through the efforts of 14 licensed clinical psychologists. The psychologists sorted SRP-A items into discrete categories that represented various forms of adolescent psychopathology. This alternative set of scales was a reflection of a purely content-based grouping of SRP-A items which was compared to the

published scales of the SRP-A. The procedure demonstrated the ability of the items to differentiate among clinical and nonclinical samples. The construct validity of the SRP was demonstrated through the correlation of SRP composites and scales with four published instruments; the Minnesota Multiphasic Personality Inventory (MMPI), the Youth Self-Report (YSR), the Behavior Rating Profile (BRP), and the Children's Personality Questionnaire (CPQ). Three of the selected instruments do purport to measure the same underlying construct, internalizing disorders, which is the basis of this review. Thus, the correlational study will also provide information on the concurrent validity of the SRP. For our purposes, the ESI was the focal point of correlation research findings. The Emotional Symptoms Index was evidenced to correlate fairly high with the BRP, the MMPI, and the internalizing broad-band of the YSR, but as being distinctly different from the CPQ, which tends to reflect normal dimensions of personality or temperament. Further evidence of the construct validity of the SRP was evidenced through factor analysis. The factor structure was demonstrated as being virtually identical at the two age levels, despite minor differences in item content and two additional scales at the adolescent level. In sum, the results support the concurrent and construct validity of the SRP. The validity of the BASC is a strong point in that its development emphasized the content and construct validity of the scales so as to formulate an assessment system with readily interpretable scores.

A limitation of the BASC resides in the manual's divergence from ease of use and clarity in the technical information chapter. This chapter on reliability and validity assumes the reader is well-versed in statistical terms (e.g., eigenvalues) and methods (e.g., principal-axis analysis). However, the BASC does offer various types of validity checks which allow the clinician to assess the consistency of informant reports.

**Summary**

The BASC is a unique and reliable assessment system of behavior and self-perceptions among children and adolescents. The three available forms (i.e., teacher & parent rating scales, self-report of personality) provide an integrated assessment system on the child's behavior in a variety of settings. Such an integrated assessment system allows for comparison and highlighting of problems that are at a clinically severe level. The BASC not only evaluates behavior problems and emotional disturbance, but it also identifies the child's positive attributes which can be capitalized on in the treatment process. It is an instrument that is highly interpretable due to its

development around specific constructs and has the potential for providing relevant information to federal regulations regarding diagnosis of severe emotional disturbance in the schools. Extensive research supports the newly developed Behavioral Assessment System for Children as a reliable and valid measure of behavior and emotional symptoms among children and adolescents. The validity of the BASC is a strong point in that its development emphasized the content and construct validity of the scales so as to formulate an assessment system with readily interpretable scores. The broad (i.e., across general, clinical, and gender populations) and geographically representative normative sample of the U.S. in addition to the available normative data also increases the applicability of this measure to a wide population of children. In general, the individual BASC components are reliable and psychometrically sophisticated instrument that provides an array of beneficial data about the complex nature of childhood problems.

## DISCUSSION

The process of conducting critical reviews and comparative evaluations of the seven self-report measures has presented the existence of many similarities in scale construction as well as the technical properties and characteristics. This analysis has also uncovered the existence of significant differences among the measures. The measures possess a common agenda to assessing the symptomology of internalizing disorders and at the same time demonstrate variations in which they attempt to achieve their objectives. The various similarities and distinctions among the seven self-report measures are most effectively presented through the considerations of the construct of internalizing symptoms, appropriate use, user-friendliness, and technical properties.

**Construct of Internalizing Type Symptoms**

The construct of internalizing disorders comprises a variety of symptoms that are specific to child and adolescent populations and generalizable to adult populations. Internalizing disorders, commonly referred to as "emotional problems", include such problems as depression, anxiety, social withdrawal, somatic complaints, and low self-esteem. The other side of this classification dichotomy is that of externalizing disorders (e.g., conduct disorder, attention-deficit hyperactivity disorder), which involve those behaviors considered as "undercontrolled", in contrast to the "overcontrolled" nature of internalizing disorders. Increasing interest has also developed around the question as to whether anxiety and depression represent two distinct states

or rather a broad-band construct termed as "negative affectivity" (Hodges, 1990). The broad construct of internalizing disorders frequently involves an element of comorbidity among internalizing problems thus further complicating the effective identification and treatment of such disorders. The coexistence of internalizing problems reinforces the necessity of conducting a multiaxial, multimethod assessment in order to effectively gauge the depth and severity of internalizing symptoms experienced by the child or adolescent. Self-report measures have the potential to play a critical role in the assessment of internalizing symptoms. Specifically, the use of self-reports in school-based settings may be the only method by which the level of psychological distress will be identified for most children and adolescents. Thus, essential for school professionals and other mental health providers to be proficient in assessment of internalizing disorders in children and adolescents. Failure to identify and treat the child or adolescent in distress has the potential to prolong the suffering and even life-threatening consequences for the individual.

Various perspectives regarding the construct of internalizing symptoms are evident in the varying item content within self-report instruments. For instance, the majority of self-report measures not only sample the negative emotional and behavioral symptomology in their item pool, but the seven self-report measures under review also incorporate positive statements that are inconsistent with the manifestations of internalizing disorders. In some instruments, the positive statements serve to not only demonstrate the child as not experiencing emotional distress but also provide information regarding the child's tendency to respond in socially desirable manner by endorsing the extreme positive statements (e.g., I'm always nice to everyone."). With respect to the most effective type of response format to be utilized in self-report measures, two primary perspectives have been evidence by the self-report measures in this review. For instance, the RCMAS and BASC utilize a true/false response format which assesses the absence or presence of internalizing symptoms in general terms. In contrast, the other five self-report measures under review (CDI, RCDS, RADS, YSR, STAIC) go a step further by assessing the depth and severity of the internalizing symptomology. Such valuable illustrative data regarding the attributes of the child's emotional distress is obtained by a simple incorporation of a response format involving a forced choice among descriptive statements. Thus, the child is allowed to

evaluate their emotional status along a continuum (e.g., "almost never" to "all the time") as related to a descriptive statement or by selecting the statement that best describes themselves.

**Appropriate Use**

The authors of the seven self-report instruments assert that their measures have potential usefulness in the screening and identification process of children and adolescents experiencing emotional distress. The seven reviewed instruments are comprised of objective and specific emotional/behavioral items which are descriptive of cognitive, psychomotor, somatic, and interpersonal problem areas. The majority of self-report measures have been developed for use with "normal" school populations, with a few being applicable to specific clinical populations. Consequently, clinicians and researchers should proceed cautiously when selecting and using an instrument which may be limited in its' generalizability to special populations encountered in clinical/research settings. In general, the use of self-report measures as the principal method in formulating a diagnosis of internalizing disorders is deemed as being inappropriate by authors and researchers of self-report instruments. Valid diagnosis is achieved through a multisource, multimethod assessment battery which may and should include self-report data from the child/adolescent. Such comprehensive assessment practices has gained increasing importance as a result of various research and clinical studies demonstrating the existence of significant associations between internalizing disorders and other problems (poor academic achievement, substance abuse, behavior/conduct disorders, truancy, anorexia) (Reynolds, 1984). Thus, the clinical utility of self-report measures is further reinforced for psychologists and other mental health professionals in routinely screening individuals referred for the above stated problems in addition to screening for emotional distress. The practice of differential diagnosis is becoming an increasingly important issue in school settings due to the complexity of many childhood problems. The complexity of childhood problems require an array of interventions tailored to individual needs. Research supports the use of self-report instruments as a cost effective and readily analyzable method in assessing the depth and severity of internalizing symptomology. Information obtained from self-reports can provide valuable insight into the inner world (i.e., thoughts, feelings, perceptions) of the child/adolescent, an area that is not readily observable by others. In general, self-reports provide school-aged children with possibly the only vehicle for expressing

their distress, thus ending their silent cry for help through prompt identification and treatment by mental health professionals.

**User Friendliness**

In determining the user friendliness of the measures several factors were considered; administration time, types of scores, and clarity of the manual. The average time necessary for completing the measures involved a 10-15 minute period, with the exception of the BASC which requires about 30 minutes for the administration process due to the large number of items in each scale. A level of consistency was evidenced among the measures' in the ease of administration and scoring process, an characteristic purported to be a unique asset of self-report measures. With the exception of the RCMAS, all measures provide an alternative computer-scored response form thus further facilitating the ease of overall administration and scoring process. None of the measures were evaluated to present a significant level of difficulty in the process of learning the administration procedures. The BASC has the potential for requiring more effort and time on the part of the clinician if the multiple components of the assessment system (i.e., teacher & parent rating scales, structured developmental history form, student observation system) were to be utilized in addition to the child or adolescent self-report forms.

Interpretative data varied among the seven norm referenced instruments. At the basic level, the seven self-report measures in this review employed normalized standard scores, commonly referred to as T-scores, and percentile ranks in their raw score conversions to interpretable data. Standard scores are useful statistics employed by self-reports as they allow for interpretation in relation to peers and other instruments. Critical cutoff scores are also provided in relation to the T-score values further delineating the clinical significance of such values in relation to peers of the same age and sex. The YSR may present some difficulty in the user's initial attempts at interpreting the significance of an individual's scores due to the varying critical T-score values stated for the syndrome scale versus the total problem and internalizing groupings of syndromes. Although the additional clinical cutpoints may pose some initial apprehension as to which range of numbers to adhere to, the rationale behind the use of two different sets of clinical cutpoint values is clearly defended in the YSR manual as a way to minimize "false positives" and "false negatives".

An integral part in an instrument's level of user friendliness and comprehensibility is that of the manual. The user's manual is the principle and initial source of information regarding the development, psychometric properties and characteristics, and applications of the instrument. Through the review and comparison of user manuals, the most comprehensible were the Reynolds Child Depression Scale (RCDS, Reynolds, 1989) and the Reynolds Adolescent Depression Scale (RADS, Reynolds, 1986). The RCDS and RADS manuals provide a complete presentation of the development, technical information, and respective areas of focus (child & adolescent depression) within each measure. The other self-report measures reviewed also presented a basically sufficient level of comprehensibility with regards to their discussion of psychometric properties and characteristics, scale development, and applications. An aspect of the RADS and RCDS which enhanced its' overall comprehensibility resides in the employment of case studies to further illustrate their clinical applications in school and clinical settings. With regards to psychometric properties, the CDI was reviewed as being the most comprehensive measure of depressive symptomology among children and adolescents. The psychometric soundness of the CDI is supported by an impressive amount of research. The vast research findings of the CDI were clearly presented in the user manual. The RCMAS, BASC, STAIC, and YSR were also reviewed to provide clearly written manuals that possessed utilitarian value. However, the STAIC manual was lacking in the inconsistent presentation of narrative and tables when discussing technical information. The YSR-1991 Profile manual clearly presented the differences between the original Youth Self-Report and the 1991 Profile and rationale for development of the 1991 version.

**Technical Properties**

Technical properties were evaluated with respect to standardization samples, validity, and reliability. A scale may be aesthetically pleasing because of the ease of use, attractive protocols, or familiarity of the instrument. A self-report instrument must also possess sound psychometric properties to be valid and effective in assessing the severity and depth of internalizing symptoms.

Evaluation of standardization samples is accomplished through the consideration of several factors: sample size, regional representation, and demographics. Large, representative, and well-stratified samples were evidenced by the RADS, RCDS, YSR, and BASC. The RCMAS was also based on a large, well-stratified sample. The RCMAS was unique in its' inclusion of an atypically large number of special education students serving to further strengthen the

representational quality of the sample. With the exception of the STAIC and RCMAS, the other scales included diverse ethnic populations (e.g., Hispanic, Asian, American Indian) in addition to the standard inclusion of black/white ethnicity. This provided an increased probability to achieving an adequately representative sample. The CDI and STAIC were based on large well-stratified samples, however they are limited in regional representation by the authors' employment of only one state in forming the standardization sample. The STAIC presents another limitation in its' normative data which was collected over 20 years ago and thus a geographically outdated normative sample. Although the STAIC possesses outdated normative data, this scale in addition to the RCMAS are the only anxiety measures with have as much available reliability and validity data.

With respect to the reviewed instruments' validity, evidence was provided for the instruments through content, construct, and criterion-related validity. These three areas comprise a complete analysis of validity. Adequate levels of content validity was evidenced by all instruments in this review. In general, content validity was supported by several sources: face validity of item selection; and item construction based on careful selection and review of emotional, cognitive, and/or behavioral descriptors by experts. The CDI, RADS, and RCDS also demonstrate adequate to strong item-total correlations thus further reinforcing the content validity of the measures. Item-total correlation data was not provided for the YSR, RCMAS, or BASC. The STAIC does possess a definite advantage with respect to the theoretical foundation of this inventory. In addition, the STAIC and the CDI were downward extensions of the adult versions, STAI and Beck Depression Inventory, respectively, which provide established item pools that are representative of the specific internalizing domains. Extensive research efforts have evidenced the CDI, RADS, RCDS, RCMAS, STAIC, and BASC as valid measures of their respective constructs or domains (e.g., depression, anxiety, internalizing symptoms). Inclusive in such research are the findings that these measures demonstrate adequate to strong correlations with similar instruments and low correlations with instruments that assess unrelated constructs. In addition, all measures in this review possessed the ability to differentiate between clinical and nonclinical samples of children. The YSR has been evidenced as a valid measure through the methods of content and criterion-related validity studies. However, the unique nature of the YSR limits the ability to test its' construct validity due to the lack of similar instruments. The

YSR has been correlated with the other components of the Achenbach assessment system (i.e., TRF, CBCL), yet again facing limitations in this attempt to demonstrate construct validity by the different perspectives of the adolescent, their parents and teachers.

## CONCLUDING REMARKS

In conclusion, the various published self-report measures evaluated through this critical review were evidenced to be reliable and valid measures of their respective internalizing syndromes (i.e., depression, anxiety) and of the breadth of internalizing symptoms as assessed through the YSR and BASC. The seven self-report measures also demonstrated clinical efficacy in their ability to differentiate between clinical and nonclinical samples. In recent years there has been a considerable increase in the availability of psychometrically sound self-report measures for children. This trend is a reaction to concerns by mental health professionals with the individual differences in emotional, social, and psychological development of a child.

Further research is necessary in studying individual characteristics that may be predictive of treatment outcome and behavioral change. In addition, limited data is available on the effects for child/adolescent in answering questions about anxiety, depression, and other internalizing symptoms. Presently, a single instrument, the Internalizing Symptoms Scale for Children (Merrell & Walters, 1996), has been recently designed to measure the multiple facets of dysfunction within the broad domain of internalizing disorders. With the increasing consensus among clinicians and researchers regarding the significant associations evidenced among various internalizing disorders, specifically between depression and anxiety, there exists a need for further development of comprehensive self-report measures of internalizing symptoms of children and adolescents.

With regards to internalizing symptoms, the covert nature prompts clinicians and researchers to seek reliable, accurate, and developmentally appropriate methods for obtaining information from children and adolescents. The seven self-report measures under this review have all demonstrated their potential utility in the assessment of internalizing symptoms among school-aged children.

Table 1
**Selected Characteristics of the Reviewed Self-Report Instruments Used to Assess Internalizing Symptoms**

| Test Name | Items | Grade Level and Normative Sample | Forms | Interpretive Profile | Response Format |
|---|---|---|---|---|---|
| **Children's Depression Inventory** (CDI, Kovacs, 1992) | 27 10-short form | Grades 1-12 (Ages 6-17 yrs) Normative sample (n=1,463)   Regular education (100%)   Special education (0%) Race/ethnicity (approx.)   White 77%   Black/Hispanic/Am.Indian 23% Male (48%) Female (52%) | 2 forms 1)hand-scored 2) computer-scored | **1 Major Scale** Depressive Symptomology **5 Factor Subscales** Negative Mood Interpersonal Problems Ineffectiveness Anhedonia Negative Self-esteem | Forced choice among 3 descriptive statements   0=absence of symptom   1=mild symptom   2=definite symptom |
| **Reynolds Adolescent Depression Scale** (RADS, Reynolds, 1986) | 30 | Grades 7-12 (Ages 13-18 yrs) Normative sample (n=2,460)   Regular education: 100%   Special education: 0% Race/ethnicity (approx.)   White 76%   Black 20%   Hispanic (incl. in "Other")   Other 4% Male (50.2%) Female (49.8%) | 2 forms 1)hand-scored 2)optical character recognition (OCR) | Severity/Frequency of Depressive Symptomology | 4-pt likert-type response format ("almost never" to "all the time") |
| **Reynolds Child Depression Scale** (RCDS, Reynolds, 1989) | 30 | Grades 3-6 (Ages 8-12 yrs) Normative sample (n=1,620)   Regular education (100%)   Special education (0%) Race/ethnicity (approx.)   White 71%   Black 18%   Hispanic 5%   Asian 4% 2% did not indicate ethnic status Male (47%) Female (57%) | 2 forms 1)hand-scored 2) OCR | Severity/Frequency of Depressive Symptomology | 4-pt response scale ("almost never" to "all the time") |

(Table 1 continues...)

Table 1
## Selected Characteristics of the Reviewed Self-Report Instruments Used to Assess Internalizing Symptoms

| Test Name | Items | Grade Level and Normative Sample | Forms | Interpretive Profile | Response Format |
|---|---|---|---|---|---|
| **Revised Children's Manifest Anxiety Scale** (RCMAS, Reynolds & Richmond, 1985) | 37 | Grades 1-12 (Ages 6-19 yrs) Normative sample (n=4,972) Regular education (88%) Special education (12%) Race/ethnicity (approx.) White 88% Black 12% Hispanic 0% Other 0% Male (50.3%) Female (49.7%) | 1 form 1)hand-scored | General Anxiety **4 Subscales** Physiological Anxiety Worry/Oversensitivity Social Concern & Concentration Lie Scale | Yes/No statements |
| **State-Trait Anxiety Inventory for Children** (STAIC, Speilberger, 1973) | 20 per scale | Grades 4-6 (Ages 9-12 yrs) Normative sample (n=1,551) Regular education (100%) Special education (0%) Race/ethnicity (approx.) White 65% Black 35% Hispanic 0% Other 0% Male (48%) Female (52%) | 1 form 1) hand-scored 2) computer-scored | **2 major scales** State Anxiety Trait Anxiety | Forced choice among 3 descriptive statements |
| **Youth Self-Report** (Internalizing + Externalizing Scales) (YSR, Achenbach & Edelbrock, 1987) | 119 *103-problem behavior *16-socially desirable | Grades 6-12 (Ages 11-18 yrs) Normative sample (n=1,315) Regular education (100%) Special education (0%) Race/ethnicity (approx.) White 72% Black 16% Hispanic 8% Other 4% Male (49%) Female (52%) | 2 forms male/female 1) hand-scored 2) computer-scored | **2 broad band scales** Internalizing Externalizing **3 Internalizing subscales** Withdrawn Somatic Complaints Anxious/Depressed **2 Externalizing subscales** Aggressive Behavior Delinquent Behavior **1 Competency Scale** Activity & Social | 3-point response format ("not true" to "very true") |

(Table 1 continues...)

**Table 1**

**Selected Characteristics of the Reviewed Self-Report Instruments Used to Assess Internalizing Symptoms**

| Test Name | Items | Grade Level and Normative Sample | Forms | Interpretive Profile | Response Format |
|---|---|---|---|---|---|
| **Behavioral Assessment System for Children** (self-report of personality form, internalizing+ externalizing symptoms, scales, Kamphaus & Reynolds, 1992) SRP-C: child form SRP-A: adolescent form | 152(SRP-C) 186(SRP-A) | Grades: Preschool - 12 (Ages 4-18 yrs) Normative sample (n=9,861) Regular education: 87.9% Special education: 12.1% Race/ethnicity (approx.) White 70% Black 16% Hispanic 11% Other 3% Male ( 48%) Female ( 52%) | 2 forms 1)hand-scored 2)computer-scored | **5 Composites & 14 scales** **Clinical maladjustment** Anxiety Atypicality Locus of control Social stress Somatization (only SRP-A) **School Maladjustment** Attitude to school Attitude to teachers Sensation seeking (SRP-A) **Other Problems** Depression Sense of inadequacy **Personal Adjustment** Relations with parents Interpersonal relations Self-esteem Self-reliance **Emotional Symptoms Index** (i.e., Anxiety, Social stress, Depression, Self-esteem, Interpersonal relations, Sense of inadequacy) | True/False statements |

Table 2

Reliability Information for Self-Report Measures Developed to Assess Internalizing

Symptoms

| Test Name | Internal Consistency | Test-Retest |
|---|---|---|
| Children's Depression Inventory (CDI, Kovacs, 1992) | .71 -.89 | .82 (2 wks) .66 (4 wks) .67 (6 wks) |
| Reynolds Adolescent Depression Scale (RADS, Reynolds, 1986) | .90-.95 | .80 (6 wks) .79 (3 mos) .63 (1 yr) |
| Reynolds Child Depression Scale (RCDS, Reynolds, 1989) | .90 | .82 (2 wks) .81 - .92 (4 wks) |
| Revised Children's Manifest Anxiety Scale (RCMAS, Reynolds & Richmond, 1985) | Total Score .79 (males) .85 (females) Subscales .50-.70 (anxiety scales) .70-.90 (lie scale) | .88 (1 wk) .77 (5 wks) .68 (9 mos.) |
| State-Trait Anxiety Inventory for Children (STAIC, Speilberger, 1973) | .78-.81 [a] .82-87 [b] | .65 - .71 [a] (6 wks) .31 - .41 [b] (6 wks) |
| Youth Self-Report (internalizing broad band) (YSR, Achenbach & Edelbrock, 1987) | .89 | .80 (1 wk) .69 (6 mos) [a] .52 (7 mos) [b] |
| Behavioral Assessment System for Children (internalizing self-report form, scales) (BASC, Kamphaus & Reynolds, 1992) | .87-.96 (Ages 8-18 yrs) .95 [c] | .81 (1 mo): SRP-A&C .55 (7 mos): SRP-C .84 (1 mo) [a] .60 (7 mos) [b] |

[a] STAIC- Trait scale

[b] STAIC- State scale

[a] YSR-Total Problem reliability for clinical population (Ages 12-17 yrs).

[b] YSR- Internalizing Scale reliability for general population (Ages 11-14 yrs).

[a] BASC- Emotional Symptoms Index reliability for SRP-A & SRP-C.

[b] BASC- Emotional Symptoms Index reliability for SRP-C.

[c] BASC- Emotional Symptoms Index reliability for general & clinical samples.

| Test Name | Content | Criterion-Related | Construct |
|---|---|---|---|
| **Children's Depression Inventory** (CDI, Kovacs, 1992) | Careful selection & review of affective, cognitive, & behavioral descriptors delineated by experts.<br><br>Initial item selection derived from established depression inventory(Beck Depression Inventory)<br><br>Adequate item-total correlation. | *Piers-Harris Children's Self-Concept Scale* (Friedman & Butler, 1979): r=.66<br><br>*Coopersmith Self-Esteem Inventory* (Green, 1980):r=.72(F); r=.67(M)<br><br>*Reynolds Adolescent Depression Scale* (Nieminen & Matson, 1989): r=.56 (Shain et.al.,1990): r=.94(F); r=.68(M)<br><br>*State-Trait Anxiety Inventory for Children* (Blumberg & Izard, 1986): r=.58<br><br>*Ctr. for Epidemiological Studies Depression Scale* (Weissman, et.al., 1980): r=.44<br><br>*Social Adjustment Scale-Self-Report* (Weissman, et.al., 1980): r=.50 | Significant differentiation b/w clinically diagnosed depressed children & nonclinical cases.<br><br>Age/Gender differences: Adolescent females endorsing greater depressive symptomology than males. |
| **Reynolds Adolescent Depression Scale** (RADS, Reynolds, 1986) | Careful selection & review of emotional/behavioral descriptors delineated by expert systems of classification.<br><br>Strong item-total correlation. | *Beck Depression Inventory* (Beck,et.al.;1961): r = .73<br><br>*Ctr. for Epidemiological Studies-Depression Scale* (Radloff, 1977): r=.75<br><br>*Children's Depression Inventory* (Kovacs, 1979, 1981): r=.73<br><br>*Zung Self-Rating Depression Scale* (Zung, 1965): r=.72<br><br>*Hamilton Depression Rating Scale* (Hamilton, 1960): r=.84 | Significant differentiation b/w clinically referred & nonreferred adolescents.<br><br>Gender differences: Females endorsing greater depressive symptomology than males<br><br>Significant differentiation b/w educably mentally retarded & regular education students. |

| Test Name | Content | Criterion-Related | Construct |
|---|---|---|---|
| **Reynolds Child Depression Scale** (RCDS, Reynolds, 1989) | Careful selection & review of emotional/behavioral descriptors delineated by expert systems of classification.<br><br><br><br>Strong item-total correlation. | *Children's Depression Inventory* (Reynolds et.al.,1984,1985): r=.72<br><br>*Children's Depression Rating Scale-R* (Tanaka-Matsumi et.al.,1986): r=.76<br><br>*Self-Esteem Inventory* (Reynolds et.al.,1984,1985;Stark,1984) r= -.65 (median)**<br><br>*Revised Children's Manifest Anxiety Scale* (Norvell, 1985;Kovacs,1983):r=.67(median) | Significant differentiation b/w clinically referred & nonreferred children.<br><br>Significant differentiation b/w gifted & nongifted children. |
| **Revised Children's Manifest Anxiety Scale** (RCMAS, Reynolds & Richmond, 1985) | Careful development & review of emotional & behavioral descriptors by experts. | *State-Trait Anxiety Inventory for Children* (Speilberger, 1973)<br>    Trait scale: r=.85    State scale: r=.24<br><br>*State-Trait Anxiety Inventory for Children* (Reynolds, 1985)<br>    Trait scale: r=.78    State scale: r=.08<br><br>*Walker Problem Behavior Identification Checklist* (Reynolds, 1982)<br>    r=.29 (total problem score;females)<br>    r=.32 (total problem score;males) | Significant differentiation b/w gifted & average children.<br><br>Significant differentiation b/w learning disabled & average children. |
| **State-Trait Anxiety Inventory for Children** (STAIC, Speilberger, 1973) | Careful development & review of emotional/behavioral descriptors as based on the STAI & other established inventories on anxiety. Also, guided by theoretical conceptions on anxiety as evidenced in the STAI. | *Children's Manifest Anxiety Scale* (Montgomery & Finch, 1974)<br>    Trait scale: r=.85<br><br>*General Anxiety Scale for Children* (Montgomery & Finch, 1974)<br>    Trait scale: r=.63 | Significant differentiation b/w anxiety-disordered & non-anxiety-disordered children.<br><br>Significant differentiation b/w anxiety levels among students w/ varying degrees of academic problems. |

| Test Name | Content | Criterion-Related | Construct |
|---|---|---|---|
| **(cont.)**<br>**State-Trait Anxiety**<br>**Inventory for Children** | | *California Achievement Test &*<br>*California Test of Mental Maturity*<br>  (Speilberger, 1973)<br>  Negative relationship b/w anxiety &<br>  ability-achievement scores. | |
| **Youth Self-Report**<br>(Internalizing broad band)<br>(YSR, Achenbach, 1987) | Careful development & review<br>of emotional/behavioral<br>descriptors by experts | Ability of Quantitative scale scores<br>  to discriminate b/w referred & nonreferred<br>  youths after demographic effects are<br>  partialled out. | Significant differentiation b/w<br>clinically referred & nonreferred<br>children/youth.<br><br>Lack of similar instruments limits<br>possibility for testing construct<br>validity regarding correlation of<br>syndrome scales derived from other<br>instruments. |
| **Behavioral Assessment**<br>**System for Children**<br>(Self-report of Personality<br>form,internalizing scales<br>(ESI);(BASC,Kamphaus &<br>Reynolds, 1992)<br><br>ESI=Emotional Symptom<br>  Index | Careful development & review<br>of emotional/behavioral<br>descriptors by experts. | *Minnesota Multiphasic Personality*<br>  *Inventory* (Hathaway & McKinley,1970)<br>  r=-.03(masculinity);r=.85(anxiety)<br><br>*Youth Self-Report* (Achenbach,1985)<br>  r=.70 (internalizing problems w/ F)<br>  r=.65 (internalizing problems w/ M)<br><br>*Behavior Rating Profile* (Brown & Hammill,<br>  1983): r= -.63 to -.67<br><br>*Children's Personality Questionnaire,*<br>  *Form A* (Porter & Cattell, 1975)<br>  r= -.59 to .6 | Significant differentiation b/w<br>clinically referred & nonreferred<br>children/adolescents.<br><br>Factor analysis demonstrated<br>factor structure of SRP at the<br>two age levels as virtually<br>identical, despite minor difference<br>in item content & additional<br>scales in SRP-A. |

Note: All information is from the respective manual unless otherwise noted.

**Negative relationship is due to positive direction of item keying on self-esteem measure.

M=male

F=female

**REFERENCES**

Achenbach, T.M. (1991). Manual for the Youth Self-Report and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.

American Psychiatric Association (1985). Standards for educational and psychological testing. Washington, DC: Author.

American Psychiatric Association (1987). Diagnostic and Statistical Manual of Mental Disorders (3rd. ed., revised). Washington, DC: Author.

Bartels, C.L., DuBois, D.L., Felner, R.D., & Silverman, M.M. (1995). Stability of self-reported depressive symptoms in a community sample of children and adolescents. Journal of Clinical Child Psychology, 24(4), 386-396.

Berndt, D.J., Schwartz, S., & Kaiser, C.F. (1983). Readability of self-report depression inventories. Journal of Consulting & Clinical Psychology, 51(4), 627-628.

Bloom, B.S. (1976). Human Characteristics and School Learning. NY: McGraw-Hill.

Blumberg, S.H. & Izard, C.E. (1986). Discriminating patterns of emotions in 10- and 11-year old children's anxiety and depression. Journal of Personality and Social Psychology, 51(4), 852-857.

Brown, L.L., & Hammill, D.D. (1983). Behavior Rating Profile. Austin, TX: PRO-ED.

Carey, G., Clark, L.A. & Watson, D. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. Journal of Abnormal Psychology, 97(3), 346-353.

Carey, M.P., Finch, A.J., & Lonigan, C.J. (1994). Anxiety and depression in children and adolescents: Negative affectivity and the utility of self-reports. Journal of Consulting and Clinical Psychology, 62, 1000-1008.

Coopersmith, S. (1981). Self-Esteem Inventory-Form B. Palo Alto, CA: Consulting Psychologists Press.

Endicott, J., & Spitzer, R.L. (1978). A Diagnostic interview: The schedule for affective disorders and schizophrenia. Archives of General Psychiatry, 35, 837-844.

Friedman, R.J., & Butler, L.F. (1979). Development and Evaluation of a Test Battery to Assess Childhood Depression. Final report to Health and Welfare, Canada, for Project #606-1533-44.

Goodenough, F.L., & Harris, D.B. (1963). Goodenough-Harris Drawing Test . New York: Harcourt Brace Jovanovich.

Green, B.J. (1980). Depression in early adolescence: An exploratory investigation of its frequency, intensity, and correlates (Doctoral dissertation, Pennsylvania State University). Dissertation Abstracts International, 41, 437-457.

Hathaway, S.R., & McKinley, J.C. (1942,1943 [renewed 1970]). Minnesota Multiphasic Personality Inventory. Minneapolis: University of Minnesota Press.

Hamilton, M. (1982). Symptoms and assessment of depression. In E.S. Paykel (Ed.), Handbook of affective disorders (pp.3-11). New York: Guilford.

Hodges, K. (1990). Depression and anxiety in children: A comparison of self-report questionnaires to clinical interview. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 376-381.

Kamphaus, R.W., & Reynolds, C.R. (1992). Behavioral Assessment System for Children. Circle Pines, MN: American Guidance Service, Inc.

Kazdin, A.E., & Petti, T.A. (1982). Self-report and interview measures of childhood and adolescent depression. Journal of Child Psychology and Psychiatry, 23,437-457.

Kovacs, M. (1980-81, 1991). The Children's Depression Inventory. North Tonawanda, NY: Multi-Health Systems.

Kovacs. M. (1983). The Interview Schedule for Children (ISC): Interrater and parent-child agreement. Unpublished manuscript.

Kovacs, M., & Beck, A.T. (1977). An empirical-clinical approach toward a definition of childhood depression. In J.G. Schulterbrandt & A. Raskin (Eds.), Depression in childhood: Diagnosis, treatment, and conceptual models. New York: Raven.

La Greca, A.M. (1990). Through the Eyes of the Child. Boston: Allyn & Bacon.

McGee, R., & Williams, S.M. (1988). Childhood depression and reading ability: Is there a relationship? Journal of School Psychology, 26, 391-394.

Merrell, K.W. (1994). Assessment of behavioral, social, & emotional problems: Direct and objective methods for use with children and adolescents. White Plains, NY: Longman Publishing Group.

Montgomery, L.E., & Finch, A.J. (1974). Validity of two measures of anxiety in children. Journal of Abnormal Child Psychology, 2(4), 293-298.

Nieminen, G.S., & Matson, J.L. (1989). Depressive problems in conduct-disordered adolescents. Journal of School Psychology, 27(2), 175-188.

Norvell, N., Brophy, C., & Finch, A.J. (1985). The relationship of anxiety to childhood depression. Journal of Personality Assessment, 49(2), 150-153.

Piaget, J. (1983). Piaget's theory. In P.H. Mussen (ed.), Handbook of Child Psychology (vol.1). New York: Wiley.

Porter, R.B., & Cattell, R.B. (1975). Children's Personality Questionnaire. Champaign, IL: Institute for Personality and Ability Testing.

Poznanski, E.O., Freeman, L.N., & Mokros, H.B. (1985). Children's Depression Rating Scale-Revised. Psychopharmacology Bulletin, 21, 979-989.

Puig-Antich, J., & Gittelman, R. (1982). Depression in childhood and adolescence. In E.S. Paykel (Ed.), Handbook of affective disorders (pp. 379-392). New York: Guilford Press.

Quay, H.C., & La Greca, A.M. (1986). Disorders of anxiety, withdrawal, and dysphoria. In H.C. Quay & J.S.Werry, (Eds.). Psychopathological disorders of childhood (3rd ed.) (pp. 73-110). New York: Wiley.

Quay, H.C., & Werry, J.S. (1972). Psychopathological Disorders of Childhood. (3rd., pp.230-259). New York: Wiley.

Reynolds, C.R. (1982). Convergent and divergent validity of the Revised Children's Manifest Anxiety Scale. Educational and Psychological Measurement, 42, 1205-1212.

Reynolds, C.R., & Richmond, B.O. (1985). Revised Children's Manifest Anxiety Scale. Los Angeles, CA: Western Psychological Services.

Reynolds, C.R. & Paget, K.D. (1981). Factor analysis of the Revised Children's Manifest Anxiety Scale for blacks, whites, males, and females within a national normative sample. Journal of Consulting and Clinical Psychology, 49, 352-359.

Reynolds, W.M. (1984). Depression in children and adolescents: Phenomenology, evaluation and treatment. School Psychology Review, 13, 171-182.

Reynolds, W.M. (1986). Reynolds Adolescent Depression Scale. Psychological Assessment Resources.

Reynolds, W.M. (1989). <u>Reynolds Child Depression Scale</u>. Odessa, FL: Psychological Assessment Resources.

Reynolds, W.M., & Coats, K.I. (1986). A comparison of cognitive-behavioral therapy and relaxation training for the treatment of depression in adolescents. <u>Journal of Consulting and Clinical Psychology, 54</u>, 653-660.

Reynolds, W.M. & Miller, K.L. (1985). Depression and learned helplessness in mentally retarded and nonmentally retarded adolescents: An initial investigation. <u>Applied Research in Mental Retardation, 6</u>, 295-306.

Shain, B.N., Naylor, M., & Alessi, N. (1990). Comparison of self-rated and clinician-rated measures of depression in adolescents. <u>American Journal of Psychiatry, 147(6)</u>, 793-195.

Spielberger, C.D. (1973). <u>State-Trait Anxiety Inventory for Children</u>. Palo Alto: CA: Consulting Psychologists Press.

Spitzer, R.L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. <u>Archives of General Psychiatry, 35</u>, 773-782.

Stark, K.D. (1984). <u>A comparison of the relative efficacy of self-control therapy and behavior therapy for the reduction of depression in children</u>. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Stark, K.D., Reynolds, W.M., & Kaslow, N.J. (1987). A comparison of the relative efficacy of self-control therapy and behavioral problem-solving therapy for depression in children. <u>Journal of Abnormal Child Psychology, 15</u>, 91-113.

Tanaka-Matsumi, J., & Kameoka, V.A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. <u>Journal of Consulting and Clinical Psychology, 53,</u>328-333.

Weiss, B., Weisz, J.R., Politano, M., Carey, M., Nelson, W.M., & Finch, A.J. (1991). Developmental differences in the factor structure of the Children's Depression Inventory. <u>Psychological Assessment, 3(1)</u>, 38-45.

Wendel, N.H., Nelson, W.M., Politano, P.M., Mayhall, C.A., & Finch, A.J. (1988). Differentiating inpatient clinically diagnosed and normal children using the Children's Depression Inventory. <u>Child Psychiatry and Human Development, 19(2)</u>, 98-108.