

THE VALIDITY OF SELF-REPORT MEASURES FOR YOUNG CHILDREN

by

KIMBERLEY BLAKER SAYE

(Under the Direction of Randy W. Kamphaus)

ABSTRACT

Questions of reliability, validity, and the ability of children to respond to child self-report measures of emotions and perceptions have historically made self-report measures inconsequential for research and clinical purposes. However, there have been many improvements in reliability and validity of these measures, and some developmental investigations indicate that children can report better than once thought. In order to add new information to this literature the current investigation assessed the validity of child self-report of emotions and behavior using the *Behavior Assessment System for Children (BASC) Self-Report of Personality (SRP-C)*, in comparison to *Teacher Rating Scales (TRS-C)*, and *Parent Rating Scales (PRS-C)*. The purpose of this study was to assess the external validity of the *BASC SRP-C* using a sample of 147 clinic-referred children ages 8-11 years and another sample of 36 school children ages 8-11 years. Predictive validity was assessed by correlating *BASC* scores with outcomes of referral for intervention and number of disciplinary reports. Child ratings were found to significantly correlate with discipline reports, while teacher and parent ratings were significantly correlated with both discipline reports and referral for intervention. Parent and child scores were correlated in the clinic sample and, overall, were found to agree at a higher-than-expected level. Predictive Discriminant Analysis was used to assess the utility of multiple

informant data for clinical diagnosis. Though child reports were found to be poor predictors of diagnostic categories when examined independently, the information obtained contributed uniquely to the results of a comprehensive assessment. Positive Predictive Power and Negative Predictive Power were also computed indicating that the *BASC SRP-C* scales did not have utility as specific inclusionary or exclusionary criteria for the diagnosis of a disruptive behavior problem. The results are discussed in terms of their additive value for clarifying the validity of child self-reports for specific research and clinical purposes.

INDEX WORDS: Behavior, Validity, *BASC*, Self-Report, Predictive Discriminant Analysis, Children, Predictive Power

THE VALIDITY OF SELF-REPORT MEASURES FOR YOUNG CHILDREN

by

KIMBERLEY BLAKER SAYE

B.A., University of Oklahoma, 1999

M.Ed., University of Georgia, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2005

© 2005

Kimberley Blaker Saye

All Rights Reserved

THE VALIDITY OF SELF-REPORT MEASURES FOR YOUNG CHILDREN

by

KIMBERLEY BLAKER SAYE

Major Professor: Randy W. Kamphaus

Committee: Roy P. Martin
A. Michele Lease
Thomas P. Hébert

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2005

DEDICATION

This work is dedicated to my parents, Scott and Cherry Blaker, who taught me the value of education early on, always believed in me, and insisted that there were no barriers I could not overcome.

I also dedicate this to my husband, Brian Saye, who has seen me through the entirety of graduate school. I cannot thank you enough for the countless things you have done to make things better for me during all of the clients, exams, and papers – particularly this one.

ACKNOWLEDGEMENTS

With much gratitude, I thank Dr. Randy Kamphaus for his efforts, encouragement, time, and teaching during my graduate program. I have learned an extraordinary amount in my time at UGA in large part due to your guidance. I would also like to thank my committee members present and past, Drs. Roy Martin, Michele Lease, Tom Hébert, and Joe Wisenbaker for their suggestions and thoughtfulness through this process. Thank you all for helping me believe I could attain my goals. Additionally, supervisors past and present, especially Drs. Dana Duggan, Carol Haisten, Barbara Wideman, and Leslie Munson provided much encouragement and many smiles throughout the process.

I am also grateful for the contributions made by my colleagues. I am especially thankful for the data sets collected by members of the Kamphaus and ACT Early research teams, as well as Clinic Coordinators past. I appreciate the comments and critiques given by my colleagues at UGA as well as in Cobb County, who have been wonderful as an open forum for ongoing research. I send out a special thank you to Beth Meisinger and Julie Bloom for always being there with brilliant ideas and supportive shoulders.

Warmest thanks go to my family. My husband, Brian, showed much patience and fortitude, often staying up into the wee hours as I typed. Thanks to his parents, Monroe and Libby Saye, for respecting the time demands of a grad student and never making me feel guilty about working; your consideration and support has meant a great deal. Thank you, also, to my parents, Scott and Cherry Blaker, for always believing in me, supporting me, and instilling in me the drive to achieve, as well as for the frequent questions about my graduation date.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
I INTRODUCTION AND REVIEW OF THE LITERATURE.....	1
Validation Studies Based on Construct or Test Content	5
Validation Studies Based on Response Processes.....	7
Validation Studies Based on Internal Structure	15
Validation Studies Based on Relations with Other Variables	17
Validation Studies Based on Consequences of Testing	25
Conclusions	26
II METHOD	28
Sample One	28
Sample Two.....	30
Instruments	32
Predictions and Analyses.....	34
III RESULTS	38
IV DISCUSSION	47
Predictive Validity of Self-Report.....	47

Parent and Child Agreement in a Clinical Sample.....	49
Differential Diagnostic Validity of Self-Report.....	51
Positive and Negative Predictive Power of Self-Report	53
Contributions to the Literature	54
Limitations of the Study	55
Conclusions and Future Directions	57
REFERENCES	60

LIST OF TABLES

	Page
Table 1: Primary Diagnoses and Percentages Associated with Each Diagnosis, Sample 1	30
Table 2: Comparisons (Validity Coefficients) To Be Obtained: Composite with Outcome	35
Table 3: Correlation of <i>BASC</i> Composite Scores with Number of Discipline Reports and Referral to SST	39
Table 4: Differences Between Parent and Teacher Ratings Using Fisher's Z Coefficient.....	40
Table 5: Correlation of SRP and PRS Ratings in a Clinic-referred Sample.....	41
Table 6: Means and Standard Deviations of SRP Composite Scores by Ultimate Diagnosis.....	41
Table 7: Concurrent Validity of SRP with Ultimate Diagnosis (Percentage of Cases Positive for Symptom Receiving a Diagnosis)	42
Table 8: Predictive Power of <i>BASC</i> Composite Scores for Discriminating between Children with or without a DSM-IV Diagnosis in the Clinical Sample.....	44
Table 9: Positive and Negative Predictive Power of the <i>BASC SRP-C</i> Maladaptive Scales	46

LIST OF FIGURES

	Page
Figure 1: The Relationship Among Diagnosis, Symptoms, and Item Statistics.....	37

CHAPTER I

INTRODUCTION AND REVIEW OF THE LITERATURE

Most children with mental health problems are never diagnosed or treated for their illness (Hennessy & Green-Hennessy, 2000). In fact, it is estimated that only one sixth to one half of children with psychiatric disorders are ever identified (Pagano, Cassidy, Little, Murphy, & Jellinek, 2000). Consequently, research on improved assessment methods and their application is crucial for identifying the presence of and risk for child psychopathology. This investigation focuses on the validity of child self-report measures as methods for identifying child psychopathology and current behavioral and emotional status.

On the positive side children are now more likely to be viewed as a valuable source of information about themselves, providing information that other sources, such as parents and teachers, cannot (Doerfler, Felner, Rowlison, Raley, & Evans, 1988). Indeed, part of the current research focus on self-report measures stems from a clinician's desire, during treatment or assessment of the child, to have insight into the child's own feelings, thoughts, and behaviors (Achenbach et al., 1987, Witt et al., 1988, Cole et al., 2000). Having the child's own insight into suspected problems can often aid in determining effective treatments or systemic interventions that can improve the child's adjustment.

Self-report measures fill the important role of defining the feelings, thoughts, and behaviors of children that are essential to assess when conducting a psychological evaluation. Kratochwill, Sheridan, Carlson, and Lasecki (1999) explain that it is important for the clinician to determine if the referral problem is actually the "primary target" for intervention. For

example, children may be referred due to behavioral concerns that are outwardly visible, such as hyperactivity, aggression, frustration, or noticeable learning difficulties. However, a particular child's responses on self-report measures may reveal that, though she exhibits learning difficulties, she has an underlying emotional problem that could be affecting her ability to participate competently in the school environment. Thus, this emotional problem could further contribute to reading or other academic difficulties. Knowing that the emotional problem exists, it would then be important to make both reading and emotional difficulties targets of further assessment and intervention.

Beyond the clinical perspective, most recently developed self-report measures aspire to meet the current standards for assessment instruments as described in the AERA/APA/NCME (1999) Standards for Educational and Psychological Testing (Eckert, Dunn, Guiney, & Coddling, 2003). Modern self-report measures typically address developmental issues, establish adequate psychometric properties, and have acceptable standardization samples (Eckert, et al., 2003). In fact, the technology of the self-report measure has improved a great deal, which has made it easier to obtain information more reliably and validly (Kamphaus & Frick, 2002).

There are numerous reasons for the attractiveness of self-report measures to clinicians and researchers. Elliott (1986) offered that children's perceptions are important predictors of treatment efficacy. Another important reason to obtain self-report measures from children is that they may report information that is subjective and hard for others to notice, such as symptoms of guilt or suicidal ideation (Reynolds, 1994). Additionally, self-report measures can be used for many purposes, including screening and diagnosing, identifying symptoms and problematic behaviors, and evaluating treatment outcomes (Silverman & Rabian, 1999).

On the negative side, self-report measures of personality and behavior have not enjoyed widespread use due to lingering concerns about their reliability and validity (Reynolds, 1993). One source of mistrust has been the fact that many of the earlier self-report measures were simply downward extensions of adult inventories (Stone & Lemanek, 1990). Children's developmental levels and life experiences necessitated more extreme changes in language and response format than was used in some of the adult inventories, thereby making the child measures suspect. Specifically, these changes were not accompanied by substantial evidence that the child scale measured the same constructs with similar validity. Child and adolescent self-report has either played a diminished role or been left out of clinical assessment procedures due to the perception that children were unreliable perceivers and reporters of events (Stone & Lemanek, 1990), or to the belief that children were not able to read and comprehend the statements on self-report measures (Kamphaus & Frick, 2002). Though Eckert et al. (2003), among others, note that self-report information is potentially biased, the consensus in the literature is that self-report information is invaluable to the assessment process. For example, Witt et al. (1988) argued that the information children provide is at least partially accurate and that the children's attributions or beliefs that may be "wrong" are equally important for clinicians to understand.

Today's focus of research regarding child self-report appears to be more developmental in nature, with researchers addressing concerns regarding the inherent pitfalls associated with downward extensions of adult measures, establishing the best approaches for eliciting information from children, and determining the degree to which children are valid reporters on self-report measures (Stone & Lemanek, 1990). Questions of reliability, validity, and utility of reports obtained from children are also being addressed, though there is a particular dearth of

studies that focus specifically on assessing the construct validity of child self-report measures. Studies regarding the validity of self-report measures with young children are lacking and, in particular, studies that directly compare the validity of parent report, teacher report, and self-report measures.

Validity, which is the most fundamental consideration in test development and evaluation, is defined as the “degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests,” (AERA, APA, NCME, 1999, p. 9). As such, validity is a process of gathering evidence to support the use of a specific measure and the inferences that can be drawn from the data obtained from that measure. Validation, or developing a sound validity argument to support the interpretation and use of scores, is a process that may involve changing of the test, the framework around the test, or the construct the test is purported to measure (AERA, APA, NCME, 1999).

The *Standards* produced by AERA, APA, and NCME provide a useful and widely accepted rubric for assessing the validity of measures. Validity evidence is organized into five categories, rather than traditionally-named concepts such as construct validity or discriminant validity. The five categories are as follows: 1) Evidence based on test content, i.e., wording, item format, administration and scoring guidelines; 2) Evidence based on response processes, i.e., the fit between the construct and the nature of response given by examinees; 3) Evidence based on internal structure, i.e., the degree to which the relationships among test items comprise the construct; 4) Evidence based on relations to other variables, i.e., the relationship of test scores to other external variables, such as convergent and discriminant evidence, test-criterion relationships, and validity generalization; and 5) Evidence based on the consequences of testing, i.e., the outcomes of a use of a test. This investigation focused on evidence based on relations

with other variables, although evidence of all five areas of validation is provided in the review of the literature.

In the review that follows and in the child assessment literature, “self-report” is a broad concept frequently involving “paper-and-pencil” rating scale responses, projective drawings, and interviews. Indeed, these are all ways of assessing a child’s thoughts and feelings, and each has its own strengths and limitations. However, this study focuses on one specific type of self-report, that of the rating scale variety. Measures like this are relatively brief and may be symptom-specific (e.g. *Children’s Depression Inventory*, Kovacs, 1992) or broad-based screeners (e.g. *Behavior Assessment System for Children*, Reynolds & Kamphaus, 1992). These “paper-and-pencil” rating scale measures require children to read statements and circle the responses provided that best match their feelings. For purposes of this study, the literature review and research questions have been specifically limited to self-report measures of the rating scale type; therefore, “self-report measures” as used in the remainder of the work should be understood to mean “rating scale self-reports.”

Validation Studies Based on Construct or Test Content

Because of the developmental issues associated with the task of reporting on one’s own behaviors (e.g., ability to self-monitor, developed emotional vocabulary, etc., see section to follow), self-report measures generally are not used with children ages 6 and below. However, some studies, including Zeltzer et al. (1988) have found that children can be better reporters at younger ages than previously thought. Witt and colleagues (1988) stated that children are probably capable of producing better and more accurate self-report data than thought previously, and much research is being done regarding modifications to the process that can help children better express what they know. Continued research into administration modifications is

important, because low reliability and validity may not necessarily reflect what the child can report, but problems with how the child is asked to report.

Zeltzer and colleagues (1988) explored the ability of very young children to use a rating scale and found that children were able to use a rating scale to quantify descriptions of somatic symptoms nearly as well as adults. Children and adults were read vignettes of persons experiencing symptoms of nausea and vomiting. They were then asked to rate the amount of time the child in the vignette spent experiencing symptoms of nausea and vomiting as well as the extent to which the child was bothered by the nausea and vomiting. Six faces were placed along a “bother” scale to illustrate varying degrees of bother. On their rating scale responses, younger children did not differentiate symptoms as well as older children, but the children’s overall responses were not strikingly different from the adults’ overall responses. All raters perceived highly distressing symptoms as occurring for a longer amount of time. Perhaps most importantly, the researchers concluded that children as young as 5 years old were able to use a rating scale to describe symptoms in a numerical fashion corresponding to severity.

Chambers and Johnston (2002) examined the effects of child age and number of response options on the tendency to complete Likert-type scales at the extremes. They noted that the younger children seemed to rate themselves at extremes, e.g., “not at all” or “always happy.” They tested this observation by assigning 60 children between the ages of 5 and 12 years to groups that received either 3-choice or 5-choice Likert-scale items on a rating scale. The tasks required involved making comparisons of a physical nature, making a determination of how strong a target child would feel under certain circumstances, and a subjective task in which children were asked to rate how they would feel under different imagined situations. Children were divided into three age groups: 5-6 years, 7-9 years, and 10-12 years. Five and 6 year olds

had more extreme scores than the other two groups. The subjective task also produced more extreme scores across ages. They found that younger children do tend to respond in an extreme manner compared to older children when rating others' and their own feelings. The number of response options (3 or 5) did not affect the tendency to respond at the extremes. Both of these studies suggested that young elementary age children were able to respond in a reliable manner to self-report items. Therefore, it appears feasible to create adequate test content for self-report measures at younger ages than previously thought.

Validation Studies Based on Response Processes

Developmental differences are important to address when discussing child self-report measures, because a child's age and/or cognitive level can add an unknown degree of error to obtained results. Children's perceptions of themselves may be different from the perceptions of adults due to diminished ability to monitor their thoughts and feelings or their diminished ability to compare themselves to others, among other difficulties. When examining validity issues, these developmental differences can help explain differences between child and adult ratings, as well as differences between child and adolescent responses on self-report measures.

On the surface, self-report measures seem like simple, relatively nondemanding tasks; however, Witt and colleagues (1988) explained that children must have the cognitive ability to properly interpret units and anchors of scales (such as Never, Sometimes, Often, and Always) and self-reflect on their experiences and emotions. For example, lower ability children were found to respond randomly to a self-report measure because they did not understand the items. Reynolds (1993) asserted that the results of self-report measures may also be affected by children's ability to control their impulses, their language skills, and their experiences with similar measures. If children are required to read the self-report, their reading level must be

adequate to the task. The use of simple language and familiar words is preferred, and there should be an option to read the items to the children if they are unable to read. Children must also have the skills to recognize their emotions and be able to report on the severity, frequency, and duration of the emotions.

Developmental Effects

Children under age 7 focus on physical characteristics or observable behavior when describing people and tend to see people as being either good or bad (Bierman, 1983). For example, when a 6-year-old child is asked to describe his gym teacher, he may say, “He’s tall and he eats Cheetos.” Around age 8, children begin to understand that others have separate beliefs and attitudes (Selman, 1980). Young children tend to have high self-concept scores and express their self-perception in overly positive terms; as they get older, however, these scores decline as the children begin to be able to make more realistic comparisons of themselves to others (Bender, 1997).

In further support of the notion of children’s overly positive self-ratings, Kolko and Kazdin (1993) also showed that there may be overreporting of symptoms by parents and underreporting of symptoms by children. Parents, teachers, and children were asked to complete, among other measures, the Achenbach *Child Behavior Checklist, Teacher Report Form*, and *Youth Self-Report* measures. They found significant but relatively low-to-moderate correlations among children, parents, and teachers, with children’s ratings being more similar to teachers’ ratings.

Other developmental characteristics important to consider in the use of self-report measures are that young children tend to use a narrow range of descriptors (Borke, 1973), categorize themselves in simple, all-or-nothing terms (Witt et al., 1988), and may be more likely

to express their self-perception in overly positive terms (Harter, 1990). In seeming support of the idea that children may be more likely to be overly positive when describing themselves, Handwerk, Larzelere, Soper, and Friman (1999) found that parents' ratings of child behavior were significantly worse than the child's or adolescent's reports on all scales of the *Youth Self Report* (Achenbach, 1991), a widely-used rating-scale report measure. It is noteworthy that this sample included only children referred for residential treatment, at an acute-care shelter, psychiatric inpatient hospital, or residential facility. The discrepancies between parent and child ratings remained relatively steady regardless of the child's age or of the child's placement. They hypothesized that there may be some overreporting of symptoms by parents and underreporting of symptoms by children due to the potential effect of placement in a treatment program. It is notable, however, that even with the discrepancies noted, parent and child ratings corresponded at $r = .24$ for the overall sample, with higher correlations for Internalizing problems.

When investigating age-mediated effects, Achenbach and colleagues (1987) found that the mean correlation across raters was significantly higher for 6-11 year olds than for adolescents. Further, the mean correlation was found to be higher for undercontrolled (externalizing) vs. overcontrolled (internalizing) problems. This result may indicate that the behavior of 6-11 year olds is easier to judge or is more consistent across situations; on the part of the adolescents, this finding could be due to an increased level of sophistication with regard to expression of emotion. Further, it makes sense that the undercontrolled problems are easier to see, and, therefore, report. Contradictory to Achenbach and colleagues' findings, Kolko and Kazdin (1993) reported that the younger age of the child led to greater parent-child disagreement, with the same overreporting by parents of overt behavior, though, overall, the

level of inter-informant agreement regarding children ages 6-9 years old was comparable to that of older children.

Another study that evaluated age effects was performed by Bondy, Sheslow, and Garcia (1985). They explored children's responses on self-reports in 2nd through 8th grade for a self-report measure of a child's fears. Mothers were asked to estimate their children's fears, and children were asked to rate their own fears. The children were administered the task twice, and were found to be reliable self-report agents. The authors indicated that their study supported the notion that young children are as reliable as older children and adults when self-reporting fearfulness.

Social-Emotional Factors and Their Effects

Social-emotional factors, specifically social-emotional symptoms leading to a clinical diagnosis, can impact a child's responses to self-report measures. These factors were studied by Epkins (1995a), who examined the contribution of inpatient-facility and community-based teachers' ratings of children's behavior. Participants were 83 inpatients at a psychiatric facility who were 8 to 12 years old. 79% had two or more diagnoses, including affective, disruptive behavior, anxiety, and adjustment disorders. Epkins assessed depression, anxiety, and aggression via child, inpatient teacher, and community teacher reports. After admission to the facility, all participants were given rating scales to complete. When examining convergence between teachers and children, Epkins discovered that the only significant convergence occurred between community teachers' and children's reports of aggression. Children reported higher levels of symptoms than inpatient teachers reported; this underreporting on the teachers' part may be because inpatient teachers have a different, more extreme, norm group with which they are faced daily. Children reported more anxiety than their community teachers reported. Epkins

reported that the poor teacher-child correspondence on trait-specific measures was inconsistent with research on children of similar ages but was consistent with other inpatient studies.

Epkins used the same inpatient sample and an additional elementary school sample for a study completed in 1993. The elementary school participants included 142 students in grades 3 through 5. Using the same self-report measure and teacher rating scale as in the 1995 study, Epkins assessed symptoms of depression, anxiety, and aggression. Inpatient children were found to report more symptoms than elementary school children, though only significantly more symptoms of depression. Younger inpatient children reported higher levels of symptoms than older inpatient children. The elementary school sample also had age effects; younger children corresponded significantly with teachers on all three dimensions rated; older children corresponded only on depression. Elementary school children reported significantly more internalizing problems than their teachers reported.

In the first study of interrater agreement based on students referred for school-based assessment, Lee, Elliott, and Barbour (1994) examined agreement between ratings on the *YSR*, *CBCL*, and *TRF* in a sample of 171 adolescent (age 12-17 years) boys who were referred for evaluation. The boys were divided by diagnosis: behavioral disorder, learning disability, or no diagnosis. Parents, teachers, and children each completed their specific Achenbach form; Total Problems, Internalizing Problems, and Externalizing Problems composite scales were then compared across respondents. Though modest, Lee and colleagues found that correlations were significant for parent and self ratings on Internalizing and Externalizing Problems and for teacher and self ratings on all composites. Regarding separate diagnostic groups, correlations between self and parent or self and teacher ratings remained modest. Self-ratings on Externalizing Problems were significantly lower than parent or teacher ratings. Self-report scores for all

composites were significantly lower than parent or teacher scores for the total sample. The authors concluded that school-referred children underreport their problems relative to parent and teacher reports. They suggested that this underreporting may be because of the threat of special education placement, the difficulty in admitting a problem exists, or developmental factors (see previous section).

Another study that examined school special education placement and self-report measures is Bussing, Zima, Belin, and Forness' 1998 study. The researchers used information from multiple informants, including child self-report, to determine what impact the diagnosis or risk for Attention-Deficit/Hyperactivity Disorder (ADHD) had on placement in Learning Disability (LD) and Serious Emotional Disturbance (SED) special education programs. Of particular interest to this review, Bussing and colleagues wanted to determine if children who meet criteria for ADHD differed on parent, teacher, or self-report measures from nondiagnosed, but high risk, peers. The total sample included 499 children; 148 children, 80% of which were male, were found to be at high risk for ADHD. The children were given the *CDI*, the *RCMAS*, the *Piers-Harris Self-Concept Scale*, and the *Diagnostic Interview Schedule for Children*. Of the children meeting criteria for ADHD, teacher ratings were elevated in 27%, whereas parent ratings were elevated in 78%. Child reports indicated small but significant differences between diagnosed and at-risk status. Specifically, children meeting the diagnostic criteria for ADHD reported slightly more symptoms of depression, anxiety, or lower self-esteem.

In one of the few studies that examined self-report measures for children diagnosed with Dyslexia, Heiervang and colleagues (2001) asked 25 fourth-grade Norwegian children diagnosed as Dyslexic to complete the *Youth Self Report*. Teachers and parents also completed rating scales regarding the child's behavior. According to both parents and teachers, significantly more

Dyslexic children than control children had behavior problem scores in the clinical range; however, child self-reports did not differ significantly from those of control children.

Montgomery (1994) studied the self-concept of children who were classified as high-achieving (HA), learning disabled (LD), or nondisabled. The sample consisted of 135 sixth through eighth grade children, their parents, and their teachers who were asked to complete rating scales addressing the self-concept of the child. Montgomery found that children with learning disabilities rated themselves as having poorer self-concept than their HA or nondisabled peers, but only in the areas of academics and competence, or ability to succeed. When asked to rate the global self-concept of the LD student, teachers rated the students lower than the children rated themselves, while parents' ratings generally matched their children's ratings.

Smith, Pelham Jr., Gnagy, Molina, and Evans (2000) conducted a single blind study of children's self-reported ADHD symptoms while undergoing several medication treatments. In all 36 adolescents aged 12 years and older who attended an intensive treatment summer camp for children with ADHD were randomly given varying amounts of medication (methylphenidate) each week. Children, their counselors, and teachers were asked to rate the child daily on the IOWA *Conners* scale and on several other sociometric variables. Additionally, trained observers took objective data regarding time on task in and out of the classroom, disruptive behavior in and out of the classroom, and frequency of teasing peers. The authors explained that their results, poor correlations between adolescent report and observer data, indicated that adolescents may be poor sources of information regarding ADHD symptoms. They further found results indicating that adolescents with ADHD were able to make meaningful reports about overt behavior but had more difficulty reporting on internal states of inattention or impulsivity. Additionally, adolescent

ratings provided no unique information beyond what was provided by the counselors and the teachers.

When examining referred vs. nonreferred adolescents, Compas and colleagues (1997) explored a large longitudinal sample whose participants completed the Achenbach *Child Behavior Checklist (CBCL)* and *Youth Self-Report (YSR)*. When scoring these measures for symptoms of depression, they found that adolescents who were referred for mental health services scored higher than nonreferred youths on both self-report measures and on parent ratings. Interestingly, clinically-referred girls scored consistently more deviantly than clinically-referred boys. The authors again pointed to the often-repeated hypothesis that adolescents can provide more sensitive reports of internalizing problems than can external observers.

In a longitudinal study investigating the emergence of depressive symptoms in children, Cole, Tram, Martin, Hoffman, Ruiz, Jacquez, and Maschman (2002) administered parent and child versions of the *Children's Depression Inventory* to a large group of school children (1,570 students) in grades 4 to 11. Their results indicated that child reports of depression may be sensitive to some depressive symptoms earlier than parent reports, particularly during early adolescence. The researchers hypothesized that this finding may be explained by the theory that awareness of such symptoms on the part of the parents requires a high degree of communication, communication which may be strained during adolescence. They also found that parents and children agreed more on the general trajectory of symptoms ($r = .65$), rather than the overall magnitude at any given time ($r = .50$). The authors suggested that clinicians might find assessment of symptom trajectories to be more reliable from informant to informant.

Once children are developmentally able to respond to the tasks of self-report, they must also be willing to disclose information in a truthful manner in order to obtain an accurate report.

Some research has demonstrated the effects of the social desirability and the environment on the self-reports of young children. For example, children with younger mental ages gave more socially desirable responses when asked to complete a self-report questionnaire (Mabe & Treiber, 1989). After reviewing the literature, James and colleagues (1994) concluded that children often feel they should “say the right thing” instead of reporting their true feelings. Further, if the children have been told their feelings are “silly” or have been brushed aside before, they may be unlikely to accurately disclose their thoughts or feelings. Bruck, Ceci, and Hembrooke (1998) concluded that children can be credible, reliable, and accurate reporters, even though their recollections may be sparse in details. They recommended a neutral tone, limited use of misleading questions, the absence of the induction of any motive or bias on the interviewer’s part, and the absence of threats, bribes, and peer pressure as being necessary to place the reports at less risk for contamination. Not surprisingly, empirical studies of response processes are sparse given that the *Standards* were just published in 1999. The current state of the literature is too theoretical and unsupported by data-based investigations to use for making generalizations about children’s response processes.

Validation Studies Based on Internal Structure

The *Behavior Assessment System for Children Self-Report of Personality (BASC SRP;* Reynolds & Kamphaus, 1992) evidenced a developmentally consistent internal factor structure. Composites were developed through the use of factor analysis of scale intercorrelations in the general normative sample of 9,861 children ages 8-18. Principal-axis analyses were completed for each age level of the SRP. Though 2, 3, and 4-factor solutions were obtained, the 3-factor solution was deemed meaningful, whereas the 4-factor solution was not. A covariance structure analysis was completed, using the principal-axis analysis results as a beginning point for the

model. The final results obtained through CSA were very similar to those obtained via principal-axis analysis and yielded three correlated but distinct factors. The following scales were found to load on the School Maladjustment Composite: Attitude to School (.69), Attitude to Teachers (.79), Locus of Control (.29), and Sense of Inadequacy (.44). Six scales loaded on the Clinical Maladjustment Composite: Anxiety (.79), Atypicality (.77), Depression (.40), Locus of Control (.65), Sense of Inadequacy (.45), and Social Stress (.87). The following scales loaded on the Personal Adjustment Composite: Depression (-.57), Interpersonal Relations (.71), Relations with Parents (.68), Self-Esteem (.69), and Self-Reliance (.61).

The *Youth Self-Report* (Achenbach, 1991), a measure similar to the *BASC* but designed specifically for children ages 11-18, has “syndromes” similar to *BASC* scales that are derived from the specific item content. The syndromes were derived from principal components analysis, in which correlations of items with other items are not reduced according to commonality. A sample of 709 boys and 563 girls were included in the study; they were seen in 26 settings, such as university psychiatric clinics, inpatient units, private practice facilities, and community mental health centers. According to the manual, eight syndromes were retained from principal components analysis. Factor analytic studies were then conducted using the same data, and Externalizing and Internalizing groupings were created based on the results of the factor loadings. Those syndromes that were found to load on the Internalizing factor were: Withdrawn (.784), Somatic Complaints (.690), and Anxious/Depressed (.650). The syndromes that loaded most heavily on the Externalizing factor were: Aggressive Behavior (.791) and Delinquent Behavior (.778). Attention Problems, Social Problems, and Thought Problems were reported not to have consistently high loadings on either factor.

Validation Studies Based on Relations with Other Variables

Convergent/Discriminant Evidence

In a study of concurrent validity performed by Doerfler and colleagues (1988), the researchers examined the validity of the *Child Depression Inventory (CDI; Kovacs, 1985)* by comparing it to the *Center for Epidemiological Studies Depression Scale (CES-D)*. The self-report ratings on both measures correlated moderately with each other but had poor correlations with teacher and parent ratings. It should be noted, however, that the poor correlations between self and teacher or self and parent reports were in the generally accepted range for parent-teacher correlations. Because the *CDI* correlated moderately with the *CES-D*, additional support was obtained for the validity of the *CDI* as an instrument to screen for depression.

In order to examine construct validity, Flanagan and Alfonso (1996) tested the convergent validity of the *BASC* and the *Social Skills Rating System (SSRS; 1990)*. They found large associations between the *BASC-PRS* and the *SSRS* parent form, though only small degrees of association between the *BASC-TRS* and the *SSRS* teacher form. However, similar scales across instruments were found to correlate well with each other.

Lonigan and colleagues (1994) investigated the validity and utility of self-reports for the diagnoses of depressive and anxiety disorders. Inpatient children ages 6 to 17 years at the psychiatric unit of a medical school were given the *Children's Depression Inventory (CDI; Kovacs, 1992)* and the *Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1985)*, two brief self-report measures designed to address specific affective states. Children were given a diagnosis based on a semistructured intake interview with the child and the child's guardian, as well as observation of the child across multiple settings. Results were factor analyzed for each self-report measure; convergent evidence was obtained if a factor

correlated more highly with other factors on the same measure. Moderately high levels of convergence were obtained for the *RCMAS*, while moderate to high levels were obtained for the *CDI*. These results led the researchers to conclude that self-reported symptoms of anxiety and depression are “at least partially discriminable” (p. 1003).

Test-Criterion Relationships

The predictive validity of the *General Behavior Inventory (GBI)* (Depue & Klein, 1988) was assessed by Klein, Dickstein, Taylor, and Harding (1989) in a study that examined the 6-7 month predictive utility of the measure. The *GBI*, a self-report inventory that can be used to screen for affective disorders in clinical settings, was specifically designed to distinguish between several types of affective disorders and identify affective conditions of varying severity. Several adult outpatients at a community mental health center and university-based training clinic were given the *GBI*, and took part in a diagnostic assessment interview. Follow-up assessments occurred six months after the interview and included several questionnaires. Patients scoring above the *GBI* Depression cutoff had significantly poorer outcomes on six of the seven outcome measures.

Morgan and Cauce (1999) examined the predictive validity of the Achenbach *Youth Self-Report (YSR)* (Achenbach, 1991), a widely used self-report measure. Homeless adolescents, ages 13-21, were given the *YSR* and the *Diagnostic Interview Schedule for Children (DISC)* as part of a study on case management. A stepwise discriminant function analysis was performed in order to find the best prediction equation for *YSR* data to *DISC* diagnoses, and several “rules of thumb” were provided that indicate high probability of diagnosis. The following sensitivity, specificity, and hit rate data was obtained:

<u>Diagnosis</u>	<u>Sensitivity</u>	<u>Specificity</u>	<u>Hit Rate</u>
Oppositional/Conduct Disorder	.70	.80	.75
Mania or Hypomania	.62	.67	.66
Depression/Dysthymia	.68	.78	.76
Schizophrenia	.74	.73	.73

The authors indicated that the *YSR* is a useful clinical screening tool and can also be used to help confirm clinical intuition. However, Carran and Scott (1992) cautioned that sensitivity and specificity rates must be higher than 0.8 to be considered “good” and 0.9 to be considered “high.”

In a study of predictive validity, Blumberg and Izard (1986) attempted to predict future depression from the self-reported emotions of children ages 10 and 11 years old. The children were given a measure to examine their anxiety symptoms, mood, and depressive symptoms, and their teachers were asked to rate the children’s emotions. The children were given the measures at two different times, with moderate correlations between measures at time one and time two. Blumberg and Izard found that attempting to predict future depression was successful when based upon children’s self-reported emotions ($r = .663$), but not when based on teacher’s ratings of the child’s emotions ($r = .219$). Again, the hypothesis used to explain this phenomenon is that, due to the internalizing nature of depression and anxiety, the child’s perceptions are more accurate. McGrath and Repetti (2002) found similar results in their predictive validity study of the *CDI*; self-reported depressive symptoms in fourth grade predicted a future increase in children’s negative thinking about themselves in sixth grade ($r = .52$).

One study that did address differences in prediction between parent and adolescent reports was carried out by Hope, Adams, Reynolds, Powers, Perez, and Kelley (1999). In their

study, Hope and colleagues evaluated whether adolescent report offered unique information beyond the parent report. Adolescents and parents were recruited for the study from outpatient clinics and from newspapers. Participants were interviewed, adolescents were given the *Youth Self-Report (YSR; Achenbach, 1991)* to complete, and parents were given both the *Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1991)* and *Conners' Parent Rating Scale-48 (CPRS; Conners, 1989)* to complete. A clinical interview was carried out using the *Diagnostic Interview for Children and Adolescents – Revised – Parent and Adolescent Forms (DICA-R-P and DICA-R-A; Reich, Shayka, & Taibleson, 1991)* and diagnoses were determined via reported symptoms meeting criteria found in the *Diagnostic and Statistical Manual, 4th edition (DSM-IV; American Psychiatric Association, 1994)*. Regression analyses were conducted in order to determine whether adolescent responses on the *YSR* contributed to the number of symptoms reported during the diagnostic interviews. They determined that the *YSR* Internalizing subscale contributed to parent report of child internalizing symptoms, though very modestly ($R^2 = 2.5\%$). When determining diagnosis based on interviews, parents and adolescents agreed only modestly regarding Internalizing disorders and agreed poorly regarding externalizing disorders (including Attention-Deficit/Hyperactivity Disorder).

An additional area of high interest validation that is not commonly assessed is the area of predictive power methods of item or scale validation. These statistics, which are conditional probabilities, can be useful in validation studies and as diagnostic aids. Statistics include sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). According to Elliott and colleagues (1993), Silverman and Rabian (1999), Laurent, Landau, and Stark (1993), and Milich, Widiger, and Landau (1987), sensitivity is the likelihood a child who has a diagnosis will exhibit a certain symptom. Specificity is the likelihood that a child who

does not have a diagnosis will not exhibit a certain behavior. Most rating scales err on the side of obtaining false positives rather than false negatives, because these types of scales are generally used as screening instruments.

The four studies cited above agreed that, though sensitivity and specificity are important statistics, when determining whether or not to make a diagnosis at the end of a psychological evaluation, the questions usually asked are, “What is the likelihood this child has the disorder, given this symptom?” and “What is the likelihood that this child does not have the disorder, given the absence of the symptom?” These questions are answered by PPP and NPP, respectively. PPP is a conditional probability that determines the usefulness of a symptom as an inclusionary criterion, whereas NPP is a conditional probability that determines the usefulness of a symptom as an exclusionary criterion. Waldman and Lilienfeld (1991) discussed how PPP and NPP are affected by the base rates of the disorder of interest. Though this affectation can add fluctuation across samples, they argued that this fluctuation could be an advantage because PPP and NPP would be able to reflect changes in diagnostic accuracy that can result from changes in base rates. Further, they explained that PPP and NPP could be used to facilitate the diagnostic decision making process; for example, symptoms with high NPP across settings could be particularly effective screening criteria.

Laurent, Landau, and Stark (1993) described the utility of PPP and NPP as being an aid to establish efficient differential diagnostic decision making. They encouraged school psychologists to compute local PPP and NPP rates for referred samples, which would then be used to determine the most efficient inclusionary and exclusionary criteria. The authors recommended looking at confirmatory and disconfirmatory data when making the decisions. The underlying theory for this process was that symptoms most descriptive of a disorder may not

be the most efficient indicators of the presence of that disorder. Because these item statistics have not been well known, there has been little research that examined these characteristics with a broad behavior rating scale. PPP and NPP have been used to validate rating scales but not self-report forms.

Interinformant Agreement

Children, adolescents, and adults responded somewhat differently when asked to rate the behavior and feelings of the same child. In a meta-analysis of 119 studies, Achenbach, McConaughy, and Howell (1987) showed that mean correlations between informants were low to moderate, at best. Mean correlations between self-report and parent report were .25; self-report and teacher report were .20; self-report and mental health worker report were .27. Younger children (ages 6-11) showed a significantly higher mean correlation than did adolescents. Even parent-parent correspondence or parent-teacher correspondence was moderate. These lower correlations indicated that respondents were completing the rating scales in different ways (for example, always marking the child's behavior at the extremes), they perceived the behaviors of the child in different ways, or that the variables differed from one situation to another. However, the authors stressed that the low correlations did not mean that the informants' information was unreliable or invalid. In fact, the low to modest correlations obtained from the meta-analysis led the authors to believe that no one informant's ratings can adequately substitute for another informant's ratings.

Crowley and Worchel (1993) provided other accounting for low correlations of source reports, namely measurement error due to instrument variation. For instance, method variance was of particular interest because not all items appeared in the same format on each scale used in an assessment. It is possible that a parent form, teacher form, and child form have all attempted

to ask the same question but ultimately may have done so in a slightly different manner. For example, when asking about depressive symptoms, the child version may have stated, “I feel sad at school,” the teacher version may have stated, “The child appears sad at school,” and the parent version may have stated, “My child appears sad when arriving home from school.” This variation, in turn, may have influenced the correlations between source reports.

Ultimately, the question remains: when examining the information obtained by the child’s self-report and reports by parents or teachers, how is a clinician to determine which report should be given more weight under which scenarios? What has been determined about the types of information children can provide with some accuracy and reliability? Between children, parents, and teachers, who does the literature say is a better reporter regarding what types of information? The answers to these questions remain unclear; there are no firm guidelines on how to resolve discrepancies and whose report to give the most weight (Stone & Lemanek, 1990). There is a fair amount of research regarding strengths and weaknesses of different raters, though few studies investigated the contributions of self-report measures, and even fewer included self-report measures completed by children younger than 10 years of age.

After a review of the literature, Kamphaus and Frick (2002) concluded that children may be better able to report internalizing problems, rather than externalizing problems. Many researchers agreed on this point (Loeber, Green, & Lahey, 1990; Garber 1984); parents and teachers were also found to be better able to report on a child’s externalizing problems or outward expressions of internalizing problems (Loeber, Green, & Lahey, 1990). This finding may have been because parents and teachers spend most of their time observing the child’s behavior. While they may have been unable to report on a child’s thoughts or feelings (other

than a child's statements of those thoughts or feelings), they were able to report with greater accuracy the child's frequency, intensity, and duration of outwardly visible behaviors.

Though these empirical differences are important to understand in order to interpret findings in a clinical setting, it is equally important to realize that simply because one reporter's information is given more weight during the assessment process does not mean that the other reporters cannot provide useful information on the same subject. In other words, though responses to self-report measures may seem suspect because of their lack of agreement with other ratings, it may be that children can provide access to information unavailable to peers, teachers, and parents. In this fashion, self-reports may represent somewhat different constructs than other-reports (Cole, Martin, Peeke, Henderson, & Harwell, 1998) and higher correlations may not be expected. Doerfler and colleagues (1988) reiterated that low correlations between respondents may have meant that variables may have differed from one situation to another, not that the informant's reports were invalid or unreliable.

Collectively, the ratings from different sources will likely provide more information about a child than if only one rater from one setting rated the child (Elliott et al., 1993). Indeed, Doerfler and colleagues (1988) as well as Kamphaus and Frick (2002) asserted that the information provided by all raters, regardless of correlation with other raters, were of value for clinical purposes due to the adequate to good psychometric properties of the measures and the raters' unique perspectives. Specifically, a review of the literature suggests that, in prior investigations, all raters possessed some evidence of valid responding. Therefore, it is perhaps best to understand whose ratings, under which situations, and for which purposes, should be primarily considered. Unfortunately, the literature to date provides no to few corroborating guidelines regarding which rater's results to use for which purpose or problem.

As seen above, there can be discrepancies between raters and errors that can occur, even with the use of the best report measures available. As Cole and colleagues (1998) suggest, instruments that are biased or nonequivalent across ethnic groups can result in the over- or underestimation of degree of psychopathology in individuals seeking treatment. It can be argued that this logic can also be applied to age ranges. For these reasons, it is important that research continue to evaluate the accuracy and validity of rating scales and the precise effects of variables that can affect the quality of the data obtained (Cole et al., 1998, Witt et al., 1988).

Validation Studies Based on Consequences of Testing

Pagano, Cassidy, Little, Murphy, and Jellinek (2000) adjusted the *Pediatric Symptom Checklist (PSC)*, a parent rating scale, to make a self-report measure, the *PSC-Y*. Pagano and colleagues found that obtaining self-report measures from children ages 9-14 as part of a wide-scale screening procedure added valuable information. In fact, when the youth form was combined with the teacher form, the “at-risk” recognition rate doubled from 10% to 20%. Further, the *PSC-Y* identified children with internalizing symptoms who were not identified by parents. This finding corroborated the previous findings that self-report measures were superior to parent or teacher reports as measures of internalizing problems, and helped validate the measure’s use through its ability to detect psychosocial dysfunction in children whose mental health needs had been previously unmet.

Using a sample of 755 children who comprised the normative and high-risk control sample in a longitudinal investigation conducted by the Conduct Problems Prevention Research Group (CPPRG), Flanagan, Bierman, Kam, and the CPPRG (2003) determined that a broad screening strategy based on child problem profiles showed superior sensitivity and overall accuracy in prediction of school outcomes. Teachers rated children in first grade on measures of

aggression, hyperactivity-inattention, and prosocial skills deficits. In third grade, school adjustment was assessed via interviews with the child, peer nominations, class grades, and teacher ratings. Behavior profiles were constructed for each child; scores one standard deviation above the mean of the normative population were considered problematic. Pearson correlations between first grade predictors and third grade outcomes and multiple regression analyses were conducted. All three first-grade behavior problem scales contributed to the prediction of behavioral adjustment and all were somewhat intercorrelated; most children in this study exhibited multi-problem profiles. The authors found that the presence of any one behavior problem, however, increased the risk for negative school outcome. A broad-based (dysregulated) screening model that identified children at risk when they exhibited high levels of aggression, hyperactivity, or prosocial skills deficits had the greatest sensitivity and the greatest odds ratio when predicting third grade difficulties (sensitivity of .63-.80, specificity of .76-.81, odds ratio of 5.49-15.06). This model did produce a slightly lower hit rate than other models based on two predictors (74-80%, compared to 77-88%).

Conclusions

There are several gaps in the current research regarding the validity of self-report measures. First, although correlational studies among raters are common, these studies do not typically address questions of validity per se (e.g. prediction of outcomes, etc.). Second, due to a dearth of validity evidence, research is not clear on whether or not information from one source is decidedly better than another and, if so, whether or not diagnostic decision making could be thereby improved. Third, studies of self-report measures that include item and scale characteristics of PPP and NPP are virtually nonexistent. Fourth, few studies have tried to examine several types of validity at once. Fifth, few studies have utilized short-term longitudinal

samples (six months or so) to assess predictive validity. Sixth, few studies used the *BASC*, a widely used relative newcomer in the behavioral assessment arena. Most importantly, when asking if the child report adds unique information over and above the parent and teacher reports, one is left with no clear guidelines from the existing scientific literature.

In order to address these shortcomings in the literature, this study; 1) directly compared the convergent, discriminant, concurrent, and predictive validity of parent, teacher, and child informants; 2) assessed the validity of assessment of both pathological behaviors and behavioral competencies; 3) assessed the hypothesis that PPP and NPP may add more to diagnostic decision making than do sensitivity and specificity; and 4) assessed predictive validity of child self-report by using a six month longitudinal investigation.

CHAPTER II

METHOD

Sample 1.

Sample 1, hereafter referred to as “S1,” was a clinic-referred sample. All participants were ages 8-11 years, or 96-143 months old who were seen at a university-based psychoeducational assessment clinic. Clinic services typically included a comprehensive psychoeducational evaluation that took place over the course of two days. Most assessments were sought due to parent or teacher referral. Parents typically waited 3-6 months after initial contact for an evaluation appointment. During the evaluation appointment, children and parents worked with two graduate students who served as the case manager and assistant in the clinic. Supervision of clinical practice was provided by doctoral-level psychologists. Children were selected for inclusion in the study based on having complete data in the clinic database, including one PRS-C and one SRP-C, as well as demographic information and diagnosis. The sample size was 147, and was approximately 70% male (N=103) and 30% female (N=44). The sample was approximately 94% Caucasian and 6% African-American. The children were ages 8 years (25%), 9 years (36%), 10 years (20%), and 11 years (19%). The socio-economic status of the sample, as measured by parent’s years of education completed, was higher than that of the surrounding community. 2.8% of the parents completed some high school; 29.1% obtained a high school diploma; 14.2% completed some college; 27.7% obtained a Bachelor’s degree; and 26.2% attended graduate school for at least one year.

Procedures. Prior to the child's assessment, informed consent was obtained from parents to use data from their child's visit in the clinic database and for possible future research. During the evaluation process, each participant received standardized measures of intelligence and achievement, the *BASC PRS-C* and *SRP-C* as measures of social-emotional functioning, and a diagnostic parent and child interview. With the parents' permission, the *BASC TRS-C* was sent to each child's teacher in order to be completed. All data was input into a database with identifying information deleted.

Outcome variables. The outcome variable not included in the *BASC* reports for this sample is the diagnosis given at the end of the evaluation. The diagnoses are presented as DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition*, APA, 1994) diagnostic codes. Space for three diagnostic codes was allowed in the database; however, only the primary diagnosis was used for the purposes of this study. Of the 147 participants, 47 (32%) were not given any diagnosis. Table 1 lists possible diagnoses and proportion of the sample with each diagnosis. Children with IQ scores below a standard score of 70 or those with low-incidence diagnoses such as an Autism Spectrum Disorder (N=8) were excluded from the study due to the cognitive demands required for responding to self-report measures.

Table 1. Primary Diagnoses and Percentages Associated with Each Diagnosis, Sample 1

Diagnosis (DSM-IV)	Diagnosis	N	% of sample
None	No Diagnosis	47	32
62.89	Borderline Intellectual Functioning	2	1.4
296.80	Bipolar Disorder NOS	1	0.7
300.00	Anxiety Disorder NOS	1	0.7
300.02	Generalized Anxiety Disorder	4	2.7
300.40	Dysthymic Disorder	3	2.0
307.60	Enuresis (Not Due to a General Medical Condition)	1	0.7
309.00	Adjustment Disorder with Depressed Mood	1	0.7
309.21	Separation Anxiety Disorder	1	0.7
309.24	Adjustment Disorder with Anxiety	1	0.7
309.40	Adjustment Disorder with Mixed Disturbance of Emotions and Conduct	1	0.7
313.00	Overanxious Disorder	1	0.7
313.81	Oppositional Defiant Disorder	2	1.4
314.00	Attention-Deficit/Hyperactivity Disorder, Predominantly Inattentive Type	26	17.7
314.01	Attention-Deficit Hyperactivity Disorder, Combined Type	25	17
314.90	Attention-Deficit/Hyperactivity Disorder NOS	4	2.7
315.00	Reading Disorder	17	11.6
315.10	Mathematics Disorder	2	1.4
315.20	Disorder of Written Expression	4	2.7
315.80	Developmental Expressive Writing Disorder	1	0.7
317.00	Mild Mental Retardation	1	0.7

Sample 2.

Sample 2, hereafter called “S2,” was a longitudinal sample, and part of the data collected by the A.C.T. Early Project (Addressing the Context of Teaching; Horne, Baker, & Kamphaus, 1996; Baker, Kamphaus, & Horne, 1999), a seven-year project. The project is a research study funded by grants through the Institute for At-Risk Children, Office of Educational Research and Improvement (OERI), U.S. Department of Education (grants R306F60158, R305T990330). The A.C.T. Early project was originally a three-year grant (ACT I) that examined behavioral characteristics of children at educational risk in kindergarten through second grade. A grant

extension was obtained in 1999 to extend research for three additional years (ACT II) and another extension year was provided (ACT III) in order to include fourth and fifth grades. This extension was applied for so that the longitudinal sample would contain students and teachers who had participated for four or more years. The primary objectives of ACT I, II, and III were to document factors at an individual, interactional, and ecological level that contribute to behavioral risk and adaptive skills in students. The study particularly addressed contextually maladaptive or problematic behaviors that contribute to educational risk. The children used in this analysis were in 1st-5th grades in three elementary schools of a small urban community in the Southeastern United States of approximately 100,000 inhabitants. All children were participants in years 6 and 7 of the ACT Early grant, during the 2001/2002 and 2002/2003 school year. The sample size used included 36 children who were 8-11 years old during the 2002/2003 school year. The sample was approximately 47% male (N=17) and 53% female (N=19). Approximately 61% of participants were African American, 33% Caucasian, and 5% multiracial, and 27% were in third grade, 42% were in fourth grade, and 31% were in fifth grade.

Procedures. In January of 2003, assistants with the ACT III project began collecting data on a subsample of children from year 6 of the grant. Permission slips were given to the child and the child's parents. Assistants collected a behavioral observation on each child using the *BASC Structured Observation System* no more than one time per week for a total of 10 observations. Assistants also interviewed the children individually each week for a total of 10 interviews. Children were also asked twice to rate their thoughts and feelings using the *BASC SRP-C*, once in February and once in May. The complete *SRP-C* was given in January, with a screener form developed for research collected in May. Following the same pattern, parents were twice asked to fill out the *BASC PRS-C* forms for their children, and teachers were twice asked to fill out the

BASC TRS-C forms. For purposes of this study, only ratings at time 1 were used. The *BASC Structured Developmental History (BASC-SDH)* was completed by parents in February. At the end of the school year, demographic information such as discipline records, referral for special education, and achievement test scores were collected from the teachers.

Outcome variables. Outcome variables for this sample were: referral for special education testing (coded 1 = referred and 0 = not referred), referral to Student Support Team (coded 1 = referred and 0 = not referred), and number of disciplinary actions taken over the school year, with numerical values available from 0 to 25.

Instruments

The *Behavior Assessment System for Children, Self-Report of Personality-Child (8-11 years)*, or *BASC SRP-C* (Reynolds & Kamphaus, 1992) is a nationally standardized measure that yields 9 problem behavior scales: Anxiety (alpha = 0.87), Attitude to School (alpha = 0.83), Attitude to Teachers (alpha = 0.75), Atypicality (alpha = 0.81), Depression (alpha = 0.88), Locus of Control (alpha = 0.87), Relations with Parents (alpha = 0.72), Sense of Inadequacy (alpha = 0.77), and Social Stress (alpha = 0.81) and 3 adaptive skills scales: Interpersonal Relations (alpha = 0.81), Self-Esteem (alpha = 0.79), and Self-Reliance (alpha = 0.71), as well as composite scores. The SRP-C composites are: School Maladjustment (alpha = 0.87; composed of Attitude to School and Attitude to Teachers), Clinical Maladjustment (alpha = 0.95; composed of Anxiety, Atypicality, Locus of Control, and Social Stress), Personal Adjustment (alpha = 0.90; composed of Relations with Parents, Interpersonal Relations, Self-Esteem, and Self-Reliance), and Emotional Symptoms Index (alpha = 0.95; comprised of Anxiety, Social Stress, Depression, Sense of Inadequacy, Interpersonal Relations, and Self-Esteem) (Reynolds & Kamphaus, 1992).

All scores are nationally normed with a mean of 50 and a standard deviation of 10 (T-Scores).

The items are answered dichotomously, either “True” or “False.”

The *BASC-Parent Rating Scales, Child (6-11 years)*, or *PRS-C* (Reynolds and Kamphaus, 1992) is a nationally standardized measure that yields nine problem behavior scales: Aggression (alpha = 0.83), Anxiety (alpha = 0.80), Attention Problems (alpha = 0.82), Atypicality (alpha = 0.58), Conduct Problems (alpha = 0.71), Depression (alpha = 0.83), Hyperactivity (alpha = 0.74), Somatization (alpha = 0.75), and Withdrawal (alpha = 0.73) and three adaptive skills scales: Adaptability (alpha = 0.77), Leadership (alpha = 0.83), and Social Skills (alpha = 0.89), as well as composite scores. The *PRS-C* Composite scores are: Externalizing Problems (alpha = 0.89, composed of Aggression, Hyperactivity, and Conduct Problems), Internalizing Problems (alpha = 0.88, composed of Anxiety, Depression, and Somatization), Adaptive Skills (alpha = 0.92, composed of Adaptability, Leadership, and Social Skills), and Behavioral Symptoms Index (alpha = 0.92, composed of Aggression, Hyperactivity, Anxiety, Depression, Attention Problems, and Atypicality). The items are rated by the parents using a four-point Likert scale from 1 (Never) to 4 (Almost Always). All *PRS-C* scores are nationally normed T-Scores with a mean of 50 and a standard deviation of 10.

The *BASC-Teacher Rating Scales-Child (6-11 years)*, or *BASC TRS-C* (Reynolds & Kamphaus, 1992), is a nationally standardized measure that yields nine problem behavior scales: Aggression (alpha = 0.95), Conduct Problems (alpha = 0.77), Hyperactivity (alpha = 0.93), Anxiety (alpha = 0.79), Depression (alpha = 0.87), Somatization (alpha = 0.77), Attention Problems (alpha = 0.93), Learning Problems (alpha = 0.90), and Atypicality (alpha = 0.84) and four adaptive skills scales: Adaptability (alpha = 0.83), Leadership (alpha = 0.89), Social Skills (alpha = 0.92), and Study Skills (alpha = 0.93), as well as composite scores. *PRS-C* Composites

available are Adaptive Skills (alpha = 0.97, composed of Adaptability, Leadership, Social Skills, and Study Skills), Externalizing Problems (alpha = 0.95, composed of Aggression, Hyperactivity, and Conduct Problems), Internalizing Problems (alpha = 0.91, composed of Anxiety, Depression, and Somatization), School Problems (alpha = 0.95, composed of Attention Problems and Learning Problems), and Behavioral Symptoms Index (alpha = 0.97, composed of Aggression, Hyperactivity, Anxiety, Depression, Attention Problems, and Atypicality). The items are rated by teachers using a four-point Likert scale from 1 (Never) to 4 (Almost Always). All *TRS-C* scores are nationally normed T-Scores with a mean of 50 and a standard deviation of 10.

The *BASC Structured Developmental History*, or *BASC SDH* (Reynolds and Kamphaus, 1992), is a structured interview questionnaire designed to address several aspects of a child's development and current functioning. Questions address pregnancy, birth complications, developmental milestones, family medical history, educational history, adaptive skills, behavioral strengths and weaknesses, the child's interests, and the family composition. Much of the demographic data (in terms of raw scores) was obtained from responses to the *BASC SDH*.

Predictions and Analyses

The *SRP* was predicted to be less predictive of school outcomes than either the *PRS* or *TRS*. Table 2 shows comparisons that were made for this portion of the study. Using *S2*, the Pearson Product Moment correlation coefficient between the Externalizing composite and the number of discipline reports, the Internalizing composite and the number of discipline reports, and the Adaptive Skills composite and the number of discipline reports was computed in order to assess predictive validity of the *PRS*, *TRS*, and *SRP*. The obtained predictive validity coefficients were then tested for the statistical significance of their differences using Fisher's *Z*

coefficient. Pearson Product Moment correlation coefficients were computed between the *BASC SRP-C* scores and the outcome variables of “referral” to obtain indices of predictive validity.

Table 2. Comparisons (Validity Coefficients) To Be Obtained: Composite with Outcome.

<i>BASC</i> Form	Composite	Outcomes
Parent	Externalizing Problems	Referral to SST Disciplinary actions
	Internalizing Problems	Referral to SST Disciplinary actions
	Adaptive Skills	Referral to SST Disciplinary actions
Teacher	Externalizing Problems	Referral to SST Disciplinary actions
	Internalizing Problems	Referral to SST Disciplinary actions
	Adaptive Skills	Referral to SST Disciplinary actions
Child	School Maladjustment	Referral to SST Disciplinary actions
	Clinical Maladjustment	Referral to SST Disciplinary actions
	Personal Adjustment	Referral to SST Disciplinary actions

Pearson Product Moment correlation coefficients between SRP and PRS ratings in a clinic-referred sample were expected to be .22 or higher, according to Achenbach and colleagues (1987). A MANOVA was used to determine if the child report had concurrent validity with the ultimate diagnosis. Differences in diagnosed and nondiagnosed subgroups were evaluated using Chi Square.

Self-reports were not expected to differentiate diagnostic groups as well as parent and teacher reports. Discriminant function analysis was used to determine which composite score, from which informant, was most meaningful for discriminating between children who were

diagnosed with a disorder vs. not diagnosed in the clinical sample. Predictive Discriminant Analysis (PDA) was used to determine whether or not specific measures or scales contributed enough information to justify their routine usage in clinical diagnostic evaluations (Lett, 1994), separately by informant and in combination.

Similarly, estimates of PPP and NPP were expected to be poor when examining self-reported disruptive behavior problems (Epkins, 1995b; Hoza, Pelham, Dobbs, Owens, and Pillow (2002), Using S1, PPP and NPP of the *BASC SRP-C* maladaptive scales (Anxiety, Atypicality, Locus of Control, Social Stress, Attitude to School, Attitude to Teachers, Depression, and Sense of Inadequacy) were computed for the sample (see Figure 1). Children who were ultimately diagnosed with any type of disruptive behavior problem (Attention-Deficit/Hyperactivity Disorder, Conduct Disorder, or Oppositional Defiant Disorder) were considered “cases” that met diagnostic criteria. Children who were not diagnosed with a disruptive behavior problem were considered “non-cases” that did not meet diagnostic criteria.

In this study, the *BASC SRP-C* T-scores for the maladaptive scales were considered to be the symptoms of interest, which is a nontraditional use of item statistics. However, previous researchers (e.g., Klein et al., 1989) have used test results as the “symptom” of interest. Because PPP and NPP are conditional probabilities, scores that approach $p = 1.0$ indicate that the scale being evaluated should be effective as an exclusionary or inclusionary criterion when considering a diagnosis of a disruptive behavior problem.

	Meets Diagnostic Criteria	Does Not Meet Diagnostic Criteria	
Symptoms Present	A		E
Symptoms Absent		C	F
	B	D	

Sensitivity (True Positives) = A/B

Positive Predictive Power (PPP) = A/E

Specificity (True Negatives) = C/D

Negative Predictive Power (NPP) = C/F

Figure 1. The Relationship Among Diagnosis, Symptoms, and Item Statistics. (Laurent et al., 1993)

CHAPTER III

RESULTS

The primary purpose of this study was to assess the external validity of the *BASC SRP-C* using a sample of 147 clinic-referred children ages 8-11 years (S1), and a separate sample of 36 school children ages 8-11 years (S2). Predictive validity was assessed by correlating *BASC* scores with outcomes of diagnosis and/or referral for pre-referral intervention or psychological testing.

Table 3 provides the correlation coefficients obtained. The highest, and most statistically significant values were obtained for the correlations between the PRS Externalizing Problems composite score and the number of disciplinary actions ($r = .74$) and referral to SST ($r = .61$). This finding indicates that as the parent's report of externalizing problems increases, so does the likelihood of the child's having discipline problems at school. Also, the likelihood that the child was referred to the SST for intervention increases. The PRS Adaptive Skills composite score was also statistically related to the number of disciplinary actions ($r = -.64$), indicating that poorer adaptive skills were correlated with increased disciplinary actions in this sample.

Table 3. Correlation of *BASC* Composite Scores with Number of Discipline Reports and Referral to SST

<i>BASC</i> form	Composite score	Variable of interest	<i>n</i>	<i>r</i>	<i>p</i>
Parent	Externalizing Problems	Disciplinary action	16	.74	<.01**
		Referral to SST	16	.61	.01*
	Internalizing Problems	Disciplinary action	16	-.14	.62
		Referral to SST	16	.39	.14
	Adaptive Skills	Disciplinary action	16	-.64	<.01**
		Referral to SST	16	-.38	.14
Teacher	Externalizing Problems	Disciplinary action	29	.53	<.01**
		Referral to SST	29	.28	.14
	Internalizing Problems	Disciplinary action	29	.11	.57
		Referral to SST	29	.10	.61
	Adaptive Skills	Disciplinary action	29	-.28	.14
		Referral to SST	29	-.43	.02*
Child	School Maladjustment	Disciplinary action	23	.46	.03*
		Referral to SST	23	.22	.32
	Clinical Maladjustment	Disciplinary action	22	.40	.07
		Referral to SST	22	.33	.13
	Personal Adjustment	Disciplinary action	23	-.38	.07
		Referral to SST	23	-.08	.72

* = $p < .05$, ** = $p < .01$

Similar to the parent ratings, two composite scores from the *TRS-C* were also found to be statistically significant. The Externalizing Problems composite score was related to the number of disciplinary actions ($r = .53$) and the Adaptive Skills composite score was found to be significantly correlated with child referral to SST ($r = -.43$).

The *SRP-C* scores were not significantly correlated to these outcomes, with the exception of the School Maladjustment composite, which was significantly correlated to disciplinary actions ($r = .46$). The School Maladjustment composite includes the Attitude to Teacher and

Attitude to School scales and indicates a pervasive pattern of dissatisfaction with the structure of the school, school personnel, and education itself (Reynolds & Kamphaus, 1992). As stated in the *BASC* manual, elevated scores are usually associated with severe problems within the school environment, so the fact that elevated scores in this sample are correlated with disciplinary actions is consistent with expectations.

Fisher's Z coefficient was calculated to determine if the correlation coefficients obtained from the analysis of the *PRS* and the *TRS* were statistically different from one another. Table 4 presents the results of these calculations. There were no statistically significant differences between the obtained correlation coefficients; this indicates that the composite scores of the Parent and Teacher forms were similarly correlated with the outcomes of interest. The small *n* of the study is a caveat, and more studies of a similar nature would need to be conducted before this finding can be generalized.

Table 4. Differences Between Parent and Teacher Ratings Using Fisher's Z Coefficient

<i>BASC</i> composite score	Variable of interest	Obtained r		<i>z</i> *
		Parents	Teachers	
Externalizing Problems	Discipline reports	.74	.53	1.04
	Referral to SST	.61	.28	1.23
Internalizing Problems	Discipline reports	-.14	.11	-.73
	Referral to SST	.39	.10	.92
Adaptive Skills	Discipline reports	-.64	-.28	-1.40
	Referral to SST	-.38	-.43	.17

* .05 level of significance requires a z coefficient of 1.65 or greater

With regard to *PRS* and *SRP* intercorrelations statistically significant correlations of $r = .22$ or above were found in five of nine calculations (see Table 5), indicating higher parent-child agreement than expected. Parents and children agreed on their ratings of adaptive skills and

externalizing problems as measured by school adjustment. Interestingly, parents' ratings of externalizing problems were correlated the highest with children's ratings of internalizing problems or poor adaptive skills. Parents' ratings of internalizing problems correlated with children's ratings of school maladjustment.

Table 5. Correlation of SRP and PRS Ratings in a Clinic-referred Sample

Parent Composite Scores	Correlation with Child's Composite Scores		
	School Maladjustment	Clinical Maladjustment	Personal Adjustment
Externalizing Problems	.22*	.33*	-.29*
Internalizing Problems	.22*	.19*	-.20*
Adaptive Skills	-.15	-.21*	.25*

* $p < .05$

To determine if differences in the child reports between diagnosed and non-diagnosed subgroups existed, two procedures were conducted. A multivariate analysis of variance (MANOVA) was conducted to determine if the SRP composite scores were significantly different for the ultimate diagnosis. No significant differences were found ($F = .66, p = .58$). The means and standard deviations of each composite score are presented in Table 6.

Table 6. Means and Standard Deviations of SRP Composite Scores by Ultimate Diagnosis

SRP Composite Scores	No DSM-IV diagnosis		One or more DSM-IV diagnosis	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
School maladjustment	49.08	10.60	51.48	10.08
Clinical maladjustment	48.22	9.98	50.29	11.11
Personal adjustment	51.49	7.97	49.86	8.71

In addition, a Chi-square analysis was used to determine if the child report correlated with the ultimate diagnosis. Table 7 presents, by ultimate diagnosis, the frequency and percentage of cases that were positive or not positive for symptoms for each composite score on the SRP. Cases were considered positive for School Maladjustment and Clinical Maladjustment symptoms when they reached the “At-Risk” range, with a T-score of 60 and above. Cases were considered positive for Personal Adjustment symptoms when they reached the “At-Risk” range, with a T-Score of 40 and below. As in the MANOVA, no significant differences were found. This cut score was used because it is one standard deviation above the mean and because the authors note that T-scores of 60 and above on the composites should be noted and investigated (Reynolds & Kamphaus, 1992).

Table 7. Concurrent Validity of SRP with Ultimate Diagnosis (Percentage of Cases Positive for Symptom Receiving a Diagnosis)

	No DSM-IV diagnosis		One or more DSM-IV diagnosis		χ^2	<i>p</i>
	<i>n</i>	%	<i>n</i>	%		
School Maladjustment						
Not positive for the symptom	35	31.5	76	68.5		
Positive for the symptom	12	33.3	24	66.7	.04	.84
Clinical Maladjustment						
Not positive for the symptom	41	34.5	78	65.5		
Positive for the symptom	6	21.4	22	78.6	1.77	.13
Personal Adjustment						
Not positive for the symptom	40	30.3	92	67.7		
Positive for the symptom	7	46.7	8	53.3	1.66	.24

Predictive discriminant function analysis (PDA) was used to examine which composite score, from which type of informant, was more accurate for discriminating, via hit rates, between children who were diagnosed with a disorder vs. not diagnosed in the clinical sample. Table 8 presents the frequency and percentage of cases correctly and incorrectly classified for each PDA conducted. All PDA analyses yielded significant functions except for that of the self-report, indicating that the self-report did not discriminate between diagnosed and nondiagnosed cases. Indeed, mean scores between children who were and were not diagnosed with a disorder were very similar. While results from the rating scales were quite accurate in correctly identifying children who were diagnosed with any type of disorder, there was a very high false positive rate amongst all raters individually. Parents alone, then teachers alone, often rated children as having high enough levels of problems to be considered disordered when they did not have a diagnosis. When models became more complicated, the accuracy in identifying those children who had a diagnosis remained high, but the false positive rate began to decline. The teacher and parent ratings combined correctly identified over 93% of those children who were diagnosed with a disorder, and the false positive rate was nearly 63%. Though the self-report yielded no significant information on its own, in combination with the parent and teacher information, it provided meaningful information that allowed correct classification of nearly 91% of true cases and reduced the false positive rate to less than 53%.

It should be noted that this analysis was run on data from a clinical sample whose diagnostic decisions could have been made, in part, based on results from the *BASC* scales. Most children were given a very extensive battery, however, which included psychoeducational testing, numerous additional rating scales, and a thorough diagnostic interview. This precludes the possibility that diagnosis was solely based on the results of the *BASC*. Future studies may

wish to gain access to children who have already been diagnosed to more thoroughly examine this question. It appears that, in a clinical sample, the self-report did not differentiate between diagnostic groups as well as other measures. However, the child ratings continued to add valuable information in conjunction with other reports that can help minimize false positives when determining diagnoses.

Table 8. Predictive Power of *BASC* Composite Scores for Discriminating between Children with or without a DSM-IV Diagnosis in the Clinical Sample

	Actual <i>n</i>	Predicted Group Membership			
		No DSM-IV diagnosis		One or more DSM-IV diagnosis	
		<i>n</i>	%	<i>n</i>	%
SRP composite scores					
No DSM-IV diagnosis	45	0	0.0	45	100.0
One or more DSM-IV diagnosis	97	0	0.0	97	100.0
PRS composite scores					
No DSM-IV diagnosis	43	5	11.6	38	88.4
One or more DSM-IV diagnosis	97	2	2.1	95	97.9
TRS composite scores					
No DSM-IV diagnosis	44	12	27.3	32	72.7
One or more DSM-IV diagnosis	90	8	8.9	82	91.1
TRS and PRS composite scores					
No DSM-IV diagnosis	40	15	37.5	25	62.5
One or more DSM-IV diagnosis	88	6	6.8	82	93.2
All composite scores					
No DSM-IV diagnosis	38	18	47.4	20	52.6
One or more DSM-IV diagnosis	86	8	9.3	78	90.7

Using S1, PPP and NPP of the *BASC SRP-C* maladaptive scales (Anxiety, Atypicality, Locus of Control, Social Stress, Attitude to School, Attitude to Teachers, Depression, and Sense

of Inadequacy) were computed for the sample. Children who were ultimately diagnosed with any type of disruptive behavior disorder (Attention-Deficit/Hyperactivity Disorder, Conduct Disorder, or Oppositional Defiant Disorder) were considered as “cases” that met diagnostic criteria. Children who were not diagnosed with a disruptive behavior disorder were considered as “non-cases” that did not meet diagnostic criteria.

Table 9 provides the PPP and NPP for each of the seven scales. No scales had PPP or NPP values approaching 1.0. Because PPP and NPP are conditional probabilities, scores that approach $p = 1.0$ indicate that the scale being evaluated should be effective as an exclusionary or inclusionary criterion when considering a diagnosis of a disruptive behavior problem. In this study, none of the *BASC* maladaptive scales appeared effective on their own as exclusionary or inclusionary criteria when making a diagnosis. In previous studies, PPP and NPP values of .75 and above were considered strong enough to be useful as inclusionary or exclusionary criteria (Laurent, Landau, & Stark, 1993). Only two scales were somewhat close to this value: the Depression scale yielded a PPP value of .63, whereas the Anxiety scale yielded an NPP value of .66. Scores in this range do not allow one to use the Depression scale as an inclusionary criterion, nor low scores on the Anxiety scale as an exclusionary criterion, for the diagnosis of a disruptive behavior problem. Thus, PPP and NPP were of little value in this sample.

Table 9. Positive and Negative Predictive Power of the *BASC SRP-C* Maladaptive Scales

	Meets	Does Not	Total	PPP	NPP
	Diagnostic	Meet			
	Criteria	Diagnostic			
	<i>n</i>	<i>n</i>			
Anxiety					
Positive for the symptom	18	13	31	.58	
Not positive for the symptom	39	77	116		.66
Atypicality					
Positive for the symptom	23	21	44	.52	
Not positive for the symptom	34	69	103		.33
Locus of Control					
Positive for the symptom	16	12	28	.57	
Not positive for the symptom	41	78	119		.34
Social Stress					
Positive for the symptom	16	12	28	.57	
Not positive for the symptom	41	78	119		.34
Attitude to School					
Positive for the symptom	19	25	44	.43	
Not positive for the symptom	38	65	103		.37
Attitude to Teachers					
Positive for the symptom	10	15	25	.40	
Not positive for the symptom	47	78	122		.39
Depression					
Positive for the symptom	15	9	24	.63	
Not positive for the symptom	42	81	123		.34
Sense of Inadequacy					
Positive for the symptom	17	22	39	.44	
Not positive for the symptom	40	68	108		.37

CHAPTER IV

DISCUSSION

As specified in previous chapters, the purpose of this study was to systematically assess the validity of a popular self-report measure for younger children ages 8 to 11 years old, the *BASC Self-Report for Children*. Results of this study will be discussed in reference to the four hypotheses posed in chapter two.

Predictive Validity of Self-Report

For this portion of the study, parents' ratings of their child's externalizing problems correlated with the child's number of discipline problems and the likelihood that the child was referred to the school SST for intervention. The adaptive skills ratings given by parents were also related to the number of disciplinary actions children received. Caution should be used before generalizing these findings to the population at large, however, as the study was limited by a very small sample size. Despite best efforts to obtain parent information, including financial incentives, only 16 students in the selected age range had information returned. Future studies may wish to consider this return rate and develop strategies to improve parent participation.

Teacher ratings showed a similar trend, with teacher ratings of externalizing problems being related to the number of disciplinary actions and teacher ratings of adaptive skills being related to the probability that the child was referred to SST. These findings suggest that, in a public school setting, child externalizing and adaptive skills problems tend to lead to disciplinary action and referral to SST more frequently than do internalizing problems. This focus on

externalizing problems may be due to difficulties in identifying internalizing problems, as has been previously suggested in the literature (Epkins, 1993). However, children of this age also may express internalizing symptoms differently (Pagano et al., 2000; Hammen & Compass, 1994) and thus may appear to have more externalizing problems in the school setting. Again, generalizability must be limited due to the small sample size of the study.

Child ratings in this study showed correlations only for the school maladjustment composite, indicating that negative feelings toward school and teachers, and dissatisfaction with the schooling process, were significantly related to the number of disciplinary actions a child received. Surprisingly, no other self-ratings were significantly associated with the school outcomes of interest. The lack of other findings may be due in part to the very small sample size used for this portion of the study; it is possible that sufficient power was not present to fully explore and find significant relationships. Full self-report data was available for only 22 subjects, which makes generalization difficult. Further studies with larger sample sizes would be needed to replicate these results and generalize findings. Furthermore, though the school outcome data of referral to SST and disciplinary action was collected over the course of the school year and is a strength of the study, the self-report data used was taken only at one time, a potential limitation, as this timing does not allow for changes in the students' behavior throughout the year. Future studies may wish to have many times at which student data are collected.

Despite the small sample sizes, all three report forms of the *BASC* showed significant correlations with external measures of importance to school psychologists, educators, and parents alike. Ratings by teachers, parents, and to some extent the children themselves, were somewhat predictive of referral to SST and the number of discipline problems that would occur

throughout the year. This finding yields support for the short-term predictive validity of the *BASC* and lends credence to its use in the schools.

Parent and Child Agreement in a Clinical Sample

In a clinical sample of 147 children, parent-child agreement was assessed for the *BASC* composite scales. Using Achenbach's (1987) inter-rater correlation of $r = .22$, the hypothesis that parent-child agreement would be similar in a clinical sample was partially supported. Parent-child agreement higher than or equal to $r = .22$ was found in five of nine areas. Parents and children agreed on their ratings of adaptive skills and externalizing problems as measured by school adjustment. Parents' ratings of externalizing problems correlated the highest with children's ratings of internalizing problems and poor adaptive skills. This finding may suggest that, though parents endorse externalizing problems, children often may be feeling underlying social stress, anxiety, or other internalizing distress of which the parent is not aware. Indeed, other studies (Pagano et al., 2000; Compas et al., 1997; Cole et al., 2002) have found that self-report measures often identify children with internalizing symptoms who were missed by parents; these results again seem to confirm the superiority of the self-report measure for knowledge of internalizing problems. Parents' ratings of the children's internalizing problems did not correlate highly with the children's ratings of internalizing problems but with the ratings of school maladjustment. Parents' ratings of adaptive skills were poorly correlated with the children's ratings of school maladjustment, suggesting that adult constructs of adaptability, social skills, and leadership may relate little to the child's feelings toward school and teachers.

Handwerk and colleagues (1999) conducted a similar study using a clinical sample of adolescents. The current study was innovative in both its use of a younger population and a different self-report measure, the *BASC SRP* and its associated parent form. Handwerk and

colleagues found a similar overall correlation of .24 between parent and child ratings; in contrast to the present study, they found that parents and adolescents had highest agreement on the Internalizing, rather than Externalizing, components. The results of the current study may differ from Handwerk's findings in part because of the psychiatric sample used in that study. Specifically, many of the youths in the Handwerk study were hospitalized and received care for depressive or anxiety disorders and would likely have sought help due to severe difficulties with those symptoms. Additionally, age may be a factor in the differing results, as the findings from Handwerk and colleagues were based on youths ages 11 to 18.

This portion of the study was analyzed on data drawn from a university clinic. Berkson's bias, states that clinical populations are more severe than nonclinical populations due to self-referral to clinical settings (Berkson, 1946). It is possible that the correlations between raters may have been higher in this clinical setting than in the general population due to the observed severity of the behaviors; because they were more severe, it is likely that the behaviors were more easily observable and more easily rated. However, because this is a clinical population, generalizability should be limited to other clinical populations. Additionally, this clinical sample was drawn from a clinic designed to see patients primarily for learning and behavioral difficulties, not necessarily psychiatric disorders. A strength of this study is that measures were taken to exclude those participants who may have had significant difficulty interpreting and responding to the items. However, it is possible that the correlations obtained may be lessened or strengthened in a psychiatric clinical setting, and future studies should explore uses with psychiatric populations. Future studies may also wish to explore examining the differences in composite scores by individual diagnoses, which would require a substantial sample size.

Differential Diagnostic Validity of Self-Report

This study found that the *BASC Self-Report* was unable to differentiate between diagnosed versus nondiagnosed subgroups in a clinical sample. Further, teacher and parent ratings were well able to identify children who were eventually diagnosed, with accuracy rates of above 90%. However, numerous false positives, between 72 and 88%, were seen when only one rater's views were taken into account. When diagnostic models became more complex, the ability to identify true positive cases was not markedly diminished, and false positive rates declined. When both teacher and parent ratings were considered, the model accurately predicted 93.2% of positive cases and had a 62.5% false positive rate. Though the self-report did not show evidence of discrimination on its own, when data were aggregated in a model with parent and teacher views, the child data added information that was able to reduce the identification of false positives, decreasing the rate to 52.6%, while preserving the true positive identification rate at 90.7%.

This study is somewhat similar to Morgan and Cauce's study, conducted in 1999, regarding the predictive validity of the *Youth Self Report*. Using discriminant function analysis, Morgan and Cauce found superior sensitivity rates between .62-.74. One reason for the lack of findings in the current study regarding the *SRP* on its own could be the difference in age; Morgan and Cauce studied adolescents, while the present study used 8-11 year old children as participants. Because many of the children were seen for reading difficulties and attentional concerns, they may have had some difficulty with the task, such as difficulty understanding some of the items. Additionally, the Morgan and Cauce study was a psychiatric sample, which may have led to more severe ratings and more easily defined differences between groups.

With regard to the child information adding unique information beyond the parent and teacher ratings, findings from the present study were similar to those of Hope et al. (1999) who found that an adolescent's responses to a rating scale offered very modest unique information beyond that provided by their parents. Similarly, this study found that self-report information alone was unable to discriminate between diagnostic groups, but did add additional, unique information beyond the teacher and parent information. This finding further supported assertions from Compas and colleagues (1997), Loeber and colleagues (1990), and Cole and colleagues (2002) that children provide unique information. Higher ratings may have been obtained overall in this sample, creating false positives, because this was a clinic-referred sample; following Berkson's bias, the sample was likely more severe than the general population.

As previously mentioned, this data set is part of a university clinic data set and diagnoses were likely made based in part on the results of the *BASC* results. However, clients received an extensive battery of assessment services, including diagnostic interviews, observations, additional rating scales, cognitive assessments, achievement tests, and processing tests. Best practices indicated that no diagnosis be made on the basis of one instrument, so it was posited that no child's diagnosis was based solely on the *BASC* scores contained in the data set. However, future studies should seek to replicate the findings with a sample whose diagnoses are already known.

A strength of this study is that few others exist that have used predictive discriminant analysis to determine a rating scale's ability to differentiate between diagnostic groups, particularly in a clinic-referred setting. A fairly large sample size allowed for good power to be maintained. A limitation is that the data set spans many years, and diagnoses were made by many different doctoral-level school psychology students under the direction of several program

faculty members, each of whom had his or her own biases when making diagnoses. Improved consistency could be achieved by having a selected team of researchers make each diagnosis, or by reviewing case files to confirm diagnoses.

Positive and Negative Predictive Power of Self-Report

The hypothesis that PPP and NPP would be poor when examining self-reports in children who were diagnosed with disruptive behavior problems was supported by this study. Most PPP and NPP statistics calculated were poor, with levels averaging .46. Highest values were in the mid .60s, a moderate level, but a level that was not high enough to enable one to use the scale for an exclusionary or inclusionary symptom when making a diagnosis. PPP and NPP were calculated at the scale level for the maladaptive scales of the *BASC SRP*, which is a nontraditional use of item statistics. This innovative design was used in order to improve reliability of the data input into the analyses. Because the T-Scores input into the probabilities were norm-referenced scores, this was considered a strength of this study. As expected, PPP and NPP were poor for this study; the poor values were expected due to previous findings (Epkins, 1995b; Hoza et al., 2002) that children with disruptive behavior problems such as AD/HD have difficulty reporting on their own disruptive behavior and overestimate their conduct and areas of weakness. Indeed, having this hypothesis supported was, so to speak, a limitation of the study, as it is unclear whether the use of scale scores would be useful in a clinical setting for other disorders. In order to further assess the utility of scale scores as “symptoms” for PPP and NPP, future research would need to address the discrimination of internalizing problems using a measure such as the *BASC*.

No other study has examined the PPP and NPP of a broad-based rating scale, nor has another study addressed the ability of rating scales to determine diagnoses of children with

disruptive behavior problems. One study that tended to obtain much higher PPP and NPP results overall was conducted by Laurent, Landau, and Stark (1993) and involved the use of trait-specific self-report measures to differentially assess depression and anxiety in a clinical sample. Contrary to this study, the researchers used specific items, rather than scaled scores, to obtain their probabilities. Future studies should attempt to replicate findings, perhaps with a larger sample size and a psychiatric clinical sample, using scaled scores as the “items” or “symptoms” of interest.

Contributions to the Literature

The current study was designed to address some of the apparent gaps and contributes to the literature on the validity of child self-report measures in several ways. First, the study used the *BASC*, a relative newcomer on the testing market which allows assessment of psychopathology as well as competencies. Second, this study extended the validity evidence for the *BASC SRP* by establishing additional test-criterion relationships. For example, this study established a link between *BASC* scores and school outcomes of number of discipline reports and referral to SST. Parent and child ratings in a clinical sample were compared to each other to add further validity evidence; parent-child agreement of magnitude greater than expected was found for more than half of the composite scores. Child ratings were compared to an outcome of diagnosis to assess the concurrent validity of the *BASC SRP*; child reports were not significant for determining whether the child was ultimately diagnosed with a disorder. While parent and teacher ratings evidenced high sensitivity when used independently, the false positive rate was unacceptably high. When parent, teacher, and child ratings were included in a model together, high sensitivity was still present, though false positives were greatly reduced.

Studies of Positive and Negative Predictive Power are very rare in the literature. This study contributed information regarding PPP and NPP in a different type of clinical setting. Further, this study used these item statistics in a novel way, using scaled scores as the “item” of interest in an attempt to increase the reliability of the statistics.

Another area in which this study contributed was in the use of two samples of children ages 8 to 11 years old. Due somewhat to the paucity of rating scales such as the *BASC* being available to children in this age range, children in this group have not been studied as frequently as their adolescent counterparts. One of the purposes of this study was to broaden the validity support for self-report measures to include this often-overlooked age group and to determine if the children’s report added unique information above that of their parents and teachers.

Rather than focusing solely on the general population, this research also contributed to the literature by using a clinical sample from a university-based school psychology clinic, which was different in format from a psychiatric clinic and accessed a different type of population. Using this sample allowed the *BASC SRP* to be studied with a group of children who were being assessed primarily for concerns of attention, learning, and behavioral difficulties. Additionally, this study used a short-term longitudinal sample in a school setting, both of which were innovative practices for validity studies.

Limitations of the Study

One limitation of this study was the fact that in the clinical sample, many diagnoses were made based in part on the results of the *BASC*. This use of the *BASC* in order to help determine diagnosis likely increased sensitivity to the diagnosis to some unknown extent, though student clinicians and psychologists did not make diagnoses solely on *BASC* data. Future studies should seek to replicate the findings with a sample whose diagnoses are already known.

Additionally, the diagnostic system used (*DSM-IV*, APA, 1994) is in itself fallible. The categorical system of diagnosis involves a great deal of clinical judgment, making the criterion variable of diagnosis also fallible. As Jensen and Hoagwood (1997) indicate, selection and construction of classification systems involves decisions and does not occur naturally; therefore, disagreements will arise and judgment must be used when a case appears to meet two categories or none at all. Because psychological disorders cannot generally be proven by medical tests, they are diagnosed based on judgment of clinicians using a categorical system which is not static. All of these things can introduce error.

Another limitation of this study was its use of a clinical sample. Clinical samples are more likely to show comorbidity and severity of a disorder simply due to the process of self-referral to the clinic (Berkson, 1946). Parents who had children with more severe impairments were more likely to bring their children in to the clinic, which required that at least one parent miss work and that the child miss school. There was also a bias in that parents were more assertive in actively seeking services for their children and were willing to wait for the appointment with the faculty member and graduate clinician. Due to the use of the clinical sample, as has been mentioned before, generalizability of these findings is limited.

Collinearity must also be mentioned as a limitation. Some of the scales of the *BASC SRP* correlate highly with one another. It is possible that this high correlation between scales may account for the poor ability of the *SRP* scales to serve as inclusionary or exclusionary criteria.

The small sample size used in the school sample is also a limitation of the study. Though significant correlations were obtained, this study would need to be replicated with a much larger sample size in order to generalize the results.

Conclusions and Future Directions

This study provided some validity support for the use of the *BASC Self-Report of Personality* in the general and clinical populations as an indicator of risk for behavior problems. In a school setting, ratings by teachers, parents, and to some extent the children themselves, were somewhat predictive of referral to SST and the number of discipline problems that would occur throughout the year. However, the extremely small sample size prohibits generalization, despite the significance of the findings. The significance of the findings with such a small sample size does indicate, however, that there may be a very strong relationship to be found, which could improve ability to predict student success. Further research should investigate the correlations obtained here along with a number of other school outcomes of interest in order to determine the ability of the *BASC SRP* to predict additional school outcomes, such as standardized test passage or failure, etc. As with this study, longitudinal data would be beneficial in order to track children's outcomes over the course of at least one school year. A significant special education population would be beneficial in order to determine if the *BASC* could discriminate between school diagnoses.

Correlations between parent and child ratings in the clinical sample were very similar to those found in the *BASC* normative sample. Concurrent validity of the *SRP* was not supported when the *SRP* scores were used to determine which children were or were not diagnosed. Though the manual from Reynolds and Kamphaus (1992) states that the *BASC* "aids in the clinical diagnosis of disorders" and "assesses a variety of symptoms that are noted in the DSM-III-R" (p. 5), it is clear that psychological diagnoses were never meant to be based solely on *BASC* results. Because of this fact, it may seem spurious to research the diagnostic differential ability of the *BASC* or of similar rating scales; however, the key to effective treatment is to

identify and treat children as early as possible before dire consequences arise (Pagano et al., 2000) and there is a need for good instruments to do so. If future studies can show that self-report measures are able to discriminate diagnostic groups from one another, the results could provide additional evidence with which to help practitioners make diagnostic decisions.

Additionally, confirming the present findings that each informant yields unique information that can aid in diagnostic decision making with a larger clinical sample would be beneficial. Perhaps the next sample could involve a more severe sample in order to fully explore the ability of the *SRP* to discriminate between groups. Additionally, if the sample were large enough, researchers could explore if child self-report ratings could discriminate not only between diagnosed and not diagnosed subjects, but also between specific disorders.

The current study yielded some validity support for the *BASC SRP*, which is used with 8-11 year old children. Of course, psychologists and other mental health professionals work with a wide range of clients, many of whom are younger than 8 years old. Zeltzer and colleagues (1988) have shown that children as young as five years old are able to use a rating scale. However, very rarely are rating scales available for children under 8 years old. Focus of future research should move to extend these tools downward so as to help practitioners who are asked to assess the internal thoughts and feelings of the very young child. Continued research into administration techniques and other issues of response processes, as well as obtaining validity information for younger ages, is very important.

The estimates presented at the beginning of this study are startling: only one sixth to one half of children with psychiatric disorders are ever identified, leading to large numbers of children with untreated mental health problems that place them at risk for serious consequences (Pagano et al., 2000). Improved validity of instruments such as self-report measures increases

practitioners' ability to identify children most in need of services and thus provide them with interventions tailored to their needs, hopefully preventing, or at least minimizing, any resulting sequelae of the disorder. This study shows that children themselves are able to play a role in their identification; self-report measures yield unique information that should be important to practitioners.

REFERENCES

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological bulletin, 101*, 213-232.
- Achenbach, T. M. (1991a.). *Manual for the Child Behavior Checklist*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (1991b.). *Manual for the Teacher Report Form*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (1991c). *Manual for the Youth Self-Report and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, D.C.: Author.
- Bender, T. A. (1997). Assessment of subjective well-being during childhood and adolescence. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 199-225). San Diego, CA: Academic Press.
- Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics, 2*, 47-53.

- Bierman, K. L. (1983). Cognitive development and clinical interviews with children. In B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 6, pp. 217-250). New York: Plenum Press.
- 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*, 852-857.
- Bondy, A., Sheslow, D., & Garcia, L. T. (1985). An investigation of children's fears and their mother's fears. *Journal of psychopathology and behavioral assessment, 7*, 1-12.
- Borke, H. (1973). The development of empathy in Chinese and American children between three and six years of age: A cross-culture study. *Developmental psychology, 9*, 102-108.
- Bruck, M., Ceci, S. J., & Hembroke, H. (1998). Reliability and credibility of young children's reports: From research to policy and practice. *American psychologist, 53*, 136-151.
- Bussing, R., Zima, B. T., Belin, T. R., & Forness, S. R. (1998). Children who qualify for LD and SED programs: Do they differ in level of ADHD symptoms and comorbid psychiatric conditions? *Behavioral disorders, 23*, 85-97.
- Carran, D. T., & Scott, K. G. (1992). Risk assessment in preschool children: Research implications for the early detection of educational handicaps. *Topics in early childhood special education, 12*, 196-211.
- Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of pediatric psychology, 27*, 27-36.

- Cole, D. A., Peeke, L., Dolezal, S., Murray, N., & Canzoniero, A. (1999). A longitudinal study of negative affect and self-perceived competence in young adolescents. *Journal of personality and social psychology, 77*, 851-862.
- Cole, D. A., Hoffman, K., Tram, J. M., & Maxwell, S. E. (2000). Structural differences in parent and child reports of children's symptoms of depression and anxiety. *Psychological assessment, 12*, 174-185.
- Cole, D. A., Tram, J. M., Martin, J. M., Hoffman, K. B., Ruiz, M. D., Jacquez, F. M., & Maschman, T. L. (2002). Individual differences in the emergence of depressive symptoms in children and adolescents: A longitudinal investigation of parent and child reports. *Journal of abnormal psychology, 111*, 156-165.
- Compas, B. E., Oppedisano, G., Connor, J. K., Gerhardt, C. A., Hinden, B. R., Achenbach, T. M., & Hammen, C. (1997). Gender differences in depressive symptoms in adolescence: Comparison of national samples of clinically referred and nonreferred youths. *Journal of consulting and clinical psychology, 65*, 617-626.
- Conners, C. (1989). *Conners' Ratings Scales Manual: Conners' Teacher Rating Scales, Conners' Parent Rating Scales*. North Tonawanda, NY: Multi-Health Systems.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.
- Crowley, S. L., & Worchel, F. F. (1993). Assessment of childhood depression: Sampling multiple data sources with one instrument. *Journal of psychoeducational assessment, 11*, 242-249.

Doerfler, L. A., Felner, R. D., Rowlison, R. T., Raley, P. A., & Evans, E. (1988).

Depression in children and adolescents: A comparative analysis of the utility and construct validity of two assessment measures. *Journal of consulting and clinical psychology, 56*, 769-772.

Eckert, T. L., Dunn, E. K., Guiney, K. M., & Coddling, R. S. (2003). Self reports: Theory and research in using rating scale measures (pp. 288-322). In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations (2nd ed.)*. New York: Guilford.

Elliott, S. N. (1986). Children's ratings of the acceptability of classroom interventions for misbehavior: Findings and methodological considerations. *Journal of school psychology, 24*, 23-35.

Elliott, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and development. *School psychology review, 22*, 313-321.

Epkins, C. C. (1993). A preliminary comparison of teacher ratings and child self-report of depression, anxiety, and aggression in inpatient and elementary school samples. *Journal of abnormal child psychology, 21*, 649-661.

Epkins, C. C. (1995a). Teachers' ratings of inpatient children's depression, anxiety, and aggression: A preliminary comparison between inpatient-facility and community-based teachers' ratings and their correspondence with children's self-reports. *Journal of clinical child psychology, 24*, 63-70.

- Epkins, C. C. (1995b). Peer ratings of internalizing and externalizing problems in inpatient and elementary school children: Correspondence with parallel child self-report and teacher ratings. *Journal of emotional and behavioral disorders, 3*, 203-213.
- Flanagan, D. P., Alfonso, V. C., Primavera, L. H., Povall, L., & Higgins, D. (1996). Convergent validity of the *BASC* and *SSRS*: Implications for social skills assessment. *Psychology in the schools, 33*, 13-23.
- Flanagan, K. S., Bierman, K. L., Kam, C.-M., and Conduct Problems Prevention Research Group (2003). Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of clinical child and adolescent psychology, 32*, 396-407.
- Garber, J. (1984). Classification of childhood psychopathology: A developmental perspective. *Child development, 55*, 30-48.
- Gresham, F. & Elliott, S. (1990). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.
- Hammen, C., & Compass, D. (1994). Unmasking masked depression in children and adolescents: The problem of comorbidity. *Clinical psychology review, 14*, 585-603.
- Handwerk, M. L., Larzelere, R. E., Soper, S. H., & Friman, P. C. (1999). Parent and child discrepancies in reporting severity of problem behaviors in three out-of-home settings. *Psychological assessment, 11*, 14-23.

- Harter, S. (1990). Causes, correlates, and the functional role of global self-worth: A life-span perspective. In J. Kolligan & R. Sternberg (Eds.), *Perceptions of competence and incompetence across the life-span* (pp. 67-98). New Haven, CT: Yale University Press.
- Heiervang, E., Stevenson, J., Lund, A., & Hugdahl, K. (2001). Behaviour problems in children with dyslexia. *Nordic journal of psychiatry*, *55*, 251-256.
- Hennessy, K. D., & Green-Hennessy, S. (2000). Estimates of children and adolescents with school-related behavior problems. *Psychiatric services*, *51*, 591.
- Hope, T. L., Adams, C., Reynolds, L., Powers, D., Perez, R. A., & Kelley, M. L. (1999). Parent v. self-report: Contributions toward diagnosis of adolescent psychopathology. *Journal of psychopathology and behavioral assessment*, *21*, 349-363.
- Hoza, B., Pelham, W. E., Dobbs, J., Owens, J. S., & Pillow, D. R. (2002). Do boys with Attention-Deficit/Hyperactivity Disorder have positive illusory self-concepts? *Journal of abnormal psychology*, *111*, 268-278.
- James, E. M., Reynolds, C. R., & Dunbar, J. (1994). Self-report instruments. In T. H. Ollendick, N. J. King, & W. Yule (Eds.), *International handbook of phobic and anxiety disorders in children and adolescents* (pp. 317-329). New York: Plenum Press.
- Jensen, P. S. & Hoagwood, K. (1997). The book of names: DSM-IV in context. *Development and psychopathology*, *9*, 231-249.
- Kamphaus, R. W. & Frick, P. J. (2002). *Clinical assessment of child and adolescent personality and behavior* (2nd ed.). Boston: Allyn & Bacon.

- Klein, D. N., Dickstein, S., Taylor, E. B., & Harding, K. (1989). Identifying chronic affective disorders in outpatients: Validation of the General Behavior Inventory. *Journal of consulting and clinical psychology, 57*, 106-111.
- Kolko, D. J. & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Journal of child psychology and psychiatry, 34*, 991-1006.
- Kovacs, M. (1992). *CDI Manual*. North Tonawonda, NY: Multi-Health Systems.
- Kratochwill, T. R., Sheridan, S. M., Carlson, J., & Lasecki, K. L. (1999). Advances in behavioral assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology (3rd ed.)*. New York: John Wiley & Sons, Inc.
- Laurent, J., Landau, S., & Stark, K. D. (1993). Conditional probabilities in the diagnosis of depressive and anxiety disorders in children. *School psychology review, 22*, 98-114.
- Lee, S. W., Elliott, J., & Barbour, J. D. (1994). A comparison of cross-informant behavior ratings in school-based diagnosis. *Behavioral disorders, 19*, 87-97.
- Lett, N. J. (1994). *Differential validity of the Behavior Assessment System for Children (BASC) for children with Attention Deficit Hyperactivity Disorder (ADHD)*. [Dissertation]. University of Georgia: Athens, Georgia.
- Loeber, R., Green, S., & Lahey, B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of clinical child psychology, 19*, 136-143.

- Lonigan, C. J., Carey, M. P., & Finch, A. J. Jr. (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.
- Mabe, P. A. & Treiber, F. A. (1989). Social desirability response tendencies in psychiatric inpatient children. *Journal of clinical psychology, 45*, 194-201.
- McGrath, E. P. & Repetti, R. L. (2002). A longitudinal study of children's depressive symptoms, self-perceptions, and cognitive distortions about the self. *Journal of abnormal psychology, 111*, 77-87.
- Milich, R., Widiger, T. A., & Landau, S. (1987). Differential diagnosis of Attention Deficit and Conduct Disorders using conditional probabilities. *Journal of consulting and clinical psychology, 55*, 762-767.
- Montgomery, M. S. (1994). Self-concept and children with learning disabilities: Observer-child concordance across six context-dependent domains. *Journal of learning disabilities, 27*, 254-262.
- Morgan, C. J., & Cauce, A. M. (1999). Predicting DSM-III-R disorders from the Youth Self-Report: Analysis of data from a field study. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 1237-1245.
- Pagano, M. E., Cassidy, L. J., Little, M., Murphy, J. M., & Jellinek, M. S. (2000). Identifying psychosocial dysfunction in school-age children: The Pediatric Symptom Checklist as a self-report measure. *Psychology in the schools, 37*, 91-106.
- Reich, W., Shayka, T., & Taibleson, C. (1991). *The Diagnostic Interview for Children and Adolescents – Revised*. St. Louis: Washington University.

- Reynolds, C. & Richmond, B. (1985). *Revised Children's Manifest Anxiety Scale*. Los Angeles, CA: Western Psychological Services.
- Reynolds, C. R. & Kamphaus, R. W. (1992). *The Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service.
- Reynolds, W. M. (1993). Self-report methodology. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child and adolescent assessment* (pp. 98-123). Boston: Allyn and Bacon.
- Reynolds, W. M. (1994). Assessment of depression in children and adolescents by self-report questionnaires. In W. M. Reynolds & H. F. Johnston (Eds.), *Handbook of depression in children and adolescents* (pp. 209-234). New York: Plenum Press.
- Selman, R. L. (1980). *The growth of interpersonal understanding*. New York: Academic Press.
- Silverman, W. K. & Rabian, B. (1999). Rating scales for anxiety and mood disorders. In D. Shaffer, C. P. Lucas, & J. E. Richters (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 127-166). New York: Guilford Press.
- Smith, B. H., Pelham Jr., W. E., Gnagy, E., Molina, B., & Evans, S. (2000). The reliability, validity, and unique contributions of self-report by adolescents receiving treatment for Attention-Deficit/Hyperactivity Disorder. *Journal of consulting and clinical psychology*, 68, 489-499.
- Stone, W. L. & Lemanek, K. L. (1990). Developmental issues in children's self-reports. In A. M. LaGreca (Ed.), *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 18-56). Boston: Allyn and Bacon.

- Waldman, I. D. & Lilenfeld, S. O. (1991). Diagnostic efficiency of symptoms for Oppositional Defiant Disorder and Attention-Deficit Hyperactivity Disorder. *Journal of consulting and clinical psychology, 59*, 732-738.
- Weissman, M., Orvaschel, H., & Padian, N. (1980). Children's symptom and social functioning self-report scales: Comparison of mothers' and children's reports. *Journal of nervous and mental disease, 168*, 736-740.
- Witt, J. C., Cavell, T. A., Heffer, R. W., Carey, M. P., & Martens, B. K. (1988). Child self-report: Interviewing techniques and rating scales. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Conceptual foundations and practical applications* (pp. 384-454). New York: Guilford Press.
- Zeltzer, L. K., LeBaron, S., Richie, D. M., Reed, D., Schoolfield, J., & Prihoda, T. J. (1988). Can children understand and use a rating scale to quantify somatic symptoms? Assessment of nausea and vomiting as a model. *Journal of consulting and clinical psychology, 56*, 567-572.