

The Pennsylvania State University

The Graduate School

College of Education

A STRUCTURAL VALIDITY STUDY OF THE ADJUSTMENT SCALES FOR
CHILDREN AND ADOLESENTS – HOME (ASCA-H)

A Thesis in

School Psychology

by

Debra Young Coffey

Copyright 2006 Debra Young Coffey

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2006

The thesis of Debra Young Coffey was reviewed and approved* by the following:

Marley W. Watkins
Professor of Education
Thesis Adviser
Chair of Committee

Sherry E. Corneal
Associate Professor of Human Development

Kathy L. Ruhl
Professor of Special Education
Head of the Department of Educational and School Psychology and Special
Education

Barbara A. Schaefer
Associate Professor of Education
Professor-In-Charge of Graduate Programs in School Psychology

*Signatures are on file in the Graduate School.

ABSTRACT

Use of behavior rating scales has been called a “best practice” in the assessment of children’s emotional and behavioral problems. The technology of behavior rating scales has expanded widely in recent years, along with their popularity and availability. The Adjustment Scales for Children and Adolescents – Home (ASCA-H) is an objective rating instrument currently in development. It was designed to be completed by a parent/caregiver as a counterpart to the Adjustment Scales for Children and Adolescents (ASCA), a teacher-completed scale. The ASCA built upon limitations of traditional behavior rating scales with a unique format and multisituational approach. Rather than judge the frequency or intensity of a symptom as on traditional rating scales, a respondent is asked to indicate which behaviors characterize a child’s functioning with respect to specific situations on the ASCA. The ASCA defines psychopathology through expression of problem behaviors across multiple contexts and situations. As evidenced by the collective research on the ASCA, the cross-situational approach advocated by McDermott and colleagues has been effectively applied to a teacher-rating instrument, yet an unanswered question is whether this approach can also be applied to a parent version. The purpose of the present study was to examine the underlying dimensions and structural validity of the ASCA-H. Parents of 426 students ranging in age from 5 to 14 participated. Parent responses were evaluated using exploratory factor analysis to determine the dimensionality of the ASCA-H. A four-factor oblique solution generated through principal axis analysis best represented the data. Four first-order factors consistently emerged across rotation methods and number of factors extracted. Higher-order exploratory analysis revealed two overarching dimensions of behavior. Factors

were discussed in relation to comparable dimensions of the ASCA and other current behavior rating scales. Results of this study provided preliminary information regarding the latent structure of the ASCA-H. Implications for scale development and future research were presented.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	viii
Acknowledgements	ix
Chapter 1. INTRODUCTION	1
Method of Choice – Behavior Rating Scales	1
Types of Rating Scales	4
Sources of Information	6
Rating Scale Items	13
Rating Scale Format	16
Defining Problem Behavior	19
Adjustment Scales for Children and Adolescents (ASCA)	22
Adjustment Scales for Children and Adolescents – Home (ASCA-H)	28
Chapter 2. METHODS	35
Participants	35
Instrument	35
Procedures	38
Data Analysis	39
Item properties	40
Sampling	44
Appropriateness of factor analysis	46
Models	47
Extraction method	49
Number of factors	51
Factor rotation	55
Factor interpretation	57
Chapter 3. RESULTS	59
Conditions	59
Number of Factors	69
Factor Solutions	71
Item-Total Correlations	84
Factor Intercorrelations	84
Chapter 4. DISCUSSION	89
First-Order Factor Structure	90
Factor I	90
Factor II	93
Factor II	94
Factor IV	96
Factor Intercorrelations	97
Higher-Order Factor Structure	98
Internal Consistency	103

Item Endorsement	106
Scale Refinement	108
Limitations	112
Conclusions	113
REFERENCES	116

LIST OF TABLES

Table 1 Student Age and Sex	36
Table 2 Population Characteristics	37
Table 3 Description and Endorsement Prevalence of ASCA-H Items	60
Table 4 Comparison of Factor Solutions	73
Table 5 Four-Factor Structure of the ASCA-H	79
Table 6 ASCA-H Factor Summary	85
Table 7 Factor Correlations	86
Table 8 One-Factor Secondary Structure of the ASCA-H	87
Table 9 Two-Factor Secondary Structure of the ASCA-H	88

LIST OF FIGURES

Figure 1 Scree Plot	70
---------------------------	----

ACKNOWLEDGEMENTS

I would like to thank several individuals who assisted me throughout this process. First, I would like to acknowledge the support and contributions of my doctoral committee members: Dr. Marley Watkins, Dr. Barbara Schaefer, Dr. Sherry Corneal, and Dr. Kathy Ruhl. As my adviser, Dr. Watkins has provided much guidance, time, and motivation from Day 1 at Penn State. I am also grateful to many exceptional colleagues, professionals, and parents in the Baltimore County Public Schools for their time and assistance at each step of this project. I would also like to thank my family and friends for their support and encouragement. The faith and confidence of my parents has been unwavering. Finally, I must express my sincerest gratitude to my husband who knows better than anyone what it took to get to this point and without whom this work would not have been possible.

A Structural Validity Study of the Adjustment Scales for Children and Adolescents – Home (ASCA-H)

School psychologists are frequently asked to evaluate and treat students who exhibit emotional and behavioral difficulties. The process of identifying whether a student shows symptoms of larger emotional and behavioral problems, whether these symptoms constitute a significant impairment in functioning, and what interventions are appropriate is a complex one (Wingenfeld, 2002). The process is complicated by questions of the most appropriate means or methods of assessment.

Rather than rely on one procedure or method of assessment, many have recommended combining information from multiple methods and multiple sources of information to obtain a meaningful understanding of a student's functioning (Knoff, 2002; McConaughy & Ritter, 2002). Six primary assessment methods have been identified: behavioral observations, interviews, self-report instruments, projective-expressive techniques (e.g., drawings), sociometric techniques, and behavior rating scales (Merrell, 2001). In conjunction with other assessment methods, use of behavior rating scales has been described as a "best practice" in the assessment of emotional and behavioral problems (Knoff; McConaughy & Ritter).

Method of Choice – Behavior Rating Scales

The availability, use, and technology of rating scales for assessing children's emotional and behavioral problems have increased dramatically in the last 20 years (Merrell, 2001). Early surveys of assessment practices of psychologists indicated frequent

and more prevalent use of projective measures (e.g., House-Tree-Person, Thematic Apperception Test, and sentence completion techniques) over other methods (Goh, Teslow, & Fuller, 1981; Prout, 1983). In comparison, more recent surveys have shown increased use of objective measures, such as behavior ratings scales. For instance, in a replication of the Goh et al. study, Hutton, Dubes, and Muir (1992) found an increase in the reported use of behavior rating scales and a decrease in the use of several projective tests. Other surveys have found a similar increase in the popularity of behavior rating scales among school psychologists (Kennedy, Faust, Willis, & Piotrowski, 1994; Reschly, Genshaft, & Binder, 1987; Shapiro & Heick, 2004; Stinnet, Havey, & Oehler-Stinnett, 1994; Wilson & Reschly, 1996). Although assessment practices have been slow to change, Wilson and Reschly described the trend away from the use of instruments of poorer technical adequacy, such as projective measures, as a step in the right direction.

Empirically-based behavior rating scales represent an objective form of assessment that provides more reliable data than information derived from unstructured interviews or projective measures (McConaughy, 1992, 1993; Merrell, 2001). Kamphaus, Petoskey, and Rowe (2000) noted that with their expanding technical sophistication, behavior rating scales measure a broad range of constructs with evidence of validity, which then adds to an improved quality of assessment services. In addition, rating scales yield norm-based standard scores, which allow judgments of the severity of problems to be made in comparison to large samples of nonreferred children (McConaughy & Ritter, 2002).

Behavior rating scales offer many other advantages for practitioners.

Emphasizing the practical benefits, Merrell (2001) argued that rating scales “provide a ‘big picture’ of the assessment problem in a short amount of time, at moderate cost, and with a good deal of technical precision and practical utility” (p. 10). As an efficient and economical assessment method, school psychologists are then able to devote time to gathering additional information not available through rating scales, such as historical information gleaned from reviewing educational records (McConaughy, 1993) or physical characteristics of target behavior (i.e., topography and intensity) obtainable through direct observation.

McConaughy (1992, 1993) also noted that the large pool of items included on many behavior rating scales covers a broad range of potentially relevant problems that might extend beyond the initially identified referral concerns. This information may help to both broaden the context for understanding the referral problem and widen the focus of the assessment, thus better defining the areas of difficulty and better understanding the child’s functioning.

In addition, behavior rating scales capitalize on observations over time in the natural environments of children made by those individuals most familiar with the child’s behavior (Merrell, 2001). Merrell also noted that rating scales help practitioners obtain information about individuals who may not or cannot readily provide information about themselves. Similarly, Kamphaus and Frick (1996) argued that using behavior rating scales helps compensate for problems related to a child’s reading ability, oral language, or writing skills.

The nature of some relatively high frequency child disorders also supports use of rating scales. Attention Deficit Hyperactivity Disorder (ADHD) and other disruptive behavior disorders are best assessed using rating scales because children with such externalizing problems are known to underreport their symptomatology (Kamphaus & Frick, 1996). Rating scales also provide data on low frequency behaviors, such as violent behavior, that might not be seen in a limited number of sessions of direct behavioral observation (Merrell, 2001). Although direct observation may be helpful in understanding overt problem behavior, it is less effective or useful in understanding covert or internalizing behavior problems, which are covered on many rating scales (Merrell).

Types of Rating Scales

“The term rating scale is broad and encompasses multiple types of measurement, including checklists, questionnaires, inventories, self-reports, other-reports, indices, and other measures” (Myers & Winters, 2002, p. 115). Behavior rating scales differ in the range of areas of adjustment they address. They can be categorized as narrowband or broad band, depending upon the number of constructs they purport to measure (Volpe & DuPaul, 2001). Narrowband scales focus only on a single domain and can be disorder specific (e.g., ADHD) or problem specific (e.g., social skills; Wingenfeld, 2002). In contrast, broadband scales assess multiple behavioral domains (Volpe & DuPaul). Many broadband scales assess specific areas of adjustment (e.g., attention problems, social problems, or anxiety) and broad areas of adjustment, such as an externalizing dimension consisting of more disruptive-type behaviors and an internalizing dimension consisting of behaviors associated with anxiety, depression, and somatic concerns.

Broadband rating scales offer advantages over more narrowly focused scales, including economy in administration time and effort and comprehensiveness (Costenbader & Keller, 1990). With their expanded range, broadband scales also serve a screening function whereby other areas of difficulty, which may have not been specified in the initial concerns, may be identified and further assessed as necessary. Narrowband measures, such as the Conners Parent/Teacher Rating Scales (Conners, 1997), are more limited and specifically assess the symptoms of ADHD. In contrast, broadband measures assess comorbid features and assist in differential diagnosis (Volpe & DuPaul). Examples of broadband measures are the Child Behavior Checklist (CBCL; Achenbach, 1991a, 1991b) and Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992).

The CBCL reflects emotional and behavioral problems in two broadband dimensions (Internalizing and Externalizing) and nine narrowband dimensions, including Aggressive Behavior, Anxious/Depressed Behavior, Attention Problems, Delinquent Behavior, Somatic Complaints, Social Problems, Withdraw, Sex Problems, and Thought Problems. Items on both the parent and teacher report forms are scored on a 3-point scale ranging from 0 (*not true*) to 2 (*very true*).

The BASC is also widely used. Similar to the CBCL, the BASC includes parent, teacher, and child/self forms that tap multiple emotional and behavioral domains and produce scale scores that represent pathological and adaptive characteristics as quantitative deviations from the mean. It is described as a multidimensional measure that assesses both adaptive and problem behaviors. Items on the parent and teacher versions

are rated on a 4-point scale ranging from 0 (*never*) to 3 (*almost always*). The BASC yields *T* scores in broad Externalizing, Internalizing, and Adaptive Skills domains, as well as 14 specific content areas. The teacher version includes two additional content areas more specific to the school setting (i.e., learning problems and study skills) and yields an additional School Problems composite or domain score.

Sources of Information

Behavior rating scales, such as the CBCL and BASC, measure perceptions of behavior rather than provide a direct measure of the behavior's existence (Merrell, 1999) and thus rely on judgments and observations of people who are very familiar with a child's behavior. To obtain a meaningful understanding of a child's overall functioning, many experts recommend combining information from multiple methods and multiple sources of information (Knoff, 2002; McConaughy & Ritter, 2002). Teachers, parents, and children themselves make unique contributions to understanding children's behavioral and emotional adjustment (Wingenfeld, 2002).

In school settings, teachers often initiate referrals for special services (Gerber & Semmel, 1984) and thus become an important source of information regarding the nature of presenting concerns and difficulties. Teachers spend large parts of the day with children and are natural observers of behavior. They observe children both in the highly structured, task-oriented classroom setting and in less structured situations, such as recess (Epkins, 1995). Their experience of observing many students across time and contexts provides teachers with a good source from which to compare behaviors of children (Canivez, Perry, & Weller, 2001). They are also familiar with the age-appropriateness of

children's academic functioning, attitude toward learning, peer relations, and behavior (Wingenfeld, 2002).

Parents are also important sources of information. They typically have the highest level of involvement with their child, have known him or her for longer periods of time, and have experience with him or her in multiple contexts (McConaughy, 1993). Parents are familiar with the child's developmental history and can also provide important information regarding family psychiatric history (Richters, 1992). Additionally, parents are in the position to observe "clinically important base rate phenomena - such as fire setting, suicide attempts, or episodes of frank psychosis - by virtue of their extended contact with the youth" (Youngstrom, Findling, Danielson, & Calabrese, 2001, p. 267).

Another potential source of information is the child. Phares, Compas, and Howell (1989) argued, "A unique feature of children's self-reports is that they may reflect a compilation of behaviors across different situations, including their behavior at home, in school, and with their peer group" (p. 68). In contrast, parents and teachers observe a child in one situation and may report on behaviors that are unique to that situation.

However, many researchers have questioned children's ability to accurately portray their feelings or describe their own behaviors. For instance, Kazdin (1981) noted that the differing levels of language ability and cognitive development in children could considerably influence the way they perceive questions or test items, as well as the answers given. In addition, severity of symptomatology and/or intellectual limitations may also influence the type of information obtained from self-report data in children. Additionally, Wingenfeld (2002) argued that younger children might show limited insight

into the problematic nature of their behavior, especially if the behavior is disruptive or aversive to others. Children may be limited in terms of their emotional awareness and self-reflection and may tend to respond in what they consider a socially desirable manner. Children and adolescents who seek social acceptance may tend to underreport symptoms, whereas those who feel overwhelmed may overreport their symptoms (Myers & Winters, 2002). Although the accuracy of data obtained from children's self-reports may be biased or unreliable (Martens, 1993), the information can be used to complement other sources of information within the context of a multimethod, multisource assessment model (Elliot, Busse, & Gresham, 1993).

By their nature, rating scales rely on an informant's perceptions of particular behaviors or symptoms. Perceptions of each possible source of information are likely to vary, and many factors can influence one's reports of a child's behavior. Strupp and Hadley (1977) argued that differences in perceptions of child behavior problems might be due to the different ways parents, teachers, and children define mental health and the types of feelings and behaviors they use to define clinical problems, with parents and teachers focusing on overt behaviors and children centering on their own subjective experiences. Informants may also differ in their sensitivity and tolerance of children's behavior problems, their ability to deal with behavioral difficulties, and their expectations regarding children's behavior. In addition, informants' reports may be influenced by the nature and amount of their contact with the child, as well as the settings and situations in which they observe and interact with the child (Edelbrock, 1983). The personality and

personal characteristics of respondents may also influence their report or ratings (McConaughy, 1993).

In particular, many investigations have examined the contribution of parents' (mostly mothers') adjustment to parental perceptions of children's maladjustment. It has been argued that the correlation between parent and other informants' reports is influenced by the parent's own pathology. Several studies have shown that depressed mothers reported more problem behaviors than nondepressed mothers (Conrad & Hammen, 1989; Richters & Pelligrini, 1989; Webster-Stratton & Hammond, 1988). However, based on a review of these studies, Richters (1992) noted that teachers also reported more problems for children of depressed mothers. Richters concluded that the difference in the mothers' reports reflected the children's actual behavior, rather than distortions in the perceptions of the depressed mothers.

Other studies have shown that parental perceptions of child behavior are a function of the combined influence of parents' psychological symptoms and child behavior (Brody & Forehand, 1986; Phares et al., 1989). Including both fathers and mothers in their analysis, Phares and colleagues found moderate and significant relationships between parents' reports of their own symptoms and their reports of their child's behavior problems. Both mothers' and fathers' self-reports of psychological symptoms (emotional and somatic distress) and their children's self-reports of behavior problems predicted maternal and paternal reports of children's problems.

In addition to personal characteristics of the informant, another factor that may influence reports of a child's behavior relates to the context within which the child is

observed. Phares et al. (1989) argued that reports of informants who observe children in only one situation (e.g., a teacher who only observes a child in the school environment) may reflect child behaviors that are unique to that situation. They noted that due to this situational specificity of behavior, one would expect only modest associations between informants in different situations with regard to ratings of behavioral and emotional problems.

Research supports this assumption, having shown only modest levels of agreement across different informants. In a frequently cited meta-analysis, Achenbach, McConaughy, and Howell (1987) demonstrated that correlations between informants varied depending upon the context in which they observed the child. Reports from informants in similar roles (e.g., two parents or two teachers) were much more strongly correlated (average correlation = .60) than reports from informants in different roles (e.g., teacher and parent; average correlation = .28). The authors argued that such modest cross-informant agreement did not mean that a particular informant's reports were biased or inaccurate. Rather, modest correlations indicated that evaluators should not substitute one source of information for others and should instead gather information from more than one source.

Verhulst, Koot, and Van der Ende (1994) argued that because each informant may provide valid, although different, information, reports from different sources should be gathered when evaluating a child's social-emotional functioning. They contended that the modest correlations between informants reported by Achenbach et al. (1987) reflected

differences in informant characteristics and differences in informant-child interactions, as well as the situational variability of children's behavior.

Additionally, Hinshaw, Han, Erhardt, and Huber (1992) asserted that the type of problem behavior must also be considered when examining the correspondence between different informants' ratings of behavior. They found that parent ratings independently predicted observed isolation and withdrawal (internalizing behavior problems). Teacher ratings, on the other hand, independently predicted observed noncompliant and aggressive behavior (externalizing behavior problems).

In another investigation, Costenbader and Keller (1990) found that correlations between the behavior problem scales of the CBCL (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986) for the same information source (either parents or teachers) were very high. However, correlations between sources both within and across scales for behavior problems and social competence measures of the CBCL were low. Measures of children's behavior problems and social competencies in home and school settings, as perceived by their parents and teachers, were relatively independent in this study. Costenbader and Keller argued that low between-source correlations indicated that reported behavior problems and social competencies were situationally specific. They then concluded that information from both home and school settings was useful for obtaining a comprehensive picture of a given child's behavior problems and social competencies.

Verhulst, Dekker, and Van der Ende (1997) made similar conclusions. In their longitudinal study of Dutch children and adolescents, the ability of parents, teachers, and

children's self-reports of problem behavior to predict signs of maladjustment over a four-year period was evaluated. The authors found that "each informant made its own unique and indispensable contribution" (p. 75) to the prediction of maladjustment over time. In another study, Verhulst et al. (1994) showed that the combination of parent and teacher information resulted in an improvement of predictive power of poor outcomes (e.g., academic and school behavior problems, police contacts, and suicidal behavior) over that of information from one source alone. They argued that "this additive effect" (p. 544) indicated the importance of obtaining assessments from both parents and teachers.

Likewise, Costenbader and Keller (1990) found that including more than one informant source and measure improved the identification rates of educationally disabled and nonreferred children. The prediction rate for emotionally disabled children was substantially higher when information from both parents and teachers was used and when both the behavior problem and social competence sections of the CBCL were used.

In a six-year longitudinal study, Achenbach, Howell, and McConaughy (1995) reported that referral to mental health services was accurately predicted by combined parent/teacher and parent/child sources of information about problem behavior. Results of each of these studies supported the argument for including information from multiple sources in a comprehensive clinical assessment.

Whereas these studies compared ratings of different types of informants (e.g., parent and teacher), other studies have focused on agreement between similar informants (i.e., mother and father). Consistent with findings of Achenbach and colleagues (1987), results of a meta-analytic review by Duhig, Renk, Epstein, and Phares (2000) indicated a

high degree of correspondence between maternal and paternal ratings of problem behaviors. The findings indicated, “One parent would typically provide a reasonable estimate of what would be provided by another parent who sees a child or adolescent under similar conditions” (p. 448). However, age, gender, and socioeconomic status were found to moderate the degree of correspondence of maternal and paternal reports. Other studies comparing maternal and paternal ratings of behavior have found that ratings from each source contained both a shared behavioral perspective as well as a component that was unique to the rater (Rowe & Kandel, 1997; Van der Valk, Van den Oord, Verhulst, & Boomsma, 2001).

In short, “it is clear that no one type of informant typically provides the same data as any other type of informant” (Achenbach et al., 1987, p. 227). Each informant provides unique and valuable information. In practical terms, these findings indicate the importance of conducting assessments using rating scales across settings and using multiple informants whenever possible, a commonly recommended practice (McConaughy, 1993; Merrell, 2000, 2001). This assessment practice is likely to produce a summative picture that helps to overcome the limitations inherent in making judgments based on information provided by single raters in specific settings.

Rating Scale Items

Informants differ in ways that influence the ratings or reports they provide. However, features of the rating instrument itself may also affect how and what is reported. Edelbrock (1983) noted that the utility of rating scales depends to a large extent upon the specific items, materials, and rating procedures that comprise the instrument.

Formal aspects of the rating scale, such as item form and content, can influence results. The central question is whether the respondent's understanding of the item corresponds with what the assessor intended (Schwarz, 1999). Characteristics of the item can lead to error. For example, difficulty with reading comprehension means that differences in raters' verbal and reading skills may confound results.

Item ambiguity indicates that different raters may be responding to different aspects of the item (Kline, 1993). "Items that are ambiguous allow respondents to respond to them systematically using their own heuristic or respond to them randomly" (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, p. 882). This in turn introduces common method variance (i.e., variance due to the measurement method rather than the construct or trait being measured). Method variance is a source of measurement error that threatens validity. It is, therefore, an important issue in scale development and use, yet it may be overlooked or discounted. According to Cote and Buckley (1987), many researchers assume that a measure is valid and has negligible measurement error when in fact measurement error often accounts for most of the variance in a measure. Thus, it is important that special consideration be given to potential sources of measurement error, such as item comprehensibility and clarity.

Individuals differ in how they perceive and integrate information (Hartel, 1993). Vague or ambiguous items "often require respondents to develop their own idiosyncratic meanings for them" (Podsakoff et al., 2003, p. 883). Therefore, the validity of the information is threatened to the extent that items do not have uniform meaning for all

respondents (Hopkins, 1998). This is particularly problematic when technical jargon is used.

“Undoubtedly, some of the complexity that exists in questionnaire measures results from the fact that some constructs are fairly complex or abstract in nature” (Podsakoff et al., 2003, p. 883). Hofstee (2002) noted that items on personality scales often revolve around “a trait-descriptive adjective” (p. 23), yet in everyday life, individuals and personalities are often discussed in terms of behaviors signified by verbs. For example, a parent would more likely describe how a child *argues frequently with adults* or *will not follow rules* rather than describe the child as *oppositional* or *antisocial*. Hofstee likened the use of such trait adjectives to official discourse such as psychological reports in which they “may serve the purpose of diplomatic vagueness” (p. 23) but noted that they are seldom used in everyday language.

When the goal is clarity and universal understanding by all respondents, items should be devoid of such technical terminology. Instead, items should be written in simple, straightforward language readily understood by people with a modest reading level (Clark & Watson, 1995). Low-frequency words, trendy expressions, and colloquial speech should be avoided (Comrey, 1988). Kline (1993) also argued against using feeling terms, whenever possible, in favor of more precise behavioral descriptions, noting that feeling terms are highly subjective and increase the problem of item ambiguity. Errors in rating are reduced when the characteristic or behavior being rated is described as objectively as possible with reference to some actually observed activity (Aiken, 1996). The more explicit the trait being measured (e.g., observable behavior) and the less

inference required by the respondent, the lower the variance in the item due to rater bias (Hoyt & Kerns, 1999).

Clearly defined items are more readily understandable and consistently interpreted by all respondents. However, when faced with vague or ambiguous items, respondents may utilize other aspects of the rating instrument scale to guide their interpretation. Schwarz (1999) described how respondents sometimes refer to the response alternatives (e.g., numeric values on the rating scale) or to preceding items to help them make sense of what is being asked. Differential interpretation again poses a potential threat to validity. However, the better defined the symptom or behavior under question, the less influence the response alternatives have on the respondents' interpretation of the item (Schwarz & Scheuring, 1992). These findings support recommendations for scale developers to define unfamiliar terms, avoid vague concepts, use examples to illustrate abstract concepts, and avoid complicated syntax (Tourangeau, Rips, & Rasinski, 2000). Behaviors or traits being rated should be described in precise behavioral terms that will be understood similarly by all respondents (Aiken, 1996).

Rating Scale Format

In addition to item characteristics, rating scale format is another critical feature of the rating instrument. Rating scale formats differ in their structure, complexity, and subjectivity (Hartel, 1993). Two common response formats are dichotomous (e.g., present/absent, yes/no, and true/false) or multipoint Likert rating scales with three or more options (Clark & Watson, 1995). Typically in multipoint scales, each numerical value in the rating format is keyed or anchored to a descriptor (for example 0 = *never*, 1 =

sometimes). Popular instruments, such as the CBCL and BASC, employ a multipoint rating scale.

Comrey (1988) argued against use of dichotomous response formats, noting that multipoint formats are “more reliable, give more stable results, and produce better scales” (p. 758). He argued that the two-choice format is more likely to produce highly unbalanced item response distributions, which then produce distorted correlational results. However, Clark and Watson (1995) reasoned that careful examination of item response rates and subsequent elimination of those items with extreme frequencies during scale development would help to avoid this problem. They also noted that dichotomous formats present an additional advantage of quicker response time.

Although multipoint formats may be reliable, they can be problematic in other ways (Reise, Waller, & Comrey, 2000). First, they may produce individual differences in individuals’ willingness to endorse the extremes of a rating scale (Chen, Lee, & Stevenson, 1995), a source of error Anastasi and Urbina (1997) referred to as an error of central tendency. While some respondents prefer the neutral or middle point on a scale, Kline (1993) also noted that a proportion of respondents endorse the extremes responses simply because they are extreme. Both situations represent response sets and potential sources of error that are more common to multipoint scales.

Another potential source of error relates to the semantic problems many response options pose. Descriptive adjectives and numerical anchors in multipoint scales are abstract in nature, and as such, a rater must provide meaning to them (Champion, Green, & Sauser, 1988). Many multipoint scales use vague or subjective rating options

such as *seldom*, *sometimes*, and *frequently*, and the explicit meanings that people attach to such words can vary considerably (Hopkins, 1998). Different respondents might assign different meanings to the same ratings. Different respondents might also interpret and use the same rating to signify different frequencies of the same behavior (Burns, Walsh, et al., 2001; Schwarz, 1999). The validity of the information is decreased to the extent that the descriptive terms applied do not have consistent meaning across individuals (Hopkins, 1998). Instead, ratings should be expressed in a form that will be uniformly interpreted by all raters (Aiken, 1996; Anastasi & Urbina, 1997).

Multipoint rating scales include response options that represent differences in degree. A higher level of inference is needed in making ratings when more possible rating options are involved, which increases the difficulty in reliably discriminating among the various rating levels (Merrell, 2000). These types of rating formats may require respondents to make distinctions that cannot be made, introducing variance that ultimately weakens the validity of the scale (Clark & Watson, 1995). As the rating task becomes more complex, the resulting data become less reliable (Kazdin, 1977). Scales requiring an excessive amount of inference in making discriminations among rating points should be avoided (Merrell). With their more simplistic format, dichotomous scales require less inference and may be less prone to problems related to comprehension and differential interpretation than multipoint scales. Dichotomous scales may minimize the potential for individual differences in scale use or response bias (Reise et al., 2000).

Thus, the response format has important implications for how and what information is reported. Aiken (1997) noted, "All test items represent procedures for

obtaining information about individuals, but the amount and kinds of information vary with the nature of the tasks posed by the items" (p. 28). Rating scales employ different response formats as well as different approaches for assessing behavior.

Defining Problem Behavior

Most behavior rating scales, such as the CBCL and BASC, present lists of problem behaviors or symptoms. A respondent, such as a parent or teacher, indicates the severity or frequency of various behaviors based on his or her observations and impressions of the child. However, there are limitations to this type of approach.

First, rating the severity of a behavior involves a complex decision-making process because perceived severity may be a combination of how often the behavior occurs, what the intensity is, and how troublesome it is to the rater. By combining these aspects of the behavior into one rating, information is lost, and unexplained variance is introduced (Wilson & Prentice-Dunn, 1981).

Burns, Walsh, et al. (2001) criticized use of vague frequency rating procedures. They argued that use of rating terms, such as *very often*, probably measure both symptom occurrence and impairment due to symptom occurrence, thereby confounding different properties of the same symptom.

Additionally, although typical scales may ask for information regarding the frequency of symptoms, variation in the occurrence and severity of problem behavior across situations cannot be determined (DuPaul & Barkley, 1992). The question of pervasiveness of problem behavior across situations cannot be assessed. Information gained from typical rating scales does not clarify whether the symptomatic behavior is

isolated to specific circumstances (situationally specific) or whether it is pervasive across varied circumstances. On typical rating scales, psychopathology is understood in terms of the number of related symptoms reported and/or the frequency or severity of symptoms reported.

An alternative approach involves cross-situational assessment of behavior. This type of approach emphasizes that behavior that is pervasive across multiple contexts or situations is more indicative of maladjustment than behavior that is only observed in isolated circumstances. Behavior seen only situationally is more likely to be random or reactive (McDermott, 1994).

A focus on cross-situational occurrence of problem behaviors is evident in recommendations for defining and identifying the educational disability of emotional disturbance (behavior disorder). For example, the Council on Children with Behavior Disorders (CCBD; 1987) noted that problematic behaviors should demonstrate cross-situational stability and cross-informant agreement. Likewise, Gresham (1985) argued that a comprehensive definition of behavior disorder should include a criterion that the problem behavior be demonstrated across situations and methods.

Several studies have shown important differences between situational and cross-situational behavior problems. For example, Tripp and Luk (1997) investigated differences between children with characteristics of ADHD. They found that “situationally hyperactive” children had fewer behavioral and social difficulties than children with pervasive hyperactivity.

Horn, Wagner, and Ialongo (1989) also emphasized the importance of pervasiveness of behavior in their study of children with ADHD. They found no sex differences in the expression of primary (impulsivity, inattention, and overactivity) and secondary (e.g., learning problems, peer problems) symptomatology of ADHD. This finding was contrary to prior work that indicated important differences between male and female participants with ADHD. The authors argued that a possible reason for this difference centered on the requirement of pervasiveness (i.e., that ADHD symptoms be pervasive across settings) for defining the presence of ADHD in this study. They argued that a group of children with nonpervasive ADHD appeared to be a more heterogeneous group with more diverse characteristics. Therefore, this increase in heterogeneity might have increased the probability that spurious factors led to the finding of gender differences in previous research.

Similarly, work by Barkley and colleagues highlighted the importance of assessing pervasiveness of problem behaviors across situations. Barkley and Edelbrock (1987) noted that an important aspect of behavioral assessment is to examine environmental factors that shape and maintain behaviors. However, these distinctions are difficult to make with typical rating scales that require informants to give global impressions of behaviors across situations. Barkley and Edelbrock argued that collapsing impressions across time and settings might blur important distinctions about environmental factors that might contribute to or influence particular behaviors. They noted that for treatment planning purposes it is important to delineate where problem

behaviors occur and for diagnostic purposes to understand the pervasive nature of problem behaviors, information that cannot be obtained from most rating scales.

Barkley and Edelbrock (1987) proposed an alternative approach that assesses variations in problem behavior across situations within home and school settings. In doing so, they developed the Home and School Situations Questionnaires (Barkley & Edelbrock). Revised versions of these instruments were developed to evaluate children suspected of having ADHD and to document the severity and pervasiveness of attention difficulties across situations within home and school (DuPaul & Barkley, 1992).

Although these instruments emphasize the value of cross-situational behavior, they are narrowband measures and are not designed to differentiate types of disturbance or problems, which is often the goal of assessment.

Adjustment Scales for Children and Adolescents (ASCA)

A behavior rating scale that utilizes a multisituational approach, as well as addressing multiple aspects of psychopathology, is the Adjustment Scales for Children and Adolescents (ASCA; McDermott, 1994). The ASCA is a standardized instrument for completion by classroom teachers. The ASCA includes 97 problem and 26 positive behavioral indicators, presented within the context of specific social, recreational, or learning situations in which the child's functioning may be observed. Rather than judge the general frequency or intensity of a symptom as on other rating scales, on the ASCA, the respondent is asked to indicate which specific behaviors (if any) characterize the child's functioning with respect to that situation. Given situations involve authority, peers, smaller or weaker youths, recreation, learning, and confrontation (McDermott,

1993, 1994). Most behavior rating scales do not specify circumstances or assess behaviors across multiple contexts (Canivez & Rains, 2002). On the ASCA, psychopathology is defined through demonstration of problem behaviors across multiple situations, rather than general or global ratings of problem behavior (McDermott, 1993, 1994).

The ASCA was developed as a major revision of the British and Canadian versions of the Bristol Social Adjustment Guides (BSAG; Stott, 1966; Stott, Marston, & Neill, 1975). Similar to the design of the ASCA, the BSAG includes descriptors of behavior in reference to specific social or learning situations in which the child's adjustment may be observed. Each behavioral indicator is scored as either present or absent. Additional item content for the ASCA was adapted from the Learning Behaviors Scale (LBS) and Study of Children's Learning Behaviors (Stott, McDermott, Green, & Francis, 1988).

Based on factor analyses conducted during its development, the 97 problem behavior items on the ASCA fall singularly on one of six core syndromes or two supplementary syndromes. The six core syndromes, which have been found to be reliable across age, gender, and race/ethnicity, are Attention-Deficit Hyperactive, Solitary Aggressive (Provocative), Solitary Aggressive (Impulsive), Oppositional Defiant, Diffident, and Avoidant. The two supplementary syndromes, which have been shown to be reliable for only certain groups, include Delinquent and Lethargic (Hypoactive). The six core syndromes also combine to form two composite or overall adjustment scales. The Attention-Deficit Hyperactive, Solitary Aggressive (Provocative), Solitary

Aggressive (Impulsive), and Oppositional Defiant syndromes combine to form the Overactivity adjustment scale. The Diffident and Avoidant syndromes form the Underactivity adjustment scale (McDermott, 1993, 1994). The Overactivity and Underactivity composite scales have been described as similar to the externalizing (or conduct problem) and internalizing (or withdrawal) dimensions found on most other measures of child psychopathology (Canivez et al., 2001).

Studies conducted during development and standardization of the ASCA indicated strong internal consistency, interrater reliability, and test-retest reliability for the core syndromes and global adjustment scales (McDermott, 1993, 1994). Canivez (2004) recently reported similar internal consistency estimates in an independent investigation.

Using a more diverse sample of exceptional students, Watkins and Canivez (1997) reported strong interrater agreement for the core syndrome and global composite scale scores of the ASCA. Strong interrater agreement has also been shown for both the Syndrome Profile (Canivez & Watkins, 2002) and Discriminant Classifications (Canivez, Watkins, & Schaefer, 2002). (Syndrome Profile Interpretation and Discriminant Classification Interpretation are two options available for facilitating interpretation of ASCA *T* scores and are described in detail in the ASCA manual [McDermott, 1994]). Canivez et al. (2001) replicated the test-retest stability of the core syndromes, supplementary syndromes, composite adjustment scales, syndrome profiles, and discriminant classifications.

The ASCA manual (McDermott, 1994) outlines substantial validity evidence. Exploratory and confirmatory analyses supported its factor structure (McDermott, 1994). In an independent factorial validity study, Canivez (2004) found that the six core syndromes of the ASCA produced a two-factor solution, similar to McDermott's (1993, 1994) findings. Results indicated factorial independence of the Overactivity and Underactivity dimensions ($r = .08$) and low to moderate intercorrelations among the core syndromes (r s ranging from $-.04$ - $.60$). Canivez asserted that similar Externalizing and Internalizing composites on the CBCL and BASC are moderately correlated and that many of the subscales on these instruments share substantial covariance. Whereas these characteristics of other scales confound analysis, Canivez argued that the relatively independent factors and lower core syndrome intercorrelations of the ASCA allow greater interpretability.

Convergent and divergent validity studies by McDermott (1993, 1994) comparing the ASCA with the Conners Teacher Rating Scale (CTRS) and the CBCL found significant correlations. Canivez and Bordenkircher (2002) and Canivez and Rains (2002) found significant convergent and divergent validity for the ASCA Overactivity and Underactivity syndromes when compared to the Preschool and Kindergarten Behavior Scales (Merrell, 1994). McDermott (1995), McDermott and Schaefer (1996), and McDermott and Spencer (1997) have presented additional evidence of construct validity.

In addition, the ASCA has been shown to detect emotional disturbance with similar accuracy for males and females, for adolescents and younger children, and for African-American and white children (McDermott et al., 1995). The ASCA was also

shown to effectively distinguish students with emotional disturbance from normal students and from those identified with learning disabilities, speech/language disabilities, mental retardation, and gifted abilities (McDermott et al., 1995).

As described, the ASCA promotes a multisituational approach for defining psychopathology. Observers indicate which specific behaviors most typically represent given circumstances. The presence and intensity of maladjustment are derived from pervasiveness across multiple situations (McDermott, 1994). However, “it could not be assumed that a high score on any dimension typically represented observation of the specified pathology over multiple situations since more than one example of the same type of behavior could be observed in some situations” (p. 40).

To explore the multisituational aspect of behavior, a subgroup of participants in the norm sample were studied. The number of different situations in which similar problems arose for a relatively maladjusted portion of the norm sample was tallied. For each core syndrome and composite scale, participants having *T* scores greater than or equal to 70 were considered maladjusted. Distributions were computed for the respective situations in which core syndrome or composite scale behaviors were observed. For the core syndromes, McDermott (1994) reported that on average participants deemed maladjusted for attention-deficit hyperactive behavior displayed that behavior in 11.9 out of 16 different possible situations. Aggressive-provocative maladjustment was evident in 5.1 out of 9 situations. Aggressive-impulsive maladjustment was seen in 2.5 out of 6 situations. Oppositional-defiant behavior was found in 6.5 out of 11 situations. Diffident behavior problems were evident in 8.9 out of 13 situations. Avoidant problems were

displayed in 5.6 out of 9 situations. For the composite scales, general overactive maladjustment was supported in 17.3 out of 26 different situations, and underactive maladjustment was evident in 10.1 out of 14 situations. McDermott (1994) concluded that these results supported the assertion that youth psychopathology as measured by the ASCA is generally multisituational in nature.

In addition to its alternative approach, the ASCA has several other notable characteristics that distinguish it from more traditional behavior rating scales. First, the behavioral indicators were designed to provide clear content meaning for respondents. As is the case with the BSAG, the items on the ASCA were drawn from the language of educators, are behaviorally oriented, and avoid clinical terminology (McDermott, 1981, 1993, 1994). McDermott noted that these characteristics helped to reduce or eliminate asking respondents to judge the meaning of behaviors or the nature of children's thoughts and feelings. Also, presenting behaviors in reference to specific situations helps provide a context for the respondent's ratings. As such, the ASCA follows previously presented recommendations regarding specificity and clarity of item content. Most other behavior rating scales do not specify circumstances (Canivez & Rains, 2002). Rather, items on other rating scales are decontextualized and call for respondents to infer meaning and judge more global or general symptoms.

Canivez (2001) described the multisituational assessment approach, along with the large nationally representative standardization sample and multivariate test interpretations (Syndrome Profile and Discriminant Classification analyses), as notable strengths of the ASCA. However, as a teacher-report instrument, it was recommended as

only one method in a multisource assessment. Additionally, in a separate review, Schowengerdt (2001) noted that other broadband rating scales, such as the BASC, have an advantage over the ASCA in that they offer corresponding versions (i.e., parent, teacher, and child/self forms), allowing for multiple perspectives and a more comprehensive assessment.

If a complementary parent version of the ASCA was introduced, it could address this limitation. It could potentially add to the growing behavior rating scale technology with merits of a unique multisituational approach and accordance with a recommended multisource, multimethod assessment model.

Adjustment Scales for Children and Adolescents – Home (ASCA-H)

The Adjustment Scales for Children and Adolescents – Home (ASCA-H; Watkins & McDermott, 2002) is currently in development. The ASCA-H is an objective rating instrument to be completed by a parent or caregiver. It is designed for use with children and adolescents ages 5 to 17. Developed as a counterpart to the ASCA, the ASCA-H has similar item content and format. The ASCA-H contains 202 behavioral descriptor items. One hundred eighty of the total items are presented in reference to 33 specific situations in which parents can observe the child's behaviors. Examples of the situations include coping with responsibility, tasks of daily living, informal or unorganized play, approach toward school, and illicit activities. Twenty additional items are presented under the heading, "Other Behaviors that Cause Concern." These items include extreme or atypical behaviors, such as suicidal ideation, specific fears, running away, setting fires, exhibiting tics, and hearing voices.

Similar to the ASCA, the ASCA-H introduces multisituational variables in an attempt to refine the measurement of psychopathology. However, thus far no studies of this instrument have been reported. As evidenced by the collective research presented on the ASCA, the cross-situational approach advocated by McDermott and colleagues has been effectively applied to a teacher rating instrument, yet an unanswered question is whether this approach can also be applied to a parent version.

This question involves the ability of parents to rate children's behavior. Results of some studies have challenged the validity of parent ratings. For instance, in a comparison of parent and teacher questionnaires of children's social-emotional functioning, Kresanov, Tuominen, Piha, and Almqvist (1998) found that both instruments had adequate reliability and discrimination power. However, teacher responses were superior to parent responses in their ability to discriminate psychiatric disturbances. The investigators suggested that parents in general might not be as well versed in rating children's behavior as teachers. They described how teachers are "educationally a more homogenic group and could be considered more reliable in understanding and answering written questions than are parents from all social classes and educational backgrounds" (p. 91).

Teachers have been described as having other characteristics that distinguish them from parents and that may contribute to their ability to rate behavior. It has been argued that "the normalized setting of school" (Myers & Winters, 2002, p. 115) and cumulative experience of teachers over time with many students in multiple contexts provides teachers with a solid framework and background from which to compare children's

behaviors (Canivez et al., 2001). Parents, on the other hand, may be more restricted or biased in their perspective, with many having little experience with children other than their own. Their frame of reference from which to compare children's behaviors may then be limited. Parents may also have less familiarity with the age-appropriateness of children's skills and behavior (Schneider & Byrne, 1989), thus making it more difficult for them to judge the normality of behavior. Teachers observe a similar class of behaviors in more structured, less ambiguous situations than parents, which might improve the reliability of their behavioral ratings (Breen & Altepeter, 1991).

Although teachers are considered by many to be more accurate raters of child behavior (Breen & Altepeter, 1991; Kamphaus & Frick, 1996), the demands of education are for multiple sources of information. Proposed definitions of behavior disorder for special education eligibility stress that behaviors should show stability across situations and informants (CCBD, 1987; Gresham, 1985). Information from many knowledgeable individuals contributes to a more comprehensive picture of the child across a variety of situations (Verhulst & Akkerhuis, 1989), which could help distinguish a truly emotionally disturbed child from one with more situation-specific problems (Wright & Piersel, 1992). In addition, the additive effect of combining parent and teacher information has been shown to improve prediction power of poor outcomes (Verhulst et al., 1994), referral for mental health services (Achenbach et al., 1995), and identification rates of emotionally disabled students (Costenbader & Keller, 1990). Experts continue to recognize obtaining information from multiple sources as a best practice in a

comprehensive assessment of behavioral and social-emotional functioning (McConaughy & Ritter, 2002).

Parents are potentially an important source of information in a multisource assessment. They have known their child for longer periods of time, typically have the highest level of involvement, and have experience with the child in different contexts (McConaughy, 1993). They have knowledge and experience with the child that others do not, information that is integral to a truly comprehensive assessment, but their ability to accurately rate child behavior has been criticized. It may not be sufficient, however, to conclude that parents are inferior raters of behavior. Rather, it may be the procedures and methods for obtaining parent report that should be questioned.

Studies comparing teacher and parent ratings (e.g., Hinshaw et al., 1992; Schneider & Byrne, 1989) have often been based on traditional measures, such as the CBCL, which require a rating of symptom frequency or severity. It has been argued that the training, knowledge, normative perspective, and experience of teachers working with different kinds of children across different contexts may contribute to the superiority of their ratings over those of parents. However, it may be that the traditional rating scales that employ frequency or severity rating formats are better suited to the characteristics of teachers. That is, because of their comparative experience and knowledge of the age-appropriateness of behaviors, teachers might be better able to judge the relative frequency or severity of child behaviors. They might be better able to make the distinctions required by such multipoint scales, leading to more reliable ratings. Given their different experiences and perspectives, parents may be better able to report on children's behavior

if provided with a different vehicle or means of assessment, particularly one that does not require the inference inherent in instruments using a frequency or severity rating format.

The original ASCA built upon limitations of traditional behavior rating scales with its unique format and multisituational approach. ASCA items are behaviorally-focused and devoid of clinical terminology (McDermott, 1981, 1993, 1994). McDermott argued that these characteristics helped to reduce the necessity of respondents to judge the meaning of behaviors or the nature of children's thoughts and feelings. Also, behaviors are presented in relation to specific situations to provide a contextual reference for the respondent's ratings. The specificity and clarity of item content and response method of the ASCA may be particularly beneficial to parents. The unique assessment approach of the ASCA could create a better parent rating instrument.

By introducing a new format for obtaining parent report and assessing behaviors across different contexts, the ASCA-H could make a unique contribution to the field of behavioral assessment. In complement to the ASCA, it would follow a recommended multisource, multimethod assessment model and proposed guidelines for defining behavior disorders (CCBD, 1987; Gresham, 1985). However, the potential of the ASCA-H must be empirically supported, and no investigations thus far have been reported.

The central concern is one of validity. According to the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), validity is defined as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p. 9). Validity is described as the most

fundamental concern in the development and evaluation of tests. Validation involves the accumulation of evidence to provide a sound, scientific basis for proposed interpretations of test scores.

Test validation is considered by many to be the most important step in the development and use of tests because it is the process by which test scores take on meaning (Benson, 1998). Within the context of test validation, construct validity concerns whether a given measure, or operational definition, really assesses the underlying conceptual variable, or construct, that the measure is intended to represent (Bryant, 2000).

Loevinger (1957) and Nunnally (1978) described three aspects of construct validity. These components were employed by Benson (1998) to illustrate the stages in the process of construct validation. First, the substantive component involves specifying the theoretical domain of the construct and defining the variables that reflect the construct. Secondly, the structural aspect relates the variables to the structure of the construct by determining the relationships between variables themselves and their relationships to the construct. According to Benson, it is the third component, the external aspect, which “begins to give meaning to test scores by determining whether or not the measures of a given construct relate in expected ways with measures of other constructs” (p. 11).

In the ASCA-H’s current stage of development, the most pertinent aspect of validity to be addressed is the structural component. “Studies conducted under the structural stage are referred to as internal domain studies because they involve only the

observed variables for a given test. The objective of an internal domain study is to determine the extent to which the observed variables covary among themselves, and how they covary with the intended structure of the theoretical domain” (Benson, 1998, p. 13). The issue of which particular dimensions underlie a measurement instrument is known as factorial validity or structural validity (Bryant, 2000) and is one of five recognized types of validity evidence defined in the Standards (AERA, APA, & NCME, 1999).

The internal structure of the ASCA-H has yet to be examined. Therefore, given the importance of this stage of construct validation and lack of exploratory studies of the ASCA-H, the proposed investigation will examine the structural validity of the ASCA-H. The specific question to be addressed is: What are the underlying dimensions of psychopathology, as assessed by the ASCA-H?

Method

Participants

Parents and caregivers of students attending four elementary schools, one middle school, and one special education school in a large suburban school district in the mid-Atlantic region were asked to provide behavior ratings of their children. A total of 426 parents/caregivers (32 males and 393 females) participated. Gender information for one participant was missing. Eighty-one percent of the respondents were mothers; 7.3% were fathers; and 11.5% were identified as “other” (e.g., step-parent, foster parent, or grandparent).

As outlined in Table 1, parents reported on a total of 426 students (240 males and 186 females). The mean age of the participants was 9.33 years ($SD = 2.57$), with a range of 5 to 15 years. Data pertaining to students’ SES and race/ethnicity could not be obtained as such information was disallowed by the school district during the research approval process. However, the schools from which these students were drawn were composed of on average 54.6% White/Caucasian students, 34.2% African-American students, 2.9% Hispanic students, 6.6% Asian-Pacific students, and .5% American-Indian students. Table 2 presents data pertaining to the school populations.

Instrument

The ASCA-H (Watkins & McDermott, 2002) is an objective rating instrument completed by a parent or caregiver of children and adolescents aged 5 to 17 years. It contains 202 behavioral descriptor items. One hundred eighty of the items fall within 33 specific situations in which a parent can observe the child’s behaviors. Examples of the

Table 1

Student Age and Sex (N = 426)

Age	Male		Female	
	N	%	N	%
5	19	4.5	12	2.8
6	21	4.9	15	3.5
7	27	6.3	22	5.1
8	34	8.0	23	5.4
9	27	6.3	25	5.9
10	33	7.7	19	4.5
11	18	4.2	22	5.1
12	25	5.9	23	5.4
13	22	5.1	15	3.5
14	10	2.3	8	1.9
15	2	0.5	0	0
Missing	2	0.5	2	0.5
Total	240	56.3	186	43.7

Table 2

Population Characteristics

School	Potential		% Students in Special Education	SES ^a	Racial Composition (Percent)				
	Number of Students	Grades			African American	Hispanic	Asian Pacific	American Indian	
1	482	K-5	12.2	27.5	88.0	11.0	0.0	0.0	0.0
2	391	K-5	12.5	10.8	62.7	19.4	3.6	13.8	0.0
3	612	6-8	15.9	61.7	59.5	36.6	0.0	0.0	0.0
4	51 ^b	K-5	14.4	25.6	44.2	19.9	9.6	26.0	0.0
5	46 ^b	Pre-K-5	9.5	79.0	14.9	79.8	3.9	0.0	0.0
6	140	Pre-K-7	100	71.6	58.0	38.4	0.0	0.0	3.0

^aSES is reported in terms of the percentage of students participating in the Free-and-Reduced Meal Program.

^bOnly students who were referred to the school's Instructional Support Team (IST) could participate.

situations include coping with responsibility, tasks of daily living, informal or unorganized play, approach toward school, and troublesome or illicit activities. For each situation, the parent indicates which, if any, of the behavioral descriptors characterized the child's behavior during the past two months. All behavioral descriptions are either marked indicating presence of the behavior or unmarked designating absence of the behavior. Twenty-nine of the 180 descriptions represent positive or adaptive behaviors. The second part of the ASCA-H is a checklist of 22 items grouped as "Other Behaviors that Cause Concern." These items include extreme or atypical behaviors, such as suicidal ideation, running away, setting fires, exhibiting tics, and hearing voices. These items were not intended to be part of the scale but were included to ensure that even rare problems could be detected by the ASCA-H (M. W. Watkins, personal communication, July 30, 2005). Thus far, no studies of this unpublished instrument have been reported.

Procedures

Data collection procedures in Schools One, Two, and Three were the same. Parents or caregivers of students attending these schools were invited to participate. They were recruited through a mass mailing of study materials via U.S. mail. Study materials included an invitation letter, an implied informed consent letter, a copy of the ASCA-H, an entry form (post-card) for a gift certificate lottery, and a self-addressed stamped return envelope. Parents were asked to return study materials within two weeks. After two weeks, a follow-up postcard was sent home to all parents. No further contacts were attempted. Parents had a chance to win one of three \$20 gift certificates to Wal-Mart for their involvement in the study. They were instructed to return the completed

rating scale to the researcher in the provided envelope and return the post-card separately with their name, address, and telephone number to enter the drawing. No personally identifiable information was collected to protect the anonymity of the students.

For Schools Four and Five, participants were parents or caregivers of students referred to the school's Instructional Support Team (IST). The IST functions as a pre-referral intervention group that links school resources to better meet the needs of students with academic, social-emotional, and behavioral problems. Team members work with students who were not identified as eligible for special education and help teachers to identify solutions to instructional issues through classroom-based assessment and collaborative problem-solving (Kovaleski, Tucker, & Stevens, 1996). In these schools, parents or caregivers were asked by the IST to complete the ASCA-H as part of the referral process in order to obtain information regarding the child's overall functioning and to screen for possible areas of difficulty. Access to this secondary data, with identifying information removed, was secured from each school administrator.

School Six is a public special education school for students with severe emotional and behavioral difficulties. The school used the ASCA-H in a schoolwide assessment project to better understand the social-emotional and behavioral functioning of the student population. Access to this secondary data, with identifying information removed, was secured from the school administrator.

Data Analysis

Benson (1998) referred to studies conducted under the structural stage of test development as "internal domain studies because they involve only the observed

variables for a given test. The objective of an internal domain study is to determine the extent to which the observed variables covary among themselves and how they covary with the intended structure of the theoretical domain” (p. 13). One of the statistical methods for conducting an internal domain study is exploratory factor analysis.

The goal of exploratory factor analysis is to identify the underlying dimensions of a domain of functioning, as assessed by a particular measuring instrument (Floyd & Widaman, 1995). “The procedure is exploratory because, presumably, the investigator has no firm a priori expectations based on theory or prior research about the composition of the subscales, and thus the analysis is used to discover the latent variables that underlie the scale” (p. 286). As the goal of this study was to examine the underlying dimensions and internal structure of the ASCA-H, exploratory factor analysis was conducted.

In accordance with procedures specified by McDermott (1994) in analyses of the original ASCA, the 29 positive behavior items of the ASCA-H were omitted from factor analysis “to preserve the clinical orientation of the measures and to avert bipolar, partially ipsative or difficulty factors” (p. 32). Also, the symptom checklist items in the second part of the ASCA-H were excluded, as they were not intended to be part of the primary scale (M. W. Watkins, personal communication, July 30, 2005). Analyses were performed using SPSS 10.0 (2001) and guided by the methods recommended by Comrey and Lee (1992), Fabrigar, MacCallum, Wegener, and Strahan, (1999), Gorsuch (1983, 1997), and Tabachnick and Fidell (1996).

Item properties. Gorsuch (1997) pointed out that methods of factor analysis were developed for analysis of scales. He argued that in contrast to scales, items have unique

properties that require different techniques for factor analysis in order to avoid misleading results. First among these properties is the lower reliability of items. Lower reliability means that items correlate lower than scales and the resulting factors are weaker with a higher percentage of the covariations being error. Secondly, items “contain confounding variance in addition to the construct being measured” (p. 535). This refers to the point that items can be interpreted or responded to differently by individuals, thereby introducing variance that has nothing to do with the construct being measured. In contrast, by adding the confounds found across items and averaging them, scales minimize this error (Gorsuch, 1997).

A third property of items is that their distributions differ from each other, which reduces the correlations among them (Gorsuch, 1997). Scales, however, sum across items thereby averaging the distributions and resulting in higher intercorrelations than items. Analyzing variables with different distributions leads to the possible creation of difficulty factors (Gorsuch, 1997). That is, items load on the same factor as other items with similar distributions. Factors result from items sharing similarly skewed distributions such that positively skewed items tend to correlate more highly with each other than they do with negatively skewed items. Therefore, factors may emerge that reflect distribution properties rather than the underlying construct being measured. These factors are more technical than substantive (Gorsuch, 1997).

A final property of items that impacts factor analysis involves the limited number of item response options (Gorsuch, 1997). This property is related to the point regarding item distributions. As previously described, response options on behavior rating scales

are limited to a certain number of categories, either dichotomous (e.g., *yes/no*) or multipoint response scales (e.g., ranging from *never* to *always*). Item scores are sets of ordered categories rather than a continuous value. It is possible for several individuals to receive exactly the same score on one item. However, if these same individuals do not receive the same score on another item, the two variables cannot correlate perfectly, even if the underlying relationship is perfect. The reduction in correlation occurs most often with dichotomous variables because many individuals receive the same score (Gorsuch, 1983). Special consideration was given to this issue in analysis of ASCA-H items, which are dichotomous variables.

Bernstein and Teng (1989) described that item distributions are nonnormal to the extent determined by the response level. Thus, the correlation between items will be affected both by similarity of their distributions as well as similarity in their content. Two items that assess the same content but differ in their response levels will correlate less than two such items that are similar in their response levels. Spurious (Bernstein & Teng) or difficulty (Gorsuch, 1983, 1997) factors may result based on dissimilarity of response level in addition to those reflecting content. Bernstein and Teng cautioned, “Categorization...can clearly produce false evidence for multidimensionality when criteria applicable to continuous data are used” (p. 467). Therefore, the same criteria or procedures used with continuous variables may not be appropriate in the analysis of categorical items.

Although it is advisable to avoid variables with extreme skews (Gorsuch, 1983), this is not always possible, such as the case of factor analysis of the ASCA-H. Because

they are dichotomous, the items of the ASCA-H are especially prone to the issues outlined by Gorsuch (1997), including low reliabilities, nonnormal distributions, and emergence of technical factors. Therefore, care must be taken in planning and carrying out the factor analysis. This will be seen in the following discussion of procedures that were followed during the different steps of the factor analysis.

In the case of dichotomous data, there is also consideration of which correlation coefficient to apply when performing factor analysis. A phi coefficient is the standard Pearson product moment correlation resulting if both variables are dichotomies. If it is assumed that the dichotomies are actually representing variables that are basically continuous, then tetrachoric correlations can be calculated as estimates of the Pearson product-moment coefficient that would have occurred if the constructs had been measured with continuous variables (Gorsuch, 1983).

Gorsuch (1997) noted that the problem of skewed distributions of dichotomous variables could not be eliminated by the use of tetrachoric correlations. Tetrachoric correlations assume that the variables are normally distributed in their underlying form (Gorsuch, 1983). Because this assumption is often inappropriate, the observed coefficients are often exceptionally large. Gorsuch (1983) argued instead for use of phi coefficients, noting that they are always within the normal range and better represent the relationships in the observed data rather than attempting to generalize to a theoretical variable. Similarly, Nunnally and Bernstein (1994) recommended the use of regular correlations over tetrachoric correlations. These recommendations support McDermott's

(1994) use of phi coefficients in the exploratory analyses of the original ASCA. Phi coefficients were also applied in this study.

Sampling. When designing a study, consideration must be given to the nature and size of the sample (Fabrigar et al., 1999). Clark and Watson (1995) advised use of a large and appropriately heterogeneous sample for the first major stage of scale development. Most importantly, the sample on whom test items are initially tried out must reflect the population for whom the test is intended (Gorsuch, 1997; Kline, 1993).

Similarly, Clark and Watson (1995) asserted that it is important to examine data from diverse samples representing the entire range of the test's target population. They reasoned that many test items would show very different response distributions across clinical and nonclinical samples. This is particularly relevant in the case of a dichotomous scale, such as the ASCA-H, where low item correlations can be expected. If only "normal" students are included in the sample, response distributions are likely to be highly skewed and unbalanced. Clark and Watson also noted that highly unbalanced items are undesirable for several reasons: "(1) When most respondents respond similarly, items convey little information; (2) owing to their limited variability, these items are likely to correlate weakly with other items in the pool and therefore fare poorly in subsequent structural analyses; and (3) items with extremely unbalanced distributions can produce highly unstable correlational results" (p. 315).

During early stages of test development, the sample does not need to be as carefully chosen as in test standardization because it is the relationships between the items rather than the mean scores that are important. "However, the nature of the sample

cannot be ignored in the quest for numbers. If there is insufficient variance in a sample, no factor can emerge. Too great homogeneity is not desirable” (Kline, 1993, p. 136). To identify replicable factors, Reise et al. (2000) argued that samples should have sufficient representation at all levels of trait dimensions in order to accurately estimate the population item intercorrelations.

As a measure of youth psychopathology, the ASCA-H is intended for use with a wide range of students. In consideration of the need for a large, heterogeneous sample in early stages of test development, it was important for the sample in this study to include students with a wide range of behavioral functioning, from typical or adjusted youth to those who are more severely maladjusted. For these reasons, schools that comprised the necessary range of students were selected.

The first three schools were selected to represent normally adjusted students with no or few problem behaviors. Particular segments of Schools Four and Five, those referred to the IST program, were chosen for inclusion in this study. Students were referred to IST due to a variety of academic and/or behavioral concerns. By nature of their having been referred to the IST program, these students were identified as experiencing some level of difficulty. Therefore, this particular group of students would be expected to exhibit a higher level of problem behaviors than a typical school population. By nature of their educational placement, students at School Six represented the upper extreme in terms of adjustment difficulties and would therefore be expected to demonstrate high levels of problem behaviors. Students from the six participating schools

were included in this study to increase sample heterogeneity and representation of the range of students on whom the ASCA-H is intended for use.

In addition to ensuring heterogeneity in the sample, the size of the sample must be adequate. A number of guidelines have been proposed for estimating an adequate sample size for exploratory factor analysis. These guidelines vary widely (Fabrigar et al., 1999). For example, Nunnally (1978) recommended a ratio of 10 participants for each measured variable. Gorsuch (1983) previously suggested that there should be at least 5 participants per variable and that a sample of 200 was preferred. Comrey and Lee (1992) provided a guide that sample sizes of 50 are *very poor*, 100 are *poor*, 200 are *fair*, 300 are *good*, 500 are *very good*, and 1,000 are *excellent*.

Based on the evidence given by Guadagnoli and Velicer (1988) regarding the stability and replicability of structural analyses, Clark and Watson (1995) recommended that a minimum of 300 respondents be assessed at this early stage of test development. This is consistent with Gorsuch's (1997) recommendation of 300 participants for item factor analysis and Tabachnick and Fidell's (1996) suggestion of at least 300 cases for factor analysis. The final sample size of 426 in this study met these recommendations.

Appropriateness of factor analysis. In addition to sample size and composition, considerations regarding the factorability of the correlation matrix must be met prior to conducting factor analysis. This condition can be analyzed by considering the randomness of the correlation matrix (Tabachnick & Fidell, 1996). Tabachnick and Fidell noted that a correlation matrix that is factorable includes "several sizable correlations" (p. 641) and suggested that at least one of the correlations in the matrix should exceed .30.

A more formal test of the randomness of the correlations is Bartlett's (1954) Test of Sphericity, which tests the hypothesis that the correlations in the matrix are zero. As part of this study, Bartlett's Test of Sphericity was conducted prior to the factor analysis. A significant result, $p < .05$, was necessary for the factor analysis to continue. In addition as suggested by Tabachnick and Fidell, at least one bivariate correlation greater than .30 was deemed necessary for this study.

Another indication of the factorability of the correlation matrix is the Kaiser-Meyer-Olkin Test of Sampling Adequacy (KMO; Kaiser, 1974). This formal test for examining the sampling adequacy of a correlation matrix involves computing a ratio of the sum of the squared correlations to the sum of the squared correlations plus the sum of the squared partial correlations. A ratio of .6 and above is considered good for conducting factor analysis (Tabachnick & Fidell, 1996). A criterion of at least .6 was used for the purposes of this study.

Decisions must be also made at each subsequent step of the factor analysis. Gorsuch (1974) emphasized the importance of making these decisions prior to conducting analysis of data in order to have maximum confidence in the results. Some of these steps include selecting a model of factor analysis and selecting methods of factor extraction, retention, and rotation.

Models. Two types of factor analysis are possible: common factor analysis (FA) and principal component analysis (PCA). According to Fabrigar et al. (1999), PCA is appropriate when the goal of analysis is data reduction, or reducing a large set of measured variables to a smaller set of composite variables that retain the maximum

amount of information from the original set as possible. In contrast, FA based on a common factor model is appropriate when the goal of analysis is to reach a parsimonious understanding of the associations between measured variables. Widaman (1993) similarly argued that FA allows researchers to interpret the patterns of covariation among variables stemming from latent factors or constructs. To do this, the effects of specific and error variance on the factor loadings and intercorrelations must be minimized.

PCA and FA differ in the type of variance that is analyzed in producing the factor solution (Tabachnick & Fidell, 1996). PCA analyzes the common, unique, and error variance whereas FA only analyzes the common variance. FA is based on the notion that including error and unique variance may make it more difficult to interpret the solution regarding the true underlying processes (Tabachnick & Fidell). The formula for PCA is identical to that of FA without the error term (Gorsuch, 1997).

Use of PCA and FA may result in different factor solutions. In a comparison study of PCA and FA, Snook and Gorsuch (1989) concluded that FA provided a more accurate solution in each of the conditions examined. The loadings obtained by the common factor method were not biased and had lower standard errors than those obtained by PCA. The principal component loadings were systematically and significantly inflated. However, Snook and Gorsuch found that the differences in solutions obtained through the two methods decreased as the number of variables and the population pattern loadings increased. Gorsuch (1997) noted that PCA gives inflated loadings in item analysis while FA provides the best estimate of correlations.

A common factor model was used in this study because the goal was to understand latent constructs underlying the ASCA-H. Therefore, the number and nature of common factors needed to account for the pattern of correlations among the items must be examined. The common factor model excludes unique and error variance and is thus more appropriate for the purposes of this study. Gorsuch (1997) also argued that using PCA “assumes that the variables are conceptualized as reproduced perfectly by the factors. Perfect reproduction generally means that the variables are almost perfectly reliable and correlate highly with at least one other variable” (p. 539). However, items are not likely to be highly reliable and may reflect more than one factor. Error is introduced by the distribution and categorization problems described earlier. Therefore, including the error term is appropriate for item analysis and so FA is preferred to PCA (Gorsuch, 1997). This model was also chosen because it has been shown to be consistently more accurate than PCA (Snook & Gorsuch, 1989; Widaman, 1993).

Extraction Method. The next decision involved in factor analysis is to decide what procedure to use to fit the common factor model to the data (Fabrigar et al., 1999). Two commonly used extraction procedures are principal axis (or principal factors) and maximum likelihood (Tabachnick & Fidell, 1996).

Maximum likelihood extraction estimates population values for factor loadings by calculating loadings that maximize the probability of sampling the observed correlation matrix from a population (Tabachnick & Fidell, 1996). The population estimates for factor loadings are calculated so that they have the greatest probability of yielding a sample with the observed correlation matrix. Fabrigar et al. (1999) noted an advantage of

maximum likelihood is that it permits calculation of a range of goodness-of-fit indices of the model. Significance tests and confidence intervals for factor loadings and correlations between factors are also possible in maximum likelihood extraction. However, maximum likelihood is based on the assumption of multivariate normality, which assumes that all variables and linear combinations of variables are normally distributed (Tabachnick & Fidell, 1996). Distorted results may occur when maximum likelihood is used when this assumption is seriously violated (Curran, West, & Finch, 1996).

Principal axis extraction, on the other hand, does not make assumptions about the distributions of measured variables in the population (Briggs & MacCallum, 2003; Fabrigar et al., 1999). In the principal axis method, estimates of communality are included in the positive diagonal of the correlation matrix. These estimates are derived through an iterative procedure that begins with squared multiple correlations. The goal of the procedure is to extract the maximum orthogonal variance from the data set with each factor (Tabachnick & Fidell, 1996).

Gorsuch (1983) argued against including nonnormal variables in maximum likelihood procedures. As mentioned, item responses inherently fall into a small number of categories, and their distributions are nonnormal to the extent determined by the response level (i.e., probability of answering in a specified direction; Bernstein & Teng, 1989). The dichotomous nature of the variables in this study increased the probability of nonnormal univariate distributions. Given its relative tolerance of multivariate nonnormality, along with its superior ability to recover weak common factors (Briggs & MacCallum, 2003), principal axis extraction was applied in the data analysis.

Communalities were initially estimated by squared multiple correlations and iterated to final estimates.

Number of factors. The next decision in the factor analysis is to determine how many factors to retain. Tabachnick and Fidell (1996) stated that it is important to keep enough factors for an adequate fit to the original data and at the same time not retain so many factors that parsimony is lost. Factoring should be stopped when additional factors account for trivial variance (Rummel, 1970). Some evidence favors overestimating rather than underestimating the number of factors (Comrey, 1988; Fava & Velicer, 1992; Rummel; Wood, Tataryn, & Gorsuch, 1996). Therefore, it is suggested that the highest to lowest number of factors be examined until the most interpretable solution is found (Ford, MacCallum, & Tait, 1986). Fabrigar et al. (1999) also argued for examining the rotated solutions of each specified number of factors to determine which produces the most readily understandable and theoretically sound results.

Several methods can be used to determine the most appropriate number of factors to extract. A commonly used method is the Unity-Eigenvalue Rule (Kaiser, 1960). Also known as Kaiser's criterion, it indicates retaining the number of factors with eigenvalues greater than one. This method is a default option in many statistical packages, such as SPSS.

Although it is popular and easy to implement (Velicer, Eaton, & Fava, 2000), use of the Unity-Eigenvalue Rule has been sharply criticized when compared to other available methods. Zwick and Velicer (1986) found that the Unity-Eigenvalue Rule seriously overestimated the actual number of factors in a comparison of factor extraction

procedures. Both Gorsuch (1983) and Kaiser (1960) reported that use of this method commonly results in the retention of one-third to one-fifth as many factors as there are original variables. It has also been shown that in cases with low communalities, the number of factors the Unity-Eigenvalue Rule indicates to retain is related to the number of variables rather than actual structure of the data (Zwick & Velicer, 1982). The Unity-Eigenvalue Rule has also been found to overestimate the number of factors with data having low correlations (Gorsuch, 1997). As previously noted, items have lower reliabilities than scales, leading to lower correlations in item factor analysis. Therefore, Gorsuch (1997) concluded that the Unity-Eigenvalue Rule would produce too many factors in analysis of items. Velicer et al. (2000) argued that the rationale and accuracy of the Unity-Eigenvalue Rule were poor, and its use was not recommended. Thus, it was not considered in this research.

The scree test, developed by Cattell (1966), is another procedure for determining the number of factors to retain. In the scree test, eigenvalues are plotted, and a straight line is fit through the smaller values. Those factors with eigenvalues falling above the line are retained (Zwick & Velicer, 1986).

Although the scree test is straightforward, several complications may arise. First, there may be no obvious break in the line, or there may be more than one apparent break. Also, more than one appropriate line may seem to fit the eigenvalues. The decision of where to draw the line and how to interpret it is up to the investigator, leading some researchers to question the objectivity of the scree test (Ford et al., 1986). Zwick and Velicer (1976) found the scree test to be more variable and less accurate than other

methods in cases with more complex patterns (i.e., patterns containing unique and complex items). However, Velicer et al. (2000) noted that the visual basis of the scree test might potentially alleviate some of the overextraction problems because “trivial factors will not be visually compelling” (p. 47). They recommended use of the scree test in supplement to other methods, such as parallel analysis (Horn, 1965).

Parallel analysis is a sample-based adaptation of the population-based Unity-Eigenvalue Rule (Zwick & Velicer, 1986). It was intended to address the inability of the Unity-Eigenvalue Rule to reflect sampling error (Velicer et al., 2000). In parallel analysis, a set of random data correlation matrices having the same number of variables and participants as the observed data is produced. The eigenvalues from the random data are compared with the eigenvalues of the observed data. The parallel analysis criterion for the number of factors is defined as the number of eigenvalues from sample data that are greater than the corresponding eigenvalues from the random data (Ford et al., 1986).

Parallel analysis is recommended for determining the number of factors to retain (Thompson & Daniel, 1996; Velicer et al., 2000). Velicer et al. noted that by comparing the eigenvalues of the sample data to the eigenvalues of random data rather than a fixed value of 1.0 (as in the Unity-Eigenvalue Rule), parallel analysis accounts for random error. However, they argued that the criticisms of the Unity-Eigenvalue Rule also weaken the rationale for parallel analysis. Zwick and Velicer (1986) noted that although parallel analysis is generally accurate, it tends to retain poorly defined factors and overextract by one or two factors. To reduce these tendencies, Glorfeld (1995) proposed a modification of the parallel analysis procedure. Rather than average the eigenvalues from a small

number of random correlation matrices, Glorfeld suggested developing a complete empirical distribution of eigenvalues from a large number of random correlation matrices so that the usual average can be replaced with any desired percentile point of the distribution (e.g., 95th or 99th).

Another recommended method is the Minimum Average Partial (MAP; Velicer, 1976) procedure. This procedure is based upon the reasoning that as common variance is partialled out of the correlation matrix for each successive value, m , the MAP criterion will continue to decrease. When the common variance has been removed and there is only unique variance left, the MAP criterion will begin to rise. As such, the MAP criterion separates common and unique variance and retains only factors consisting of common variance (Velicer). Zwick and Velicer (1982) found the MAP procedure to be comparable to the scree test and more accurate than the Unity-Eigenvalue Rule and Bartlett's significance test. Zwick & Velicer (1986) later showed that the MAP criterion was more accurate than the Unity-Eigenvalue Rule, scree test, and Bartlett's significance test with more complex data.

For this study, multiple procedures were used to determine the number of factors to retain. Based on the recommendations of Velicer et al. (2000), parallel analysis and MAP were used, supplemented by the scree test. The scree test was generated by SPSS 10.0; however, parallel analysis and MAP are not included in common statistical packages (Fabrigar et al., 1999). The modified parallel analysis procedure (Glorfeld, 1995) and MAP analysis were conducted using a program developed by O'Connor (2001) for SPSS.

Factor rotation. The next decision in factor analysis involves whether or not to rotate factors and if so, what type of rotation to perform. The guiding principle in factor rotation is simple structure (Gorsuch, 1983, 1997). Rotation of factors helps to clarify the factor structure and makes the solution easier to interpret without changing the underlying mathematical properties (Tabachnick & Fidell, 1996). The number of variables that load on a factor is minimized in factor rotation as are the number of factors that load the same variables, helping to simplify the factor structure (Gorsuch, 1983, 1997).

Two types of rotations are possible: orthogonal and oblique. The fundamental difference between them is that it is possible for the factors to be correlated in oblique rotation whereas factors must be uncorrelated in orthogonal rotation (Tabachnick & Fidell, 1996). Some common types of orthogonal rotation include varimax, quartimax, and equamax.

The factors resulting after an orthogonal rotation are often easier to interpret, yet this is not the most appropriate approach if it is theorized that the underlying factors are interdependent. If there is reason to suspect that factors are correlated, an oblique rotation should be performed (Tabachnick & Fidell, 1996). Several researchers (Horn, 1965; Meredith, 1965) have argued that even if attributes are orthogonal in the population, correlations between them can be expected in samples drawn from the population, especially if the samples are not perfectly representative of the population.

Although oblique rotations allow factors to correlate, they do not require factors to be correlated. As discussed by Harman (1976), a successful oblique rotation will

produce correlations of approximately zero and a very similar solution to that of a successful orthogonal rotation if the best simple structure solution involves uncorrelated factors. Similarly, Gorsuch (1997) noted that nonrestricted solutions, such as those produced by oblique rotations, would give uncorrelated factors when that provides a reasonable solution.

The most common oblique rotation is direct oblimin (Tabachnick & Fidell, 1996). However, promax has been described as an “efficient procedure giving good solutions” (Gorsuch, 1983, p. 205) and of “demonstrated quality” (Gorsuch, 1983, p. 190). According to Comrey and Lee (1992), the promax procedure “is designed to give an oblique solution that represents a modification of an orthogonal solution toward an even better oblique simple structure” (p. 188). The promax procedure uses an orthogonal rotation, typically varimax or quartimax, for the initial solution. The factor loadings from the orthogonal rotation are then raised to a higher power with the sign of each loading replaced. Although the loadings decrease in size, the high loadings remain relatively high whereas the moderate and low loadings become closer to zero (Gorsuch, 1983).

Because personality constructs or dimensions of such constructs are likely to be correlated, oblique rotations afford a more sensible and accurate representation of how constructs are likely to be related (Fabrigar et al., 1999). Oblique (promax) rotations and, for comparison, orthogonal (varimax) rotations were performed in this analysis. According to Gorsuch (1983), a comparison of orthogonal and oblique solutions encourages selection of the simpler uncorrelated factor model, if relevant. If correlations among oblique factors are negligible, then the varimax rotation solution can be accepted

as a reasonable solution. However, if the correlations are sizable, then the oblique solution is more appropriate (Gorsuch, 1983).

Factor interpretation. Interpreting a factor involves understanding the underlying construct or process that unifies the group of variables that load on it. Each factor is based upon a set of highly correlated items (Gorsuch, 1997). The loading matrix from orthogonal rotation splits into two matrices for oblique rotation: a structure matrix that includes correlations between factors and variables and a pattern matrix containing values that represent the unique contributions of each factor to the variance in the variables (Tabachnick & Fidell, 1996). Tabachnick and Fidell argued for the use of the pattern matrix in factor interpretation. They noted that correlations between variables and factors are inflated by an overlap between factors, a problem that becomes more severe as the correlations among factors increase. Thus, using the structure matrix makes it more difficult to determine which variables are related to a factor. In contrast, the pattern matrix presents “unique relationships uncontaminated by overlap among factors” (Tabachnick & Fidell, p. 637). Shared variance is omitted, but the set of variables that composes a factor is usually easier to see. For this study, the pattern matrix was interpreted. The term *loading* will hereafter refer to pattern coefficients.

Generally, the higher the correlation between the factor and the variable, the more that variable is a pure measure of that factor (Tabachnick & Fidell, 1996). Researchers have established rules to guide interpretation. Comrey and Lee (1992) stated that factor loadings $> .71$ are *excellent*, $> .63$ are *very good*, $> .55$ are *good*, $> .45$ are *fair*, and $< .32$ are *poor*. Gorsuch (1997) defined a salient loading as one that is greater than .40 and is

the highest loading for that variable. Gorsuch (1974, 1983, 1997) recommended interpreting factors having three or more salient loadings. Comrey and Lee argued that three loadings would represent the very minimum, but five or more variables are considered preferable for defining a factor (Comrey & Lee; Fabrigar et al., 1999).

In studies with the original ASCA, McDermott (1994) specified that five or more items with loadings greater than or equal to .30 were used in defining a factor. For purposes of consistency, a criterion of a loading of .30 was set for defining and interpreting factors in this study. A minimum of five salient loadings was also used for defining a factor.

The internal consistency reliability (coefficient alpha) for each factor was also considered. DeVellis (2003) suggested that for research scales alpha values below .60 were *unacceptable*; between .60 and .65 were *undesirable*; between .65 and .70 were *minimally acceptable*; between .70 and .80 were *respectable*; and between .80 and .90 were *very good* (p. 96).

Results

Conditions

As anticipated, 27 of the 29 positive or adaptive items were endorsed by a majority of the respondents (see Table 3). Item 49 (Accepts correction without a fuss) and item 96 (Willing or eager to do household chores) were endorsed by 22% and 43% of the responding parents, respectively. As recommended by McDermott (1994), all 29 positive items were omitted from subsequent analyses.

Given the severity of the non-situationally specific items (i.e., numbers 181 through 203), it was anticipated that they would be relatively infrequently endorsed by parents. That expectation was confirmed: the endorsement rate of those 23 items ranged from 0.2% to 16.0% and included 12 out of the 20 most sparsely endorsed items on the entire ASCA-H. These 23 items were not intended to be part of the ASCA-H situational assessment of childhood adjustment and, consequently, were omitted from subsequent analyses.

It was also anticipated that more problem behaviors would be reported for students from the special education school than for students in regular education schools or those referred to the IST. One-way ANOVAs were calculated to assess differences in group mean scores. As expected, parents reported significantly more problem behaviors for students in the special education school group than for students in the other groups ($F(2, 423) = 71.91, p < .05$).

Descriptive analyses were first conducted to examine the univariate distributions of the 151 situationally specific or maladaptive items. Item 167 (Supplies dangerous

Table 3

Description and Endorsement Prevalence of ASCA-H Items

Item	Description ^a	Prevalence ^b	Rank ^c
1*	Greets parent as most other children do	78.2	8H
2	Clings to parent or shows tears	6.1	
3	Waits for parent to greet him/her first	13.1	
4	Seems too unconcerned about parent to greet	6.6	
5	Responds with an angry look or turns away when greeted	4.2	
6	Greets parent loudly	18.8	
7	Answers question before s/he has had time to think	35.9	
8	Not shy but rarely offers an answer	8.9	
9*	Answers if s/he can	62.0	20H
10	Freezes up and has trouble answering	12.4	
11	Lacks the energy to ask for help	4.0	
12*	Asks for help when s/he needs it	82.6	5H
13	Too timid to ask for help	2.8	
14	Never seeks help even if needed	7.0	
15	Seeks help when not needed	20.4	
16	Much too talkative with parent	21.6	
17	Tends to be in untalkative moods	19.5	
18	Speaks in a rude or angry tone	22.1	
19	Distant, seldom says anything	4.7	
20	Uses dirty words or offensive language	6.1	
21	So timid it's difficult to get him/her to speak to you at all	2.1	
22*	Talks freely	76.8	10H
23	Wants your interest but holds back	11.7	
24	Uses various ways to get your attention	56.6	
25*	Likes to be praised	70.2	15H

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
26	Quite unconcerned whether s/he gets parents' attention	8.5	
27	Sometimes seems to seek disapproval	13.8	
28	Tells tall tales about him/herself or his/her friends	16.4	
29*	Does not generally tell untruths	55.2	
30	Sometimes lies to avoid blame or punishment	51.6	
31	Doesn't hesitate to lie	8.5	
32*	Friendly, smiles readily with parent	81.0	6H
33	Distant, makes not effort to relate to parent	4.0	
34	Rarely smiles with parent	1.6	
35	Has a dejected look with parent	3.8	
36	Overly dependent with parent	8.5	
37	Sometimes in an unfriendly mood with parent	30.0	
38	Shy but not unfriendly with parent	8.9	
39	Misbehaves when parent is attending to other things	39.4	
40*	Well behaved at home	60.8	
41	Never any trouble because she/he's so timid	1.2	10L
42	Disobedient, difficult to control at home	13.1	
43	Doesn't seem to have enough energy to misbehave	1.6	
44	Does things in front of parent s/he knows are wrong	23.9	
45	Improves for the moment when corrected but does not last	25.8	
46	Takes correction badly	35.4	
47	Argues and talks back but will eventually do what is asked	52.6	
48	His/her feelings are easily hurt when corrected	43.2	
49*	Accepts correction without a fuss	22.1	
50	Answers back aggressively, threatens	15.5	
51*	Greets other adults as most children do	61.5	

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
52	Waits for adults to greet him/her first	33.8	
53	Seems too unconcerned about people to greet	7.0	
54	Responds with an angry look when greets other adults	1.9	
55	Welcomes other adults loudly	14.6	
56	Seems fearful of other adults	3.1	
57	Too shy to greet other adults	11.5	
58	Uses bad language that offends other adults	3.5	
59	Too shy to interact with other adults	7.5	
60	Is rude with other adults	4.5	
61*	Well behaved with other adults	79.1	7H
62	Uses various ways to get adults' attention	24.6	
63	Seems unconcerned whether s/he gets adults' attention	11.3	
64	Sometimes wanders off by self	19.7	
65	Prefers adults to own age peers	11.0	
66	Associates with troublesome children	10.3	
67*	Has one or more companions with whom s/he gets along	82.9	4H
68	Prefers younger children	12.2	
69	Is a loner, prefers not to associate with other children	6.1	
70	Clowns around, plays silly tricks	38.5	
71*	A good mixer with others same age	69.5	18H
72	Cannot keep a friend for long	10.6	
73	Unkind to weaker children	4.5	
74	Quarrels, provokes others	9.6	
75	Tries to dominate with others his age	17.1	
76	Fights physically with others	6.6	
77	Takes others' things without permission	16.0	

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
78	Snatches objects away from other children	8.5	78
79	Has deliberately destroyed others' belongings	5.4	
80	Too uninterested in others to notice their belongings	1.4	16L
81*	Respects others' property	83.8	3H
82	Does not stand up for self	17.4	
83	Attacks others physically if provoked	16.2	
84*	Stands up for self when s/he has to	76.3	11H
85	Allows him/herself to be bullied	18.8	
86*	Takes care of his/her things	59.2	
87	Often loses his/her belongings	42.3	
88	Overly fussy about his/her things	12.7	
89	Destroys belonging purposely	11.0	
90	Charges in without taking time to think	25.2	
91	Won't attempt a new task if she/he senses difficulty	17.4	
92	Cannot work up the energy to face anything new	1.2	10L
93*	Approaches a new task with caution but gives it a try	52.1	
94	Becomes visibly upset if does not perform well	37.1	
95	Helps with household jobs unless in a bad mood	41.5	
96*	Willing or eager (with household jobs)	42.7	
97	Asks to be given jobs but often doesn't finish them	28.2	
98	Appears to withdrawn to offer help	5.9	
99	Refuses help with chores	12.9	
100*	Waits patiently in line	52.8	
101	Becomes restless and fidgety in line	49.3	
102	Argues and complains about waiting	20.2	
103	Tries to push ahead in line	6.6	

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
104	Will let others push ahead of him/her	5.9	
105	Gets into fights with others in line	2.8	
106	Throws tantrums at bedtime	11.5	
107	Stays in bed to avoid responsibilities	5.4	
108	Has trouble sleeping unless s/he is near parent	17.8	
109*	Goes to bed without too much fuss	75.4	14H
110	Sleeps more than most children but is still tired	4.0	
111	Refuses to eat	11.0	
112*	Appropriately eats the food served	77.7	9H
113	Starts arguments or fights at meal times	4.9	
114	Constantly restless at meal times	16.9	
115	Loses temper if doesn't get what s/he wants at meals	12.4	
116*	Personal cleanliness is similar to other children	86.4	2H
117	Seems to lack the energy to take care of him/herself	8.7	
118	Refuses to care for his/her hygiene	7.0	
119	Needs encouragement to join in games	18.8	
120	Too disinterested to play games	3.3	
121	Disrupts games by fooling around	9.6	
122	Won't get involved with games	3.8	
123	Has trouble waiting for his/her turn during games	15.3	
124	Inclined to cheat	8.5	
125	Poor loser	23.2	
126*	Joins in games eagerly	62.9	17H
127	Doesn't seem to have fun in games he/she used to enjoy	2.6	
128*	Enjoys club or team activities	76.1	12H
129	Refuses to take part in any formal organized activity	9.4	

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
130	Lacks the energy to take part in organized activities	5.6	
131	Has been expelled from organized activities for behavior	3.3	
132	Rather loud but not disruptive in informal play	27.5	
133	Is too timid to join in informal play	5.9	
134	Disturbs others' fun in informal play	7.5	
135	Wants to dominate and his/her own way in informal play	21.6	
136*	Engages in sensible play activities	59.9	
137	Starts fights and rough play	4.5	
138	Tends to go off by self and play alone	15.5	
139	Overly rough with smaller or weaker children	5.9	
140	Seems afraid to try work with his/her hands	4.0	
141*	Sticks to his/her activity	66.9	19H
142	Fools around when working with hands	13.6	
143	Has ruined his/her work on purpose	5.9	
144	Very slow and never seems to finish things	10.8	
145	Lacks physical energy when working with hands	1.2	10L
146	Very concerned about mistakes when working	35.7	
147*	Appropriately gets ready for school	76.1	12H
148	Appears fearful of going to school	3.5	
149	Moody and uncooperative when going to school	16.2	
150	Too disinterested to get ready for school	3.3	
151	Constantly distracted so s/he is not ready on time	22.3	
152	Skips school with friends	0.7	6L
153	Stays home rather than attend school regularly	1.9	
154*	Attendance has been good	96.9	1H
155	Listless, unconcerned with homework	8.0	

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
156*	Generally does homework on his/her own	70.0	
157	Asks for help with homework even it is not needed	24.9	
158	Lies about assignments	13.4	
159	Destroys or defaces his/her school materials	6.6	
160	Often loses or forgets books or materials	21.1	160
161	Makes sexually offensive gestures or remarks	4.9	
162	Has intentionally hurt animals	3.3	
163	Participates in fights with a gang	0.7	6L
164	Threatens to hurt others if s/he doesn't get what s/he wants	10.6	
165	Steals from home	4.2	
166	Uses dangerous drugs	0.2	1L
167	Supplies dangerous drugs to others	0	
168	Regularly drinks alcoholic beverages	0	
169	Has carried dangerous weapons	1.4	16L
170	Has stolen from a store, building, or car	5.2	
171	Has deliberately damaged property	5.6	
172	Makes sudden inappropriate noises	10.6	
173	Hurts self by scratching, picking, or biting	6.8	
174	Often cries for no apparent reason	6.6	
175	Actions are interrupted by involuntary movements	2.8	
176	Panics in response to unexpected noises or events	7.7	
177	Has made unprovoked attacks on parent	3.1	
178	Has made sudden unprovoked attacks on other children	4.5	
179	Without warning may throw an object across the room	9.4	
180	Rushes about shouting madly	10.3	
181†	Has been suspended or expelled	11.7	

table continues

Table 3 (continued)

Item	Description ^a	Prevalence ^b	Rank ^c
182†	Speaks normally at home but refuses to speak at school	2.8	
183†	Regularly smokes or chews tobacco	0.5	4L
184†	Says that other people can read mind or control thoughts	0.2	1L
185†	Accuses others of spying on him/her	3.1	
186†	Responds to voices that non one else can hear	1.2	10L
187†	Controls weight by starving, throwing up, etc.	0.5	4L
188†	Persistently eats nonedible materials	1.2	10L
189†	Claims s/he is famous or has special powers	0.2	1L
190†	Talks about hurting him/herself	4.5	
191†	Gets very agitated if doesn't check things over and over	0.7	6L
192†	Expresses extreme fear of common animals	2.8	
193†	Expresses extreme fear of medical situations	6.3	
194†	Expresses extreme fear of injury or death	3.3	
195†	Expresses extreme fear of certain places	1.4	16L
196†	Gets very upset if things are not arranged in the right order	4.7	
197†	Often says that s/he is dumb	16.0	
198†	Runs away from home	1.6	20L
199†	Sets fire without permission	0.7	6L
200†	Constantly talks about germs or disease	1.4	16L
201†	Has noticeable tics he/she can't control	1.2	10L
202†	Wets the bed	7.0	

Note. From *The Adjustment Scales for Children and Adolescents – Home*, by M. W.

Watkins and P. A. McDermott, 2002, Unpublished.

*Positive or adaptive items

†Non-situationally specific items not included in the ASCA-H scale

^aItem descriptions are abbreviated with situational context integrated into wording.

^bEach value corresponds to the item endorsement rate multiplied by 100.

^cValues represent items with the 20 highest endorsement prevalence (designated by the corresponding numerical ranking and “H”) and the 20 lowest endorsement prevalence (designed with a corresponding numerical ranking and “L”).

drugs to others) and item 168 (Regularly drinks alcoholic beverages) were unendorsed in the sample and were excluded from further analysis. As shown in Table 3, 58 of the 202 items had endorsement rates below 5%. One item had an endorsement rate above 95%. It is sometimes recommended to eliminate such items during scale development (Schwartz, 1978). However, as argued by McDermott (1994) in the development of the ASCA, the ASCA-H as a measure of behavioral and social-emotional adjustment problems must have enough variability to support reliability and discrimination. Therefore, even low prevalence items were included in exploratory factor analysis. In all, 149 problem behavior items were included in subsequent analysis.

Due to the dichotomous nature of the variables, nonnormality was anticipated and was indeed found in the item distributions. Fabrigar et al. (1999) noted that skewness values exceeding 2 and kurtosis values exceeding 7 are indicative of severe nonnormality. According to these standards, descriptive analyses of items indicated that 100 of the remaining 149 items (67%) were skewed. Seventy-one items (48%) had kurtosis values exceeding 7. Such significant deviations from normality are often encountered on variables measuring rare, pathological behavior (Floyd & Widaman, 1995) and with dichotomous data (Gorsuch, 1997).

Data transformation is often recommended for failures of normality. However, it is not always appropriate because an analysis is interpreted from the variables in it, and transformed variables are sometimes more difficult to interpret (Tabachnick & Fidell, 1996). Also, because the goal of this study was to examine the underlying factor structure of the ASCA-H, items were not omitted from subsequent analyses due to nonnormality. However, a maximum likelihood extraction procedure was not utilized in the factor analysis due to the nonnormal distributions of many of the items and the sensitivity of ML procedures to such deviation from normality.

Another consideration prior to factor analysis involved the factorability of the correlation matrix. First, as suggested by Tabachnick and Fidell (1996), the correlation matrix was examined for correlations greater than .30. Several bivariate correlations exceeded .30. Also, the significant results of Bartlett's Test of Sphericity (Bartlett, 1954) indicated that the correlations in the matrix were not random. Next, the final sample size of 426 met the aforementioned guidelines for item factor analysis. Lastly, the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) statistic was .763, above the minimum standard suggested by Tabachnick and Fidell (1996) for conducting factor analysis. Taken together, results of each criterion suggested that the correlation matrix was factorable.

Number of Factors

Several procedures were considered in determining the number of factors to retain. In the scree test, a point is sought where there is a break in the eigenvalues (Comrey & Lee, 1992). Visual inspection of the scree plot (Figure 1) indicated a large

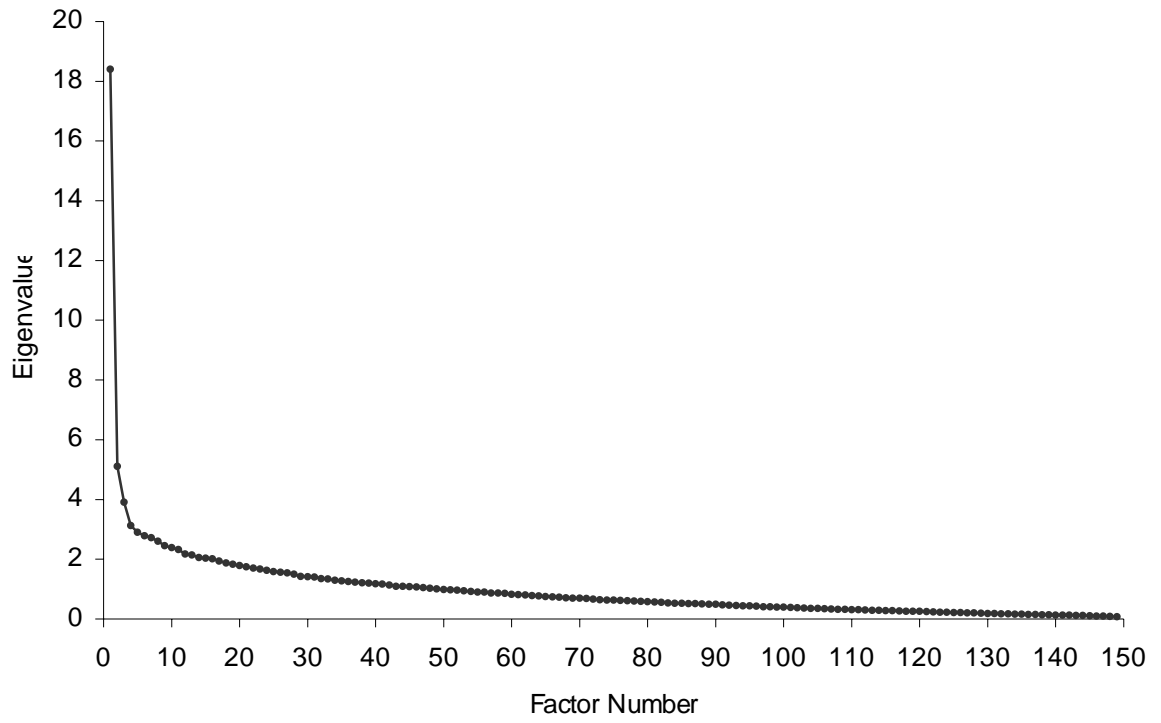


Figure 1. Scree plot of number of factors to retain.

break at the fourth factor and smaller breaks at the eighth and eleventh factors. Results of a regression-based (AUTOSCREEN; Barrett & Kline, 1982) confirmed the visual scree test results.

A modification of the parallel analysis procedure described by Glorfeld (1995) was used with the desired percentile of the distribution and random data eigenvalues set at the 99th percentile. The modified parallel analysis procedure generated a criterion of 19 factors. In comparison, results of the MAP indicated 10 factors should be retained. To examine which factor solution best explained the resulting correlations among the observed variables, iterative extractions of 4 to 19 factors were performed.

Using a common factor model, principal axis factor analyses with promax rotation were conducted. Varimax rotation was also performed for comparison of the stability of the factor solutions. However, several moderate correlations between factors suggested that an oblique solution better represented the data. Results of the promax rotation are, therefore, reported.

Factor Solutions

Each factor solution was evaluated according to a prior criteria for statistical suitability and substantive meaningfulness. As indicated in Table 4, each extraction resulted in many items that did not load on any factor, as well as many complex items. In contrast to “pure-factor-data variables” (Comrey & Lee, 1992, p. 206), which have major loadings on only one factor, complex data variables load highly on more than one factor. Such factorial complexity makes it difficult to understand what constructs are tapped by the items and how to interpret each factor (Streiner, 1994). According to Comrey and

Lee, when complex variables are included in an analysis, they should not be used to define a factor. Therefore, complex items were excluded.

As shown in Table 4, the solutions resulting from extraction of 9 to 19 factors were unable to produce more than six factors that met a prior criterion for number of salient items. Additionally, these solutions were marked by many items that were complex (11-17 items) or failed to load saliently on any factor (36-56 items). Consequently, these solutions were rejected.

When the eight-factor solution was considered, the first factor contained 34 salient items and seemed to tap an aggressive, disruptive dimension (e.g., gets into fights, snatches objects from others, threatens to hurt others, and difficult to control). Its coefficient alpha was .92. Twenty items that tapped impulsive and attention-seeking behaviors (e.g., answers before has time to think, uses various ways to get attention, asks for help even when not needed) comprised the second factor. Its internal consistency reliability was .81. Eleven items involving timid and shy behaviors (e.g., too timid to join organized play, too shy to greet adults, needs encouragement to join in games) saliently loaded the third factor. Its coefficient alpha was .70. The fourth factor contained eight salient items that tapped a distant, avoidant dimension (e.g., distant/makes no effort to relate, seems too unconcerned about parent to greet, stays in bed to avoid responsibilities). The alpha coefficient for this factor was .68, which was minimally adequate according to a prior criterion (DeVellis, 2003). The fifth factor contained five items that dealt with disparate content: two items focused on delinquent behaviors (e.g., stealing, fighting with a gang) and three dealt with timidity and excessive sleep. Thus,

Table 4

Comparison of Factor Solutions

Number of Factors Extracted	Number of Factors with 5 Salient Loadings	Number of Complex Items	Number of Items Failing to Load	% Non-redundant Residuals > .05
19	4	17	42	13
18	5	15	38	13
17	6	16	39	14
16	6	11	36	16
15	5	11	40	17
14	4	13	37	18
13	5	12	46	19
12	5	12	44	21
11	6	11	49	22
10	5	11	52	23
9	5	12	56	24
8	6	9	51	26
7	5	9	51	27
6	6	6	55	29
5	4	6	61	30
4	4	3	59	31
3	3	2	63	34

Note. Analyses were conducted using principal axis extraction and promax rotation.

this factor appeared to contain incongruous content. This observation was supported by the .55 alpha coefficient. Thus, the fifth factor was rejected as inadequate. Five salient items also formed the sixth factor. This factor seemed to tap school avoidance, disinterest, and depressed behaviors (e.g., dependence and no longer experiencing fun in previously preferred activities). The coefficient alpha for the sixth factor was only .40, falling far below the a priori criterion. Therefore, this factor was also rejected. The seventh and eighth factors failed to meet the criterion of having at least five salient items, with 4 and 2 salient items respectively. As these were deemed inadequate to form a stable factor, they were also rejected. Therefore, the eight-factor solution produced only four adequate factors.

Given the inadequacy of the eight-factor solution, a seven-factor solution was then considered. Very similar to the first factor from the eight-factor extraction, the first factor in this solution contained 34 salient items reflecting aggressive, disruptive behaviors. Its coefficient alpha was .92. The second factor was made up of 21 salient items dealing with impulsive and attention-seeking behaviors. Its internal consistency reliability was .82. Fourteen items reflecting timid and shy behaviors saliently loaded the third factor. Its coefficient alpha was .70. Eleven items tapping a detached, avoidant dimension loaded on the fourth factor. The alpha coefficient for this factor was .68, which was minimally adequate (DeVellis, 2003). Similar to results of the eight-factor extraction, the fifth factor in the seven-factor solution seemed to contain inconsistent content, with two of the five salient items involving delinquent behaviors and three dealing with timidity and excessive sleep. The low alpha coefficient of .55 supported the

disparate content. On this basis, the fifth factor was rejected as inadequate. The sixth and seventh factors contained only three items and one salient item, respectively. Therefore, they did not meet statistical criteria for meaningful factors and were subsequently rejected. The seven-factor solution produced only four adequate factors.

A six-factor extraction was then performed. Comparable to results of the other extractions, the first factor in the six-factor solution contained 33 salient items that reflected aggressive, disruptive behaviors. Its coefficient alpha was .92. Twenty-two items comprised the second factor and reflected impulsive and attention-seeking behaviors. Its coefficient alpha was .83. The third factor was made up of 14 salient items that tapped timidity and shyness. With an alpha value of .69, the internal consistency of the third factor was minimally acceptable (DeVellis, 2003). The fourth factor contained eight items. These items dealt with distant, unconcerned, and avoidant behaviors. Its coefficient alpha was .65, which was barely adequate to meet a prior criterion. The fifth factor included five salient items related to angry mood and negative responses (e.g., speaks in rude or angry tone, takes correction badly, unfriendly mood). Although this factor's internal consistency reliability of .67 was minimally acceptable and it satisfied the criterion of five items with salient loadings, it was deemed inadequate due to inconsistency across solutions. The dimension was not stable across extractions, as it was unique to the six-factor extraction. Furthermore, although the items loading on this factor had shared meaning, no such factor was identified in previous work with the original ASCA (McDermott, 1993, 1994). Given these considerations, this factor was rejected as inadequate. Five items also loaded saliently on the sixth factor; however, the sixth factor

was not substantively sound. As in previous extractions, this factor was a mix of items involving minor delinquent activities, timidity, excessive sleep, and involuntary movements. The disparate content was reflected in the low coefficient alpha of .48. Examination of item distributions indicated that each item was highly negatively skewed with very similar means. This suggested that the sixth factor was possibly a technical or difficulty factor, reflecting distribution properties rather than the latent construct being measured (Gorsuch, 1997). Therefore, the sixth factor was also rejected. Overall, the six-factor extraction resulted in four meaningful factors.

In consideration of the inadequacy of the six-factor solution, a five-factor model was next evaluated. A first factor highly consistent with previous extractions resulted that contained 33 items and tapped an aggressive, disruptive component. Its coefficient alpha was .92. The second factor consisted of 26 salient items representing attention-seeking and impulsive behaviors. Its internal consistency reliability was .84. The third factor was comprised of 12 salient items that dealt with timid and shy behaviors. The coefficient alpha was .67. The fourth factor contained eight items that tapped a distant, unconcerned, and avoidant dimension. The coefficient alpha was .65. The internal consistency reliabilities for the third and fourth factors were minimally acceptable (DeVellis, 2003). Only three items loaded saliently on the fifth factor, which was inadequate to form a stable factor. It was therefore rejected. Thus, the five-factor extraction produced only four adequate factors.

Extraction of four factors also produced four factors with at least five items with salient loadings. Three items were factorially complex, and 59 items failed to load on any

factor. Together these factors accounted for only 18.42% of the variance and left 31% of the residual matrix $\geq |.05|$. Several sizable residual correlations were found, which could suggest the presence of an additional factor (Cudeck, 2000; Gorsuch, 2003). However, the fifth factor in the five-factor extraction was not found to be a stable or meaningful factor.

The first factor in the four-factor solution contained 37 items with salient loadings and tapped an aggressive, disruptive dimension evident across extractions. The internal consistency reliability of the first factor was .92. Twenty-six items that dealt with attention-seeking and impulsive behaviors comprised the second factor. Its alpha coefficient was .85. Thirteen items saliently loaded on the third factor. This factor tapped a distant, avoidant component and had an alpha value of .70. The fourth factor included 11 salient items that reflected shy and timid behaviors. Its internal consistency reliability was minimally acceptable with an alpha level of .65.

The four-factor model met criteria of having dimensions with a minimal number of items with salient loadings and reasonable internal consistency. Additionally, the factors were substantively sound and were consistent and stable across extractions. Therefore, the four-factor model appeared to be a potentially viable factor solution.

Three factors were also extracted for statistical and substantive comparison. This extraction resulted in three factors with five or more items with salient loadings, two complex items, and 63 items failing to load on any factor. The first factor included 41 salient items and had a coefficient alpha of .93. In comparison to results of the other extractions, the first factor in the three-factor model reflected more generalized

externalizing or conduct problems, including aggression, oppositionality, and delinquency. Items related to rudeness, lying, and stealing loaded on the first factor in this solution. Consistent with previous models, the second factor, which contained 25 items, dealt with attention-seeking and impulsive behaviors. The internal consistency reliability of the second factor was .84. Eighteen salient items comprised the third factor. Its coefficient alpha was .72. The third and fourth factors from the four-factor solution collapsed into a combined factor in this solution that included both distant, avoidant behaviors as well as timid, shy behaviors. In doing so, this factor seemed to contain some contradictory material. For example, it was composed both of items reflecting a lack of concern for attention as well as items reflecting a desire for attention but reluctance or hesitance (e.g., wants parent's interest but holds back). This combined factor was also inconsistent with previous work with the original ASCA. McDermott (1993, 1994) identified two unique, separate dimensions, with the Diffident scale characterizing timid and fearful behavior and the Avoidant scale distinguishing withdrawn, aloof, and uncommunicative behavior. Two similar dimensions were evident across all previous extractions in this study as well.

The changes in the factor structure resulting from extraction of three factors suggested underfactoring. According to Wood, Tataryn, and Gorsuch (1996), extracting too few factors "can muddy the genuine factors" (p. 360), such that variables that should load on unextracted factors may incorrectly show loadings on extracted factors. Gorsuch (1997) argued that underextraction "can radically change one or more factors but extracting an additional factor when 'in the right range' leaves the earlier ones

unchanged” (p. 42). Genuine factors emerge almost as clearly when overextraction occurs as when the correct number of factors is extracted.

Extraction of four to 19 factors consistently produced four distinct, meaningful factors. The four factors showed stability and consistency in the pattern of items loadings across the number of factors extracted and the rotation method (i.e., promax, direct oblimin, and varimax). Similarity of the factors across extractions and rotations suggested that those factors were the most robust. Therefore, a three-factor solution was rejected in favor of the four-factor solution.

The four-factor model and factor loadings are presented in Table 5. Only pure-factor items are included (Comrey & Lee, 1992). Consequently, only 87 of the 149 problem behaviors (58%) are represented in the factor solution. Items are ordered and grouped according to the size of the loading to facilitate interpretation. Only salient loadings are included.

Factor I (Aggressive-Oppositional) related to confrontational, threatening, destructive, and disruptive behaviors and ill temper. This factor encompassed aspects of three scales of the original ASCA, Solitary Aggressive (Provocative), Solitary Aggressive (Impulsive), and Oppositional Defiant. Factor II (Impulsive-Attention-Seeking) consisted of items dealing with inattention, impulsivity, and a desire for attention. It was very similar to the Attention-Deficit Hyperactive scale of the ASCA. Factor III (Detached) included distant, aloof, and avoidant behaviors as well as lying and stealing. This factor reflected the ASCA’s Avoidant scale but included elements of the ASCA’s Delinquent scale as well. Factor IV (Diffident) related to timid and fearful behaviors and was

Table 5

Four-Factor Structure of the ASCA-H

Item	Item Description	Pattern Loading	Item-Total Correlation ^a
Factor I – Aggressive-Oppositional			
105	Gets into fights with others in line	.82	.54
178	Has made sudden unprovoked attacks on other children	.66	.50
164	Threatens to hurt others if doesn't get what s/he wants	.64	.59
78	Snatches objects away from other children	.63	.56
60	Is rude	.63	.55
42	Disobedient, difficult to control	.61	.74
83	Attacks others physically if provoked	.60	.61
58	Uses bad language that offends other adults	.60	.51
79	Has deliberately destroyed others' belongings	.59	.59
103	Tries to push ahead in line	.57	.50
76	Fights physically with others	.55	.49
50	Answers aggressively, makes threats	.54	.64
74	Quarrels, provokes others	.54	.61
20	Uses dirty words or offensive language	.53	.58
137	Starts fights and rough play	.51	.51
171	Has deliberately damaged property	.47	.46
89	Destroys belonging purposely	.46	.56
113	Starts arguments or fights	.42	.28
180	Rushes about shouting madly	.41	.43
179	Without warning or apparent reason, may throw object	.41	.49
72	Cannot keep a friend for long	.40	.42
161	Makes sexually offensive gestures or remarks	.40	.32
134	Disturbs others' fun	.40	.49

table continues

Table 5 (continued)

Item	Item Description	Pattern Loading	Item-Total Correlation ^a
88	Overly fussy about his/her things	.40	.29
143	Has ruined his/her work on purpose	.39	.53
123	Has trouble waiting for his/her turn	.36	.47
115	Loses temper if doesn't get what s/he wants	.36	.42
77	Takes others' things without permission	.35	.50
44	Does things in front of parent s/he knows are wrong	.35	.57
149	Moody and uncooperative	.35	.51
139	Overly rough with smaller or weaker children	.34	.43
135	Wants to dominate and his/her own way	.32	.47
73	Unkind to weaker children	.32	.43
174	Often cries for no apparent reason	.31	.28
106	Throws tantrums at bedtime	.31	.39
159	Destroys his/her school materials or defaces them	.31	.51
131	Has been expelled from organized activities for behavior	.31	.27
Factor II – Impulsive-Attention-Seeking			
7	Answers before s/he has had time to think	.55	.49
62	Uses various ways to get other adults' attention	.51	.51
24	Uses various ways to get parent's attention	.51	.42
157	Asks for help with homework even it is not needed	.50	.38
45	Improves when corrected but it does not last long	.47	.42
87	Often loses his/her belongings	.45	.39
15	Seeks help when not needed	.44	.43
39	Misbehaves when parent is attending to other things	.43	.46
16	Much too talkative	.42	.42
90	Charges in without taking time to think	.42	.34

table continues

Table 5 (continued)

Item	Item Description	Pattern Loading	Item-Total Correlation ^a
55	Welcomes other adults loudly	.41	.43
47	Argues and talks back but will eventually do it	.40	.36
70	Clowns around, plays silly tricks	.40	.37
142	Fools around when working with hands	.39	.45
101	Becomes restless and fidgety	.38	.40
6	Greets parent loudly	.37	.33
97	Asks to be given jobs but often doesn't finish them	.35	.24
94	Helps with household jobs unless in a bad mood	.35	.35
121	Disrupts by fooling around	.35	.46
114	Constantly restless, gets out of seat, plays with utensils	.35	.33
75	Tries to dominate with same age peers	.34	.38
46	Takes correction badly	.33	.31
36	Overly dependent with parent	.33	.34
102	Argues and complains about waiting	.32	.39
27	Sometimes seems to seek disapproval	.32	.42
144	Very slow and never seem to finish working with hands	.31	.31
	Factor III – Detached		
33	Distant, makes no effort to relate to parent	.58	.56
4	Seems too unconcerned about parent to greet	.43	.44
158	Lies about assignments	.42	.34
14	Never seeks help even if needed	.39	.34
107	Stays in bed to avoid responsibilities	.35	.40
66	Associates with troublesome children	.34	.30
26	Quite unconcerned whether gets parent's attention or not	.33	.30
31	Doesn't hesitate to lie	.33	.40

table continues

Table 5 (continued)

Item	Item Description	Pattern Loading	Item-Total Correlation ^a
19	Distant, seldom says anything	.33	.27
63	Seems unconcerned whether gets adults' attention or not	.32	.20
165	Steals from home	.32	.29
8	Not shy but rarely offers an answer	.31	.20
99	Refuses help with chores	.30	.32
Factor IV – Diffident			
133	Is too timid to join in informal play	.53	.46
57	Too shy to greet adults	.52	.40
119	Needs encouragement to join in games	.40	.31
59	Too shy to interact with other adults	.38	.33
10	Freezes up and has trouble answering questions	.38	.28
38	Shy but not unfriendly with parent	.37	.33
69	Is a loner, prefers not to associate with other children	.35	.35
104	Will let others push ahead of him/her	.35	.33
138	Tends to go off by self and play alone	.33	.27
23	Wants parent's interest but holds back	.32	.21
11	Lacks the energy to ask for help	.31	.19

Note. Analyses were conducted using principal axis extraction and promax rotation.

^aEach value is a corrected item-total correlation, representing the correlation between the respective item with all factor items excluding itself.

congruent with the Diffident scale of the ASCA. Factor descriptions and properties are summarized in Table 6.

Item-Total Correlations

Corrected item-total correlations, which correlate the item being evaluated with all the factor items excluding itself, were computed within each factor (DeVellis, 2003). According to Kline (1986), all items should correlate beyond .2 with the total score. As indicated in Table 5, three items (Items 63 and 8 from Factor 3 and Item 11 of Factor 4) had low item-total correlations (i.e., $\leq .20$) and could therefore be considered for exclusion in further scale development, per recommendations of Smith and McCarthy (1995). However, elimination of each of these items had no effect (i.e., appreciable gain) on coefficient alpha in this sample.

Factor Intercorrelations

Correlations between the factors were computed and are presented in Table 7. Correlations were positive and ranged from very low to modest (Cohen, 1982). The strongest correlation was between Factors I (Aggressive-Oppositional) and II (Attention-Seeking-Impulsive; $r = .59$). Factor II (Attention-Seeking-Impulsive) also correlated modestly with Factor III (Detached; $r = .41$).

If factors are correlated, then those correlations can be factored (Gorsuch, 1983, 2003). The factors from the items themselves are called primary or first-order factors. Those extracted from the primary factors are referred to as secondary or higher-order factors. Higher-order factor analysis reveals the extent to which the first-order factors share common variance (Briggs & Cheek, 1986).

Table 6

ASCA-H Factor Summary

	Factor I	Factor II	Factor III	Factor IV
	Aggressive- Oppositional	Attention-Seeking- Impulsive	Detached	Diffident
Nature of Factor:	Confrontational, threatening, destructive, and disruptive behaviors and ill temper	Inattention, impulsivity, and a desire for attention	Distant, aloof, and avoidant behaviors as well as lying and stealing	Timid, fearful behaviors
# Items:	37	26	13	11
Alpha:	.92	.85	.70	.65
Comparable Dimensions of ASCA:	Solitary Aggressive (Provocative), Solitary Aggressive (Impulsive) Oppositional Defiant	Attention-Deficit Hyperactive	Avoidant, Delinquent	Diffident
Sample Items:	Gets into fights in line	Answers before had time to think	Distant, makes no effort to relate to parent	Too timid to join in informal play

Table 7

Factor Correlations

Factor	Aggressive- Oppositional	Impulsive- Attention-Seeking	Detached	Diffident
Aggressive- Oppositional	-	-	-	-
Impulsive- Attention-Seeking	.59	-	-	-
Detached	.41	.38	-	-
Diffident	.06	.18	.15	-

Accordingly, the correlation matrix of the primary factors was factor analyzed using principal axis extraction. Results of the scree test indicated that one factor should be extracted. The one higher-order factor solution accounted for 36.75% of the variance and left two residuals (33% of the residual matrix) $\geq |.05|$, which suggested that an additional factor should be extracted (Cudeck, 2000, Gorsuch, 2003). As indicated in Table 8, the Aggressive-Oppositional, Impulsive-Attention-Seeking, and Detached first-order factors showed moderate to high loadings on the higher-order factor while the Diffident factor did not.

Table 8

One-factor secondary structure of the ASCA-H

First-Order Factor	Second-Order Factor
Aggressive-Oppositional	.56
Impulsive-Attention-Seeking	.60
Detached	.28
Diffident	.03

Note. Analyses were conducted using principal axis extraction.

Given the sizable residuals, a two-factor principal axis extraction with promax rotation was then conducted. The two-factor secondary structure accounted for 44.75% of the variance and left zero residuals $|\gt;.05|$. As reported in Table 8, only one primary factor loaded on the second higher-order factor. The Aggressive-Oppositional, Impulsive-Attention-Seeking, and Detached first-order factors were strongly associated with the first higher-order factor while the Diffident factor was strongly associated with the second higher-order factor.

Table 9

Two-factor secondary structure of the ASCA-H

First-Order Factor	Second-Order Factor I	Second-Order Factor II
Aggressive-Oppositional	.89	-.16
Impulsive-Attention-Seeking	.69	.14
Detached	.47	.15
Diffident	.04	.48

Note. Analyses were conducted using principal axis extraction and promax rotation.

Discussion

The ASCA-H (Watkins & McDermott, 2002) is an objective rating instrument completed by a parent or caregiver of children and adolescents aged 5 to 17 years that is currently in development. It was designed as a counterpart to the ASCA, a teacher-completed scale. The ASCA built upon limitations of traditional behavior rating scales with its unique format and multisituational approach. Rather than judge the general frequency or intensity of a symptom as on other rating scales, the respondent is asked on the ASCA to indicate which behaviors characterize a child's functioning with respect to specific situations. The ASCA defines psychopathology through demonstration of problem behavior across multiple situations and contexts.

By introducing a new format for obtaining parent report and assessing behaviors across different contexts, the ASCA-H could make a unique contribution to the field of behavioral assessment. In complement to the ASCA, it would follow a recommended multisource, multimethod assessment model and proposed guidelines for defining behavior disorders (CCBD, 1987; Gresham, 1985). As evidenced by the collective research on the ASCA, the cross-situational approach advocated by McDermott and colleagues has been effectively applied to a teacher-rating instrument, yet an unanswered question is whether this approach can also be applied to a parent version.

Thus far, no investigations of the ASCA-H have been reported. Given its early stage of development, the relevant aspect of construct validity to be addressed is the structural component. The purpose of the present study was to examine the underlying dimensions and structural validity of the ASCA-H.

One of the statistical methods for examining the latent structure of a scale is exploratory factor analysis. Exploratory analyses of the ASCA-H indicated that a four-factor model best represented the data. Four factors consistently emerged across rotation methods and number of factors extracted.

First-Order Factor Structure

Factor I. Factor I (Aggressive-Opositional) related to confrontational, threatening, destructive, and disruptive behaviors and ill temper. It encompassed aspects of three scales of the original ASCA - Solitary Aggressive (Provocative), Solitary Aggressive (Impulsive), and Oppositional Defiant. As such, Factor I of the ASCA-H did not distinguish between subtypes of aggression as on the ASCA.

The aggression scales of the ASCA relate to forms of aggression outlined by Dodge and Coie (1987). The Solitary Aggressive (Provocative) scale relates to proactive aggression. According to Dodge and Coie, proactive aggression is a deliberate aversive act aimed at achieving a desired goal, such as gaining dominance. In contrast, reactive aggressive behavior is a defensive response to perceived threat, frustration, or provocation. Reactive aggression relates to the ASCA Solitary Aggressive (Impulsive) scale.

Studies involving a teacher-rating instrument designed by Dodge and Coie (1987) to assess these forms of aggression have shown them to be highly correlated (Connor, Steingard, Cunningham, Anderson, & Melloni, 2004; Dodge & Coie; Poulin & Boivin, 2000; Price & Dodge, 1989). Using a more homogenous sample (i.e., aggressive boys only), Day, Bream, and Pal (1992) reported a modest correlation ($r = .41$) between the

two scales, in comparison to the stronger correlations of .76 and .83 observed by Dodge and Coie and Price and Dodge, respectively. Given the sizable correlations, factor analyses have not consistently shown reactive and proactive aggression to be two distinct dimensions of behavior as measured by this instrument (Dodge & Coie; Poulin & Boivin). Dodge and Coie concluded that teachers use primarily a single dimension to characterize aggressive behavior but also contended that there was modest support for distinguishing between reactive and proactive aggression based on higher within-scale correlations than between-scale correlations.

Other studies have supported the discriminant validity of reactive and proactive aggression. Differences between reactively and proactively aggressively children have been found with regard to social information processing (Dodge & Coie); peer status, social withdrawal, leadership, and victimization (Poulin & Boivin); temperament, delinquency, and depression (Vitaro, Brendgen, & Tremblay, 2002); and substance use, family history of substance abuse, and family violence (Connor et al). Concerning their equivocal findings with regard to discriminant validity, Dodge and Coie (1987) argued, “Teachers (like most adults) do not sufficiently attend to, or at least store in memory and utilize, subtle structural and functional differences in subtypes of aggression. Because teacher attention is usually directed toward aggressive interaction only at the end of aggression sequences, they may have limited opportunity to observe the elements of social interactions that are the distinguishing features of the types of behavior. Thus, they may tend to chunk the two types into one broader category of aggression” (p. 1156).

In contrast to the instrument developed by Dodge and Coie (1987), sub-types of aggression have been clearly identified using the ASCA. The Solitary Aggressive (Provocative) and Solitary Aggressive (Impulsive) scales of the ASCA have been shown to be reliable and distinct dimensions of child psychopathology with modest inter-scale correlations ($r = .45$; McDermott, 1993, 1994). The specificity of the aggression scales of the ASCA seems to suggest that the ASCA, which differed in item breadth (i.e., number of related items) and format (i.e., present/absent versus 1- to 5-point frequency scale) might be a better means for teachers to report on such behaviors.

In a related study involving trained observers as well as teachers, Dodge and Coie (1987) argued that adult observers could be taught to reliably distinguish between the two forms of aggression. However, given their lack of training and comparative experiences, parents might not be as capable as teachers in making such distinctions, even if given a different method for reporting, such as the ASCA-H. This consideration, along with the collective research clearly showing the close association between reactive and proactive aggression, could then explain the high correlations between these behaviors and subsequent factor composition found in this study.

Moreover, parent-rating measures of aggression have been shown to be substantially correlated with measures of oppositional, defiant behaviors and conduct problems. For example, Reynolds and Kamphaus (1992) reported inter-scale correlations of .66 in the children's and .56 in the adolescent's general norm samples for the Aggression and Conduct Problems scales of the BASC – PRS. Also, in exploratory factor analytic work with the Disruptive Behavior Rating Scale – Parent Version (DBRS-P;

Erford, 1993), the original, rationally-derived oppositional and aggression items combined to emerge as an Oppositional factor (Erford, 1996, 1998). (Inter-scale scale correlations were not reported.) Similarly, oppositional and aggressive behaviors constituted a single dimension on the ASCA-H, which is also consistent with the structure of the CBCL.

Factor II. Factor II (Impulsive-Attention-Seeking) consisted of items dealing with inattention, impulsivity, and a desire for attention. Factor II of the ASCA-H was comparable to the ASCA's Attention-Deficit Hyperactive scale. On both the ASCA and ASCA-H, attention problems and hyperactivity constituted one dimension, similar to the structure of the CBCL. The BASC, on the other hand, differentiates between the two with separate scales of Hyperactivity and Attention Problems. However, similar to the previous discussion of high intercorrelations between scales, the Hyperactivity and Attention Problems scales of the BASC-PRS have been shown to be moderately correlated (i.e., average correlation of .55 across the three age-levels; Reynolds & Kamphaus, 1992).

Confirmatory factor analytic studies involving narrowband measures of ADHD symptoms (e.g., ADHD Rating Scale IV; DuPaul, Anastopoulos, Power, Murphy, & Barkley, 1994) have found that a two-factor model consisting of inattention and hyperactivity/impulsivity provided a better model for the organization of parent and teacher ratings of these symptoms than a one-factor model (Burns, Boe, Walsh, Sommers-Flanagan, & Teegarden, 2001; DuPaul et al., 1998; DuPaul et al., 1997). At the same time, these studies have also shown high correlations between the two dimensions.

For instance, Burns and colleagues (Burns, Walsh, Owen, & Snell, 1997; Burns, Walsh, Patterson, et al., 1997; Burns, Boe, et al., 2001) reported correlations ranging from .68 to .80. DuPaul et al. (1997, 1998) found a correlation of .74 for teacher ratings and .92 for parent ratings. DuPaul et al. argued that the two-factor model allowed for identification of sub-types of ADHD consistent with DSM-IV diagnostic formulations of ADHD. However, the high intercorrelations found across studies seem to raise questions regarding the independence of the dimensions and the appropriateness of differential interpretation.

Factor III. Factor III (Detached) consisted of distant, aloof, and avoidant behaviors, as well as lying and stealing. This factor reflected the ASCA's Avoidant scale but also included elements of the ASCA's supplementary Delinquent scale. Although intercorrelations between the Avoidant and Delinquent scales have not been reported (McDermott, 1993, 1994), a relationship between the Avoidant items has been established with another Delinquent scale. That is, in a concurrent validity study, the ASCA Avoidant scale was most highly associated with the Delinquent ($r = .59$; girls' norm group) and Cruel ($r = .46$; girl's norm group) scales of the 1983 version of the CBCL (McDermott, 1994). With its inclusion of Delinquency items, the Detached factor of the ASCA-H differs from the composition of the Avoidant factor of the ASCA.

Both factors, however, appeared to be unique aspects of psychopathology when compared to the structure of other broadband parent rating instruments. The Avoidant syndrome is one of the ASCA's Underactivity scales. The Underactivity scales have been associated with the Internalizing dimension of other instruments, such as the CBCL

(McDermott, 1993, 1994; Canivez, 2004), yet no other instrument has an internalizing behavior scale comparable to the Avoidant scale. The BASC, for example, includes Anxiety, Depression, Somatization, and Withdrawal scales. The CBCL has Anxious/Depressed, Withdrawn, and Somatic Complaints scales.

Achenbach and McConaughy (1997) discussed a link between the Withdrawn syndrome of the CBCL and Avoidant Disorder of Childhood or Adolescence included in the DSM-III and DSM-III-R. (The DSM-IV does not include this diagnostic category but instead includes a category of Avoidant Personality Disorder intended primarily for adults.) According to Achenbach and McConaughy, Withdrawn items reflect “low engagement from other people, low energy, and some negative affectivity” (p. 63). Avoidant Disorder, as described by Barrios and Hartmann (1988), consists of a mixture of persistent and excessive withdrawal from strangers, clear and sufficient disturbance in peer relationships, healthy and satisfying relationships with family members, and sufficient desires for affection and acceptance.

This description, particularly the last component, is incongruent with the composition of the ASCA-H Detached and ASCA Avoidant factors. In fact, the ASCA-H and ASCA factors contain items reflecting a lack of concern or desire for attention from others. Furthermore, even though they include withdrawn or disengaged behaviors, both lack the element of affectivity (i.e., shy, looks sad, sulks) described by Achenbach and McConaughy (1997) in the CBCL Withdrawn syndrome.

This disparity might be attributed to the different assessment approaches of the ASCA/ASCA-H and other traditional instruments. In contrast to the CBCL and BASC

that were purportedly designed to measure both internalizing and externalizing behavior problems (Achenbach, 1991a; Reynolds & Kamphaus, 1992), McDermott et al. (1995) noted that the ASCA “is not a measure of behaviorally indistinct pathology such as anxiety or depression – constructs requiring inferences about feelings or emotions” (pg. 349). The behavioral indicators and situations of the ASCA were designed to provide clear content meaning and intended to reduce the need for respondents to make inferences regarding children’s thoughts or feelings (McDermott, 1993, 1994). In comparison, the CBCL and BASC require the respondent to rate the frequency or severity of symptoms. In the case of internalizing problems, which “mainly involve inner distress” (Achenbach & McConaughy, 1997, pg. 56), these instruments sometimes require the respondent to judge or speculate about the child’s feelings, thoughts, or moods. Thus, it appears that the ASCA and ASCA-H are not tapping the same Internalizing constructs as other instruments.

Factor IV. Factor IV (Diffident) related to timid and shy behaviors. Factor IV was very similar to the Diffident scale of the ASCA. As with the Detached factor of the ASCA-H and the Avoidant factor of the ASCA, no other current parent rating instruments have a parallel scale to the Diffident factor. The comparable Diffident scale of the ASCA was shown to correlate highest with the Uncommunicative ($r = .59$; boys’ norm group) and the Obsessive-Compulsive ($r = .46$; boys’ norm group) scales of the 1983 version of the CBCL (McDermott, 1994). These scales, however, were not identified as core syndromes common to most of the age and gender groups in the 1991 revision of the CBCL and were therefore not included in the revision (Achenbach,

1991a). The 1991 version of the CBCL includes separate Withdrawal and Anxious/Depressed scales. No known studies have examined the empirical relationship between the ASCA and 1991 CBCL. Conceptually, however, the Diffident factor shares the shyness aspect of the CBCL Withdrawal scale, but consistent with the behavioral focus and intent of the ASCA (McDermott, 1993, 1994), it does not include similar Anxious/Depressed items that reflect subjective psychological states, such as worry or fear.

Factor Intercorrelations

Correlations between ASCA-H factors were positive and ranged from very low to modest (Cohen, 1982). The strongest correlation was between Factors I (Aggressive-Oppositional) and II (Attention-Seeking-Impulsive; $r = .59$). The weakest correlation was between Factors I (Aggressive-Oppositional) and IV (Diffident; $r = .06$).

Other parent rating measures have shown inter-scale correlations of similar magnitude. For example, intercorrelations of the maladaptive behavior scales of the BASC-PRS ranged from .05-.67 for the preschool level, .07-.66 for the child level, and .24-.63 for the adolescent level (Reynolds & Kamphaus, 1992). On the CBCL, problem scale correlations ranged from .13-.62 for boys aged 4-11, .19-.64 for boys aged 12-18, .09-.57 for girls aged 4-11, and .29-.65 for girls aged 12-18 (Achenbach, 1991a).

Factor I of the ASCA-H, as mentioned, did not represent a dimension of distinctly aggressive behavior but rather was composed of aggressive and oppositional behaviors. It correlated highest with the ASCA-H factor that reflected attention problems and impulsive behavior (Factor II). As such, these results corresponded with patterns of inter-

scale correlations of other parent rating scales. For example, the strongest intercorrelations of the BASC-PRS maladaptive scales across age levels were between the Hyperactivity and Aggression scales (mean $r = .65$; Reynolds & Kamphaus, 1992). Attention Problems and Aggression were the most highly correlated scales of the CBCL among girls aged 4-11 and 6-12 (Achenbach, 1991a) as well. ADHD and aggression have been shown to be separate but highly correlated syndromes (Hinshaw, 1987). However, the moderate to high inter-scale correlations has caused some to criticize rating scales for overlapping too extensively to allow for differential interpretation (McDermott, 1993).

Higher-Order Factor Structure

Given the moderate correlations between the ASCA-H's primary factors, higher-order factor analysis was also conducted. Multivariate analyses of child psychopathology have frequently identified a two-factor structure. The two-factor dichotomy has been referred to as Personality Problem versus Conduct Problem (Peterson, 1961), Inhibition versus Aggression (Miller, 1967), Overcontrolled versus Undercontrolled (Achenbach & Edelbrock, 1978), Internalizing versus Externalizing (Achenbach, 1991a, 1991b; Reynolds & Kamphaus, 1992), and Overactivity versus Underactivity (McDermott, 1993, 1994; Canivez, 2004). A two-factor higher-order structure was also found in this study.

The first higher-order factor was saliently loaded by Factors I (Aggressive- Oppositional), II (Impulsive-Attention-Seeking), and III (Detached). This higher-order factor seemed to tap an overarching externalizing behavior dimension similar to the Overactivity factor of the ASCA as well as the Externalizing scale of the CBCL and the Externalizing Problems composite of the BASC. The CBCL and BASC include a

separate, first-order Attention Problems scale that is not included on the higher-order Externalizing dimension, although it has showed moderate loadings with this dimension on both instruments (Achenbach, 1991a; Reynolds & Kamphaus, 1992).

Only one primary factor (Factor IV - Diffident) loaded on the second higher-order factor, which seemed to tap an Internalizing dimension congruent with other measures of behavior pathology. Factor III (Detached), which would have seemed more theoretically related to an Internalizing or Underactivity higher-order factor, actually loaded on the first higher-order factor. This result might have been due to the fact that the Detached factor was comprised of items from the Delinquent scale of the ASCA in addition to items from the ASCA's Avoidant scale. The Avoidant scale, which loaded on the higher-order Underactivity dimension (similar to the Internalizing dimension of other scales), has been shown to correlate with the Delinquent scale of the 1983 version of the CBCL, an Externalizing behavior scale. Thus, the Detached factor seems to contain a strong externalizing component.

With its single loading of the Diffident factor, the second higher-order factor was not well represented. This result might be attributed to the manner in which internalizing/underactive behaviors are assessed by the ASCA-H. As previously discussed, the ASCA and by association the ASCA-H were not intended to measure syndromes such as anxiety and depression per se (McDermott et al., 1995). As argued by Canivez (2004), their "internalized nature" makes such problems "difficult or impossible for third parties to observe and report" (p. 198). Items on the ASCA and ASCA-H were written in objective, behavioral terms in order to be readily and easily observed. This type

of approach would then preclude items pertaining to feelings and moods, which are part of diagnostic criteria for anxiety and depressive disorders (APA, 1994). For instance, the Diffident factor tapped shy and timid characteristics that are related to anxiety but do not directly measure internalized or subjective states symptomatic of depressive and anxiety disorders.

In contrast, traditional measures, such as the CBCL and BASC, which are intended to assess anxiety and depression, include items that require the respondent to speculate about a child's thoughts or feelings. Semrud-Clikeman, Bennett, and Guli (2003) argued that the Anxious/Depressed scale of the CBCL "tends to focus on the inner life of the child rather than on environmental stressors or contributions to problem behavior" (p. 274). They also noted that the items of the CBCL "tap information that can be quite inferential in nature and may require the informant to make judgments based on the child's behavior. For this reason, the reliability of the scale is lower than desired" (p. 274). Achenbach (1991a) reported an average inter-rater reliability of .66 for the Anxious/Depressed scale. In the same way, the Anxiety and Depression scales of the BASC-PRS have lower inter-rater reliabilities (i.e., average inter-rater reliabilities of .54 and .61, respectively, across age levels; Reynolds & Kamphaus, 1992). Achenbach acknowledged the subjectivity implicit in some of the CBCL items (e.g., Item 35 *Feels worthless or inferior* and Item 52 *Feels too guilty*) but argued that parents "are usually in a better position than most other people to judge and report their children's expressions of feelings" (pp. 226-227).

Still, the use of feeling terms in rating scales has been criticized due to the ambiguity and subjectivity they present for the respondent (Kline, 1993). The more explicit the trait being measured (e.g., observable behavior) and the less inference required by the respondent, the lower the variance in the item due to rater bias (Hoyt & Kerns, 1999). However, measuring internalizing problems in behavioral terms given their “covert symptomatology” (Reynolds, 2004, pg. 232) seems to have its limitations.

Mash and Terdal (1988) contended that depression and anxiety require direct assessment of children’s thoughts and feelings. Children’s self-reports represent the most direct and accurate expression of personal thoughts and subjective feelings (Michael & Merrell, 1998). Whereas adults are generally considered better informants for observable behavior, child reports may be more informative for subjective phenomena such as internal states and feelings (Piacentini, Cohen, & Cohen 1992; Silverman & Rabian, 1999). However, young elementary-school-age children may be unable to describe themselves adequately through response to questionnaires, as they are unlikely to have mastered the range of vocabulary necessary to adequately describe aspects of adjustment (Lachar & Gruber, 2003). Thus, the use of children’s self-reports in assessment also has its shortcomings.

From this discussion, the complexity of assessing internalizing behavior problems is evident. More difficult still is assessing internalizing problems in an objective, observable manner. As such, although results of this study support a two-factor dichotomy of psychopathology, the internalizing dimension was not as well represented as the externalizing dimension.

The two-factor model of the ASCA-H explained 48.75% of the variance among the factors, which was somewhat higher than the 40.8% reported by McDermott (1993, 1994) for the second-order structure of the ASCA. Even though this result leaves the majority of the variance unexplained, it is not uncommon, particularly in personality assessment.

Kline (1998) argued that the abstract nature of psychological phenomenon, particularly personality variables, makes measurement of such variables more difficult than in the natural sciences. In the case of intelligence, which Kline described as “the easiest” (p. 28) of psychological variables to measure, there is some agreement as to what intelligence tests measure. However, only 45% of the total test variance is accounted for by the Verbal-Performance second-order structure of the widely used and researched cognitive measure, the Wechsler Intelligence Test for Children – Third Edition (WISC-III; Wechsler, 1991). According to Kline, measurement of a personality construct, such as dominance, is even more difficult, as “it is by no means clear what items might measure dominance or how it might be shown that a group of items did in fact measure that variable. Actually, the very definition of the dominance might be hard to agree upon” (p. 28). If cognitive instruments cannot fully tap the more concrete construct being measured, then personality measures will likely fall short and leave a high percentage of variance unexplained as well.

Test manuals for commonly used parent rating scales including the BASC and CBCL did not report this information nor was it included in factor analytic studies of these instruments. However, results of an exploratory study of the Revised Conners

Parent Rating Scales (CPRS-R) indicated that its seven factors accounted for 50.8% of the total variance (Conners, Sitarenios, Parker, & Epstein, 1998). Similarly, father responses to the Disruptive Behavior Rating Scale (Erford, 1993) generated a four-factor solution that accounted for 49% of the test variance (Erford, 1998). Likewise, Brown, Aman, and Havercamp (2002) reported that parent ratings on the Aberrant Behavior Checklist – Community (ABC-C; Aman & Singh, 1994) produced a four-factor solution that accounted for 48% of the total common variance. Results of these studies indicate that often more of the variance in personality scales is left unexplained than explained. Even with its unique format and behavioral focus, the ASCA-H is no exception.

Internal Consistency

The internal consistency reliability of the ASCA-H factors ranged from .65 (Factor IV) to .92 (Factor I). Factor I (Aggressive-Oppositional) had the highest alpha value, which was consistent with conceptually-similar scales of other instruments. For example, the Aggression scale of the CBCL had the highest alpha values (.92) for all gender and age groups (Achenbach, 1991a). Likewise, Naglieri, LeBuffe, and Pfeiffer (1994) reported median alpha estimates of .95 and .94 for males and females aged 5-12 respectively and values of .99 and .96 for males and females aged 13-18 respectively for the Conduct scale of the Devereux Scales of Mental Disorders (DSMD; Naglieri et al.). The relatively higher internal consistency reliability of Factor I, which was comprised of 37 items, the most of any of the ASCA-H factors, was also anticipated given that reliability increases with test length (Kline, 1998).

Clark and Watson (1995) advised that once internal consistency reliabilities of .80 are achieved, there is no need to strive for any substantial increases. Similarly, Bracken (1987) also recommended a minimum reliability of .80 for subscales. However, Factors III and Factor IV of the ASCA-H did not meet this standard.

Alpha values for internalizing behavior subscales of other rating instruments have met or fallen just below the minimal standard of .80. For instance, across age levels in the general norm sample, the Depression and Anxiety scales of the BASC-PRS have average coefficient alphas of .80 and .82, respectively. The average alpha values for the Somatization and Withdrawal scales are .76 and .74, respectively (Reynolds & Kamphaus, 1992). The average coefficient alpha for the Anxious/Depressed scale of the CBCL is somewhat higher (.87), but the internal consistency reliabilities of the Withdrawn and Somatic Complaints scales of the CBCL are .79 and .74, respectively when averaged across gender and age groups (Achenbach, 1991a).

Higher alpha values could be expected for instruments, such as the BASC and CBCL, which utilize a multipoint rating scale because the multipoint scale generates more item variance and increases coefficient alpha. In contrast, the dichotomous nature of the ASCA-H items often contributes to lower alpha values. Alpha values for Factors III and IV of the ASCA-H were considered minimally adequate for research purposes (DeVellis, 2003) but failed to meet the higher standard required for assessment and decision-making, thereby indicating the need for scale refinement. According to Clark and Watson (1995), if the alpha value of a new scale falls below .80, then a scale should be revised to try to increase the reliability to an acceptable level.

Coefficient alpha is related both to the number of items on a scale and the extent of covariation among the items (DeVellis, 2003). Consequently, it can be improved by increasing the number of items or by selecting items that are more cohesive in order to improve the average inter-item correlation (Briggs & Cheek, 1986). Briggs and Cheek advocated that the mean inter-item correlation of a scale should be in the range of .2 to .4. They argued that it is unlikely that a single total score could adequately represent the complexity of the items if the average inter-item correlation was lower than .1. When the average correlation is greater than .5, the items tend to be highly redundant and the construct measured too specific (Briggs & Cheek). The mean inter-item correlations for both Factors III and IV fell below the recommendation, with correlations of .16 and .15, respectively.

The intercorrelations of the items of these factors were examined in another way. Corrected item-total correlations correlate each item with all the factor items excluding itself (DeVellis, 2003). Three items (Items 63 and 8 from Factor III and Item 11 of Factor IV) had low item-total correlations (i.e., $\leq .20$; Kline, 1986) and could therefore be considered for exclusion in further scale development, per recommendations of Smith and McCarthy (1995). However, elimination of each of these items had no appreciable effect on coefficient alpha in this sample.

Similarly, when the inter-item correlations of Factors III and IV were inspected, it was found that both Item 8 of Factor III and Item 11 of Factor IV negatively correlated with another item in the factor and had very low (Cohen, 1982) or negligible relationships with most of the items. Specifically, Item 8 correlated .18 or below with

other items of Factor III, and Item 11 correlated .20 or below with all items of Factor IV. If these items were deleted, however, as advised by DeVellis (2003), again no improvement in coefficient alpha resulted in this sample.

Smith and McCarthy (1995) defined poor items as “those that have low item-total correlations, low correlations with other items, and their removal would lead to an increase in coefficient alpha” (pg. 305). Although properties of items 8, 11, and 63 were cause for consideration for elimination in scale revision, their deletion did not improve the reliability estimate in this sample. Therefore, at this point, these items should not be deemed “poor” and worthy of exclusion without further support.

Item Endorsement

Another item property examined during scale development is the percentage of the sample selecting the keyed response (item endorsement rate or prevalence). Two items were unendorsed in the sample and thus excluded from factor analysis. The unendorsed items related to drug and alcohol use, behaviors more associated with adolescents and adults. Consequently, lack of endorsement is most likely attributable to the nature of the sample, which was comprised of only two students above the age of 14, and not the scale itself. Therefore, these items should not be eliminated based on the current results.

Forty of the 149 situationally-specific problem items had endorsements rates below 5%. One item had an endorsement rate above 95%. It is sometimes recommended to eliminate such items with unbalanced response distributions during scale development (Schwartz, 1978). However, items might perform differently across samples such that an

item with a low endorsement rate in one sample might discriminate in a clinical population (Reise et al., 2000). Clark and Watson (1995) thus noted that before excluding an item based on its distribution, it is essential to examine data from diverse samples representing the entire range of the scale's target population. Although efforts were made to obtain a heterogeneous sample that included students with a range of problem behaviors, not all age ranges for which the ASCA-H is intended were part of the final sample. Therefore, items with unbalanced distributions should be retained at this point due to limitations of the sample and not the performance of the items themselves. Additionally, the ASCA-H is a measure of behavioral and social-emotional adjustment problems and, so, as argued by McDermott (1994), must have enough variability to support reliability and discrimination. Thus, the most and least frequently endorsed items should not at this point be eliminated.

To summarize, based on examination of inter-item relationships and endorsement rates, no items should be eliminated from the ASCA-H at this time. However, this recommendation is made without benefit of more extensive item analyses. That is, formative information regarding "good" and "bad" items was limited in this study because the size of the sample precluded further assessment of items, such as analyzing differences in responding relative to gender, age, or ethnic groups. At this point, given the limited conclusions that could be drawn, it seems advisable that all items be retained for further investigation.

Scale Refinement

According to Clark and Watson (1995), “good scale construction typically is an iterative process involving several periods of item writing, followed in each case by conceptual and psychometric analysis. These analyses serve to sharpen one’s understanding of the nature and structure of the target domain as well as to identify deficiencies in the initial item pool” (p. 311). As was the case in this study, exploratory factor analysis might indicate that the items can be separated into several subscales but also reveal that the initial item pool does not contain enough items to assess each of the hypothesized domains reliably. Initial research may demonstrate that a measure is missing an important aspect of a construct or not tapping a researcher’s intended dimension.

Not all hypothesized dimensions of the original ASCA were identified with the ASCA-H using the current sample. The ASCA is comprised of six dimensions that generalized across subgroups and two supplementary dimensions that did not generalize across all groups. Two of these eight dimensions were identified clearly in the ASCA-H: Factor II (Impulsive-Attention-Seeking) was comparable to the ASCA’s Attention-Deficit Hyperactive scale, and Factor IV (Diffident) was very similar to the Diffident scale of the ASCA. Factor I (Aggressive-Oppositional) represented elements of three ASCA scales (i.e., Solitary Aggressive (Provocative), Solitary Aggressive (Impulsive), and Oppositional Defiant). Factor IV (Detached) was similar to the ASCA’s Avoidant scale but also contained aspects of its Delinquent scale as well. A dimension comparable

to the ASCA's supplementary scale, Lethargic (Hypoactive), was not identified on the ASCA-H.

Clark and Watson (1995) advised, "If items that reflect the theoretical core of the construct do not correlate strongly with it in preliminary analyses, it is not wise to simply eliminate them without consideration of why they did not behave as expected" (p. 314). One such consideration is whether the sample is nonrepresentative in some important way. As mentioned, this study's sample did not include a sufficient number of older adolescents so as to truly represent the entire range of students on whom the ASCA-H is intended, as is suggested during early test development (Clark & Watson). As such, the lack of a unidimensional conceptualization of a Delinquency factor on the ASCA-H could be due to the lack of antisocial behaviors generally displayed by children age 5-14 years who made up this study's sample. Conduct problems are more prevalent during adolescence (APA, 1994). Thus, a sample that included older students would likely be necessary to establish the presence of such a factor.

Also, this study's sample was not nationally representative as was the ASCA's standardization sample. Therefore, results of this study would not be expected to mirror those of the ASCA exactly. However, these results can be used to help in refining the instrument. They help to clarify the underlying nature of the ASCA-H and can be used to help modify it to make it more comprehensive and more consistent with the literature (Segal & Collidge, 2004).

In the current study, a four-factor solution best represented the data; however, the four factors accounted for only 18.42% of the total variance. Heterogeneity among the

items was also indicated by a considerable number of variables that failed to load on any factor (Tabachnick & Fidell, 1996). According to Floyd and Widaman (1995), results of factor analysis can be used to reduce the number of items by eliminating items that load comparably on more than one factor or failed to load on a factor. Complex variables should be avoided because of the ambiguity they create in interpretation of factors (Tabachnick & Fidell). Kline (1986) argued that if two factors account for the variance among items, those items loading on one factor should be separated and other similar items written. Items should be selected that load on only one factor (Kline). Only three ASCA-H items were complex. However, 59 items failed to load on any factor, and to simply eliminate them does not solve the problem. The goal is to create unidimensional factors that are internally consistent and reduce the unexplained variance (Smith & McCarthy, 1995).

As mentioned, exploratory analyses have consistently identified a two-factor higher-order structure of externalizing and internalizing dimensions. While the current study supported a two-factor dichotomy, the second, internalizing dimension was not well represented. Relatedly, the two primary factors conceptually related to internalizing behaviors (Diffident and Detached) did not meet accepted standards for adequate internal consistency reliability. The Detached factor in particular was complex, comprised of aspects of both internalizing and externalizing problems. Therefore, one implication from the present study is refinement of the internalizing behavior items.

Because conclusions regarding what items could be deemed “poor” were limited in this study, it was suggested that all items be retained for subsequent research. Thus,

scale refinement would likely involve addition of new items that would then be subjected to psychometric analyses (Clark & Watson, 1995). Reise et al. (2000) recommended writing multiple sets of “content homogeneous items” (p. 288) that tap different aspects of the target construct. They advised that the new item pool be substantially overinclusive, as it is easier to subtract rather than add items. This practice would help to ensure adequate construct representation, which was not indicated by the factor structure of the current ASCA-H. Developing additional items would help better define the hypothesized factors, as increasing the number of marker variables of each factor improves factor stability (Floyd & Widaman, 1995). Subsequent factor analysis would then be better able to provide empirical evidence of what type of item content belongs in the construct and what belongs somewhere else.

Refinement is also suggested with regard to externalizing behavior items. Aggression and oppositional, defiant behaviors constituted a single dimension on the ASCA-H. Although this is consistent with the structure of some existing measures, such as the CBCL and DSMD, other test developers have argued that aggression and conduct problems are correlated but distinct facets of behavior that can be measured reliably and independently (Reynolds & Kamphaus, 1992). Furthermore, sub-types of aggression have been defined (e.g., Dodge & Coie, 1987). McDermott (1993, 1994) found empirical support for two distinct aspects of aggression as well as a separate dimension of oppositional-defiant behavior. Distinct dimensions, however, were not identified using the current ASCA-H with respect to the current sample.

To further investigate the dimensionality of externalizing behavior problems among parent ratings, parallel items may be written. The item pool should then be retested on a heterogeneous sample that represents the entire target population (Clark & Watson, 1995) and all items factor analyzed together (Kline, 1986). If distinct dimensions were found in another, more representative sample with a revised scale, results would suggest that aggression and oppositionality could be reliably measured and distinguished using a multisituational approach and parents as respondents, as was done successfully with the ASCA and teachers as raters.

Limitations

Although demographic information obtained for the school communities suggested a diverse, heterogeneous pool of students, information regarding race/ethnicity and SES for the actual students could not be obtained due to the school district's requirements for research approval. In addition, all students were from a limited geographic area. Therefore, the students in this sample cannot be considered representative of the U. S. population and generalization of these results cannot be made.

Another limitation was that the sample did not include the entire range of students for whom the scale is intended, as is suggested during early stages of test development (Gorsuch, 1997; Kline, 1993). The sample was heavily weighted with elementary-age students (75% of the students rated were younger than 12 years of age). The group chosen to represent the more extreme range of problem behavior was also likely overrepresented in the final sample (17% of the sample), which likely influenced the results.

Also, the size of the sample precluded more extensive analyses. For instance, separate analyses by gender and age could not be conducted to examine the invariance of the factor structure across subgroups. A larger sample size would have allowed the data set to be divided in half and analyses performed on both groups. Further item analyses and stability of the factor structure could then be examined on split halves of the data set to see if the solution was robust across subsamples. In scale development, “one of the chief considerations should always be the replicability of the factor structure. Factors that do not replicate are of little value” (Briggs & Cheek, 1986, p. 120). It is important to ensure that resultant factors are replicable and generalizable to subgroups. Such analyses, however, would require a larger sample than was available in the present study.

Conclusions

Evidence of construct validity requires an accumulation of research results that support hypotheses consistent with the construct being measured and the “nomological net” surrounding it (Cronbach & Meehl, 1955). One step of the validation process involves establishing support for the factor structure of the scale. The present study presented preliminary information regarding the latent structure of the ASCA-H.

Results suggested the need for additional items and further exploratory work. In future investigations, the new item pool should be retested on a large, heterogeneous sample that represents the entire target population of the ASCA-H (Clark & Watson, 1995). With a sufficient number of participants, each one could be randomly assigned to one of two groups and item analyses performed for each sample. Analyses should be done to demonstrate the robustness of the factor solution across split-halves of the sample

as well as checking for differences in responding across gender, age, and ethnicity. The consistency of the two resulting factor solutions could then be evaluated by computing a coefficient of factor similarity, such as the coefficient of factor congruence (Wrigley & Neuhaus, 1955).

Kline (1986) emphasized the need for replication of factors and invariance of factor patterns across samples and studies. Further investigation of all ASCA-H items collectively should therefore be conducted. Factor replicability can also be examined using confirmatory factor analysis procedures. Replication across groups would provide evidence of the ASCA-H's structural integrity (invariance) and reliability generalization across ages, genders, and ethnicity.

Although replication of factor structure is vital in scale development, identifying a replicable factor structure does not ensure that the importance or utility of those factors (Briggs & Cheek, 1986). A second level of validation that examines the relationships between factors and other measures is then necessary. Construct validity requires "a test be anchored in a conceptual framework or theory that delineates the meaning of the construct, its uniqueness, and its relationship to other variables measuring similar domains" (Segal & Collidge, 2004, p. 9).

Simply stated, once an instrument has been developed and internal analyses performed, subsequent research is needed to show that the instrument correlates with variables it should and does not correlate with variables with which it should not correlate. Future research should examine the question of how the ASCA-H is related to "measures that bear conceptual resemblance" (Briggs & Cheek, 1986, p. 120), such as the

ASCA, BASC, and CBCL. It is also important to know how the ASCA-H correlates with important variables such as age, socioeconomic status, and IQ and how its mean and standard deviation vary across groups.

In addition to showing what the ASCA-H is measuring, future research should examine how it can be used in practice. For example, studies looking at its clinical utility, such as its ability to differentiate among groups and its ability to add to the predictive power of demographics and cognitive functioning to predict future academic achievement will be important.

Ultimately, the important question to be addressed is whether the ASCA-H helps improve upon the ability of parents to report on their children's behavior. Teachers have often been considered better raters of behavior, but the contribution of parents to a comprehensive assessment is also recognized and recommended. Using teachers as respondents, the ASCA helped advance the assessment of youth psychopathology with its emphasis on the multisituational expression of problems behaviors and objective, behaviorally-oriented format. This approach, which provides contextual information to assist in ratings, less clinical terminology, and less subjectivity (i.e., inferring meaning regarding feelings and thoughts), may prove to be a better means for parents to report. As such, it has the potential to build upon limitations of traditional parent-rating instruments, and as part of a recommended multimethod, multisource, multisetting assessment battery, the ASCA-H could not only compliment other assessment methods but also provide a unique contribution. The potential, however, remains to be realized through the support of future work.

References

- Achenbach, T. M. (1991a). *Manual for the CBCL/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. S. (1986). *Manual for the Teacher's Report Form and Teacher Version of the Child Behavior Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., Howell, C. T., & McConaughy, S. H. (1995). Six-year prediction of problems in a national sample of children and youths. *Journal of American Academy of Clinical and Adolescent Psychiatry, 34*, 336-347.
- Achenbach, T. M., & McConaughy, S. H. (1997). *Empirically based assessment of child and adolescent psychopathology* (2nd ed.). Thousand Oaks, CA: Sage.
- 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.
- Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behavior, personality, and attitudes*. New York: Wiley.

- Aiken, L. R. (1997). *Psychological testing and assessment* (9th ed.). Boston: Allyn & Bacon.
- Aman, M. G., & Singh, N. N. (1994). *Supplement to the Aberrant Behavior Checklist manual*. East Aurora, NY: Slosson.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Barkley, R. A., & Edelbrock, C. (1987). Assessing situational variation in children's problem behaviors: The Home and School Situations Questionnaires. In R. Prinz (Ed.), *Advances in behavioral assessment of children and families* (pp. 157-176). Greenwich, CT: JAI.
- Barrett, P. T., & Kline, P. (1982). Factor extraction: An examination of three methods. *Personality Study and Group Behavior*, 3, 84-98.
- Barrios, B.A., & Hartmann, D.P. (1988). Fears and anxieties. In E. J. Mash & L. G. Terdal (Eds.), *Behavioral assessment of childhood disorders* (pp. 196-262). New York: Guilford Press.
- Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximation. *Journal of the Royal Statistical Society*, 16, 296-298.

- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*, 10-22.
- Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105*, 467-477.
- Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*, 313-326.
- Breen, M. J., & Altepeter, T. S. (1991). Factor structures of the Home Situations Questionnaire and the School Situations Questionnaire. *Journal of Pediatric Psychology, 16*, 59-67.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106-148.
- Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 28*, 25-56.
- Brody, G. H., & Forehand, R. (1986). Maternal perceptions of child maladjustment as a function of the combined influence of child behavior and maternal depression. *Journal of Consulting and Clinical Psychology, 54*, 237-240.
- Brown, E. C., Aman, M. G., & Havercamp, S. M. (2002). Factor analysis and norms for parent ratings on the Aberrant Behavior Checklist – Community for young people in special education. *Research in Developmental Disabilities, 23*, 45-60.

- Burns, G. L., Boe, B., Walsh, J. A., Sommers-Flanagan, R., & Teegarden, L. A. (2001). A confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: What is the best model for the organization of these symptoms? *Journal of Abnormal Child Psychology, 29*, 339-349.
- Burns, G. L., Walsh, J. A., Owen, S. M., & Snell, J. (1997). Internal validity of the ADHD, CD, and overt CD symptoms in young children: Implications from teacher ratings for a dimensional approach to symptom validity. *Journal of Clinical Child Psychology, 26*, 266-275.
- Burns, G. L., Walsh, J. A., Patterson, D. J., Holte, C. S., Sommers-Flanagan, R., & Parker, C. M. (1997). Internal validity of the disruptive behavior disorder symptoms: Implications from parent ratings for a dimensional approach to symptom validity. *Journal of Abnormal Child Psychology, 25*, 305-319.
- Burns, G. L., Walsh, J. A., Patterson, D. J., Holte, C. S., Sommers-Flanagan, R., & Parker, C. M. (2001). Attention deficit and disruptive behavior disorder symptoms: Usefulness of a frequency count rating procedure to measure these symptoms. *European Journal of Psychological Assessment, 17*, 25-35.
- Bryant, F. B. (2000). Assessing the validity of measurement. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 99-146). Washington, D.C.: American Psychological Association.
- Canivez, G. L. (2001). Review of the Adjustment Scales for Children and Adolescents. In J. Impara and B. Plake (Eds.), *The fourteenth mental measurements yearbook*

- (pp. 22-24). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.
- Canivez, G. L. (2004). Replication of the Adjustment Scales for Children and Adolescents core syndrome factor structure. *Psychology in the Schools, 41*, 191-199.
- Canivez, G. L., & Bordenkircher, S. E. (2002). Convergent and divergent validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales. *Journal of Psychoeducational Assessment, 20*, 30-45.
- Canivez, G. L., Perry, A. R., & Weller, E. M. (2001). Stability of the Adjustment Scales for Children and Adolescents. *Psychology in the Schools, 38*, 217-227.
- Canivez, G. L., & Rains, J. D. (2002). Construct validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales: Convergent and divergent evidence. *Psychology in the Schools, 39*, 621-633.
- Canivez, G. L., & Watkins, M. W. (2002). Interrater agreement for syndromic profile classifications on the Adjustment Scales for Children and Adolescents. *Assessment for Effective Intervention, 28*, 39-46.
- Canivez, G. L., Watkins, M. W., & Schaefer, B. A. (2002). Interrater agreement for syndromic profile interpretations on the Adjustment Scales for Children and Adolescents. *Psychology in the Schools, 39*, 375-389.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

- Champion, C. E., Green, S. B., & Sauser, W. (1988). Development and evaluation of short-cut behaviorally anchored rating scales. *Educational and Psychological Measurement, 48*, 29-41.
- Chen, C., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparison of rating scales among East Asian and North American students. *Psychological Science, 6*, 170-175.
- Clark, L. A., & Watson, D. (1995). Construct validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.
- Cohen, J. (1982). *Statistics for social scientists*. London: Harper and Row.
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56*, 754-761.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conners, K. C. (1997). *Conners' Rating Scales (Revised)*. North Towanda, NY: Multi-Health Systems.
- Conners, K. C., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The Revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26*, 257-268.
- Connor, D. F., Steingard, R. J., Cunningham, J. A., Anderson, J. J., & Melloni, R. H. (2004). Proactive and reactive aggression in referred children and adolescents. *American Journal of Orthopsychiatry, 74*, 129-136.

- Conrad, M., & Hammen, C. (1989). Role of maternal depression in perceptions of child maladjustment. *Journal of Consulting and Clinical Psychology, 57*, 663-667.
- Costenbader, V. K., & Keller, H. R. (1990). Behavioral ratings of emotionally handicapped, learning disabled, and nonreferred children: Scale and source consistency. *Journal of Psychoeducational Assessment, 8*, 485-496.
- Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24*, 315-318.
- Council for Children with Behavior Disorders. (1987). Position paper on definition and identification of students with behavioral disorders. *Behavioral Disorders, 12*, 9-19.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Cudeck, R. (2000). Exploratory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 265-296). San Diego: Academic Press.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.
- Day, D. M., Bream, L. A., & Pal, A. (1992). Proactive and reactive aggression: An analysis of subtypes based on teacher perceptions. *Journal of Clinical Child Psychology, 21*, 210-217.

- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
- Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146-1158.
- Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice, 7*, 435-453.
- DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Murphy, K., & Barkley, R. A. (1994). *ADHD Rating Scale-IV*. Unpublished.
- DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., McGoey, K.E., & Ikeda, M. J. (1998). Parent ratings of attention deficit hyperactivity disorder symptoms: Factor structure, normative data, and psychometric properties. *Journal of Psychopathology and Behavioral Assessment, 20*, 83-102.
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Reid, R., McGoey, K. E., & Ikeda, M. J. (1997). Teacher ratings of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. *Psychological Assessment, 9*, 436-444.
- DuPaul, G. J., & Barkley, R. A. (1992). Situational variability of attention problems: Psychometric properties of the revised Home and School Situations Questionnaires. *Journal of Clinical Child Psychology, 21*, 178-188.
- Edelbrock, C. (1983). Problems and issues in using rating scales to assess child personality and psychopathology. *School Psychology Review, 12*, 293-299.

- Elliot, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and interpretation. *School Psychology Review, 22*, 313-321.
- Epkins, C. C. (1995). 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-report. *Journal of Clinical Child Psychology, 24*, 63-70.
- Erford, B. T. (1993). *Disruptive Behavior Rating Scale*. East Aurora, NY: Slosson.
- Erford, B. T. (1996). Reliability and validity of mother responses to the Disruptive Behavior Rating Scale – Parent Version (DBRS-P). *Diagnostique, 21*, 17-33.
- Erford, B. T. (1998). Technical analysis of father responses to the Disruptive Behavior Rating Scale – Parent Version (DBRS-P). *Measurement and Evaluation in Counseling and Development, 30*, 199-210.
- Fabrigar, L. R., MacCallum, R. C., Wegener, D. T., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.
- Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*, 301-322.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.

- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291-314.
- Gerber, M., & Semmel, M. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*, 137-146.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393.
- Goh, D. S., Teslow, C. J., & Fuller, G. B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology, 12*, 696-706.
- Gorsuch, R. L. (1974). *Factor analysis*. Philadelphia: Saunders.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68*, 532-560.
- Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. Velicer (Eds.), *Handbook of psychology: Vol. 2 research methods in psychology* (pp. 143-164). Hoboken, NJ: Wiley.
- Gresham, F. M. (1985). Behavior disorder assessment: Conceptual, definitional, and practical considerations. *School Psychology Review, 14*, 495-509.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*, 265-275.

- Harman, H. H. (1976). *Modern factor analysis* (2nd ed.). Chicago: University of Chicago Press.
- Hartel, C. E. J. (1993). Rating format research revisited: Format effectiveness and acceptability depend on rater characteristics. *Journal of Applied Psychology, 78*, 212-217.
- Hinshaw, S. P. (1987). On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Bulletin, 101*, 443-463.
- Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. *Journal of Clinical Child Psychology, 21*, 143-150.
- Hofstee, W. K. B. (2002). The questionnaire construction of personality: Pragmatics of personality assessment. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological measurement* (pp. 19-35). Mahwah, NJ: Erlbaum.
- Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Boston: Allyn & Bacon.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

- Horn, W. F., Wagner, A. E., & Ialongo, N. (1989). Sex differences in school-aged children with pervasive attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology, 17*, 109-125.
- Hoyt, W. T., & Kerns, M. D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*, 403-424.
- Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review, 21*, 271-284.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36.
- Kamphaus, R. W., & Frick, P. J. (1996). *Clinical assessment of child and adolescent personality and behavior*. Boston: Allyn & Bacon.
- Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice, 31*, 155-164.
- Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis, 10*, 141-150.
- Kazdin, A. E. (1981). Assessment techniques for childhood depression. *Journal of the American Academy of Psychiatry, 20*, 358-375.
- Kennedy, M. L., Faust, D., Willis, W. G., & Piotrowski, C. (1994). Social-emotional assessment practices in school psychology. *Journal of Psychoeducational Assessment, 12*, 228-240.

- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen.
- Kline, P. (1993). *The handbook of psychological testing*. New York: Routledge .
- Kline, P. (1998). *The new psychometrics: Science, psychology, and measurement*. New York: Routledge .
- Knoff, H. M. (2002). Best practices in personality assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1281-1302). Washington D.C.: National Association of School Psychologists.
- Kovaleski, J. F., Tucker, J. A., & Stevens, A. J. (1996). Bridging special and regular education: The Pennsylvania initiative. *Educational Leadership*, 53, 44-47.
- Kresanov, K., Tuominen, J., Piha, J., & Almqvist, F. (1998). Validity of child psychiatric screening methods. *European Child and Adolescent Psychiatry*, 7, 85-95.
- Lachar, D., & Gruber, C. P. (2003). Multisource and multidimensional objective assessment of adjustment: The Personality Inventory for Children Second Edition, Personality Inventory for Youth, and Student Behavior Survey. In C.R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 337-367). New York: Guildford Press.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694.
- Martens, B. K. (1993). Social labeling, precision of measurement, and problem-solving: Key issues in the assessment of children's emotional problems. *School Psychology Review*, 22, 308-312.

- Mash, E. J., & Terdal, L. E. (1988). Behavioral assessment of child and family disturbance. In E. J. Mash & L. G. Terdal (Eds.), *Behavioral assessment of childhood disorders* (2nd ed., pp. 3-65). New York: Guilford Press.
- McConaughy, S. H. (1992). Objective assessment of children's behavioral and emotional problems. In C. E. Walker & M. C. Roberts (Eds.), *Handbook of clinical child psychology* (pp. 163-180). Oxford, England: Wiley.
- McConaughy, S. H. (1993). Advances in empirically based assessment of children's behavioral and emotional problems. *School Psychology Review*, 22, 285-307.
- McConaughy, S. H., & Ritter, D. R. (2002). Best practices in multidimensional assessment of emotional and behavioral disorders. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1303-1320). Washington D.C.: National Association of School Psychologists.
- McDermott, P. A. (1981). The manifestation of problem behaviour in ten age groups of Canadian school children. *Canadian Journal of Behavior Science*, 13, 310-319.
- McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment*, 5, 413-424.
- McDermott, P. A. (1994). *National profiles in youth psychopathology: Manual of Adjustment Scales for Children and Adolescents*. Philadelphia: Edumatic and Clinical Science.
- McDermott, P. A. (1995). Sex, race, class, and other demographics as explanations for children's ability and adjustment: A national appraisal. *Journal of School Psychology*, 33, 75-91,

- McDermott, P. A., & Schaefer, B. A. (1996). A demographic survey of rare and common problem behaviors among American students. *Journal of Clinical Child Psychology, 25*, 352-362.
- McDermott, P. A., & Spencer, M. B. (1997). Racial and social class prevalence of psychopathology among school-age youth in the United States. *Youth & Society, 28*, 387-414.
- McDermott, P. A., Watkins, M. W., Sichel, A. F., Weber, E. M., Keenan, J. T., Holland, A. M., & Leigh, N. M. (1995). The accuracy of new national scales for detecting emotional disturbance in children and adolescents. *The Journal of Special Education, 29*, 337-354.
- Meredith, W. (1965). A method for studying differences between groups. *Psychometrika, 30*, 15-29.
- Merrell, K. W. (1994). *Preschool and Kindergarten Behavior Scales*. Brandon, VT: Clinical Psychology Publishing.
- Merrell, K. W. (1999). *Behavioral, social, and emotional assessment of children and adolescents*. Mahwah, NJ: Erlbaum.
- Merrell, K. W. (2000). Informant report: Rating scale measures. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Conducting school-based assessments of child and adolescent behavior* (pp. 203-234). New York: Guilford Press.
- Merrell, K. W. (2001). Assessment of children's social skills: Recent developments, best practices, and new directions. *Exceptionality, 9*, 3-18.
- Michael, K. D., & Merrell, K. W. (1998). Reliability of children's self-reported

- internalizing symptoms over short to medium-length time intervals. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 194-201.
- Miller, L. C. (1967). Louisville Behavior Checklist for males 6-12 years of age. *Psychological Reports*, 21, 885-896.
- Myers, K., & Winters, N. C. (2002). Ten-year review of rating scales: Overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 114-122.
- Naglieri, J. A., LeBuffe, P. A., & Pfeiffer, S. I. (1994). *Devereux Scales of Mental Disorders (DSMD)*. San Antonio, TX: Psychological Corporation.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.
- O'Connor, B. P. (2001). Extension: SAS, SPSS, and MATLAB programs for extension analysis. *Applied Psychological Measurement*, 25, 88.
- Peterson, D. R. (1961). Behavior problems of middle childhood. *Journal of Consulting Psychology*, 25, 205-209.
- Phares, V., Comas, B., & Howell, D. (1989). Perspectives on child behavior problems: Comparisons of children's self-reports with parent and teacher reports. *Psychological Assessment*, 1, 68-71.
- Piacentini, J. C., Cohen, P., & Cohen, J. (1992). Combining discrepant diagnostic information from multiple sources: Are complex algorithms better than simple ones? *Journal of Abnormal Child Psychology*, 20, 51-63.

- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method bias in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- Poulin, F., & Boivin, M. (2000). Reactive and proactive aggression: Evidence of a two-factor model. *Psychological Assessment, 12*, 115-122.
- Price, J. M., & Dodge, K. A. (1989). Reactive and proactive aggression in childhood: Relations to peer status and social context dimensions. *Journal of Abnormal Child Psychology, 17*, 455-471.
- Prout, H. T. (1983). School psychologists and social-emotional assessment techniques: Patterns in training and use. *School Psychology Review, 12*, 377-383.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287-297.
- Reschly, D. J., Genshaft, J., & Binder, M. S. (1987). *The 1986 NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues*. Washington D.C.: National Association of School Psychologists. (ERIC Document Reproduction Service No. ED30073).
- Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children (BASC)*. Circle Pines, MN: American Guidance Services.
- Reynolds, W. M. (2004). The Reynolds Adolescent Depression Scale – Second Edition (RADS-2). In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment: Vol. 2 personality assessment* (pp. 224-236). Hoboken, NJ: Wiley.

- Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*, *112*, 485-499.
- Richters, J. E., & Pelligrini, D. (1989). Depressed mothers' judgment about their children: An examination of the depression-distortion hypothesis. *Child Development*, *60*, 1068-1075.
- Rowe, D. C., & Kandel, D. (1997). In the eye of the beholder? Parent ratings of externalizing and internalizing symptoms. *Journal of Abnormal Child Psychology*, *25*, 265-275.
- Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.
- Schneider, B. H., & Byrne, B. M. (1989). Parents rating children's social behavior: How focused the lens? *Journal of Clinical Child Psychology*, *18*, 237-241.
- Schowengerdt, R. V. (2001). Review of the Adjustment Scales for Children and Adolescents. In J. Impara & B. Plake (Eds.), *The fourteenth mental measurements yearbook* (pp. 24-26). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.
- Schwartz, S. A. (1978). A comprehensive system for item analysis in psychological test construction. *Journal of Educational Measurement*, *15*, 117-123.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.
- Schwarz, N., & Scheuring, B. (1992). Selbstberichtete verhaltens- und symptomhaeufigkeiten: Was befragte aus Antwortvorgaben des fragebogens lernen

- (Frequency reports of behaviors and symptoms: What respondents learn from the response alternatives of the questionnaire). *Zeitschrift fuer Klinische Psychologie Forschung und Praxis*, 21, 197-208.
- Segal, D. L., & Collidge, F. L. (2004). Objective assessment of psychology and psychopathology: An overview. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment: Vol. 2. personality assessment* (pp. 3-13). Hoboken, NJ: Wiley.
- Semrud-Clikeman, M., Bennett, L., & Guli, L. (1993). Assessment of childhood depression. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (2nd ed., pp. 259-290). New York: Guildford Press.
- Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools*, 41, 551-561.
- Silverman, W. K., & Rabian, B. (1999). Rating scales for anxiety and mood disorders. In D. Shaffer & C. P. Lucas (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 127-166). New York: Guildford Press.
- Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, 7, 300-308.
- Snook, S. C., & Gorsuch, R. L. (1989). Common factor analysis vs. component analysis. *Psychological Bulletin*, 106, 148-154.

- SPSS. (2001). *SPSS 10.0 for Macintosh*. Chicago: Author.
- Stinnett, M. L., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12*, 331-350.
- Stott, D. H. (1966). *The social adjustment of children*. London: University of London Press.
- Stott, D. H., Marston, N. C., & Neill, S. J. (1975). *Taxonomy of behavior disturbance*. Toronto: Musson.
- Stott, D. H., McDermott, P. A., Green, L. F., & Francis, J. M., (1988). *Learning Behaviors Scale and Study of Learning Behaviors research edition manual*. San Antonio, TX: Psychological Corporation.
- Streiner, D. L. (1994). The use and misuse of factor analysis. *Canadian Journal of Psychiatry, 39*, 135-140.
- Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes with special reference to negative effects in psychotherapy. *American Psychologist, 32*, 187-196.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.

- Tripp, G. & Luk, S. L. (1997). The identification of pervasive hyperactivity: Is clinic observation necessary? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38, 219-234.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.
- Van der Valk, J. C., Van den Oord, E. J., Verhulst, F. C., & Boomsma, D. I. (2001). Using parent ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *Journal of Child Psychology and Psychiatry*, 42, 921-931.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). Boston: Kluwer.
- Verhulst, F. C., & Akkerhuis, G. W. (1989). Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4-12. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 30, 123-136.
- Verhulst, F. C., Dekker, M. C., & Van der Ende, J. (1997). Parent, teacher, and self-reports as predictors of signs of disturbance in adolescents: Whose information carries the most weight? *Acta Psychiatrica Scandinavica*, 96, 75-81.

- Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. *Journal of Abnormal Child Psychology*, 22, 531-546.
- Vitaro, F., Brendgen, M., & Tremblay, R. E. (2002). Reactively and proactively aggressive children: Antecedent and subsequent characteristics. *Journal of Child Psychology and Psychiatry*, 43, 495-505.
- Volpe, R. J., & DuPaul, G. J. (2001). Assessment with brief behavior ratings scales. In J. Andrews, D. Saklofske, & H. Janzen (Eds.), *Handbook of psychoeducational assessment: Ability, achievement, and behavior in children* (pp. 357-387). San Diego, CA: Academic Press.
- Watkins, M. W., & Canivez, G. L. (1997). Interrater agreement of the Adjustment Scales for Children and Adolescents. *Diagnostique*, 22, 205-213.
- Watkins, M. W., & McDermott, P. A. (2002). *The Adjustment Scales for Children and Adolescents – Home*. Unpublished.
- Webster-Stratton, C. & Hammond, M. (1988). Maternal depression and its relation to life stress, perceptions of child behavior problems, parenting behaviors, and child conduct problems. *Journal of Abnormal Child Psychology*, 16, 299-315.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. New York: Psychological Corporation.
- Widaman, K. F. (1993). Common factor analysis versus principal components analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28, 263-311.

- Wilson, D. R., & Prentice-Dunn, S. (1981). Rating scales in the assessment of child behavior. *Journal of Clinical Child Psychology, 10*, 121-126.
- Wilson, M. S., & Reschly D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25*, 62-80.
- Wingenfeld, S. A. (2002). Assessment of behavioral and emotional difficulties in children and adolescents. *Peabody Journal of Education, 77*, 85-107.
- Wood, S. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*, 354-365.
- Wright, D., & Piersel, W. C. (1992). Components of variance in behavior ratings from parents and teachers. *Journal of Psychoeducational Assessment, 10*, 310-318.
- Wrigley, C. S., & Neuhaus, J. O. (1955). The matching of two sets of factors. *American Psychologist, 1*, 354-365.
- Youngstrom, E. A., Findling, R. L., Danielson, C. K., & Calabrese, J. R. (2001). Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychological Assessment, 13*, 267-276.
- Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*, 253-269.
- Zwick, W. R., & Velicer, W. F. (1986). A comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

VITA

Debra Young Coffey

25 Sawgrass St.
Jackson, NJ 08527

Education:

Ph.D.	School Psychology	Penn State University	May 2006 (anticipated)
Ed.S.	School Psychology	College of William and Mary	May 1995
M.Ed.	School Psychology	College of William and Mary	May 1994
B.S.	Communications and Psychology	Bradley University	May 1992

Professional Experience:

07/00 – 06/05	School Psychologist	Baltimore County Public Schools, Towson, MD
01/01 – 05/00	Teaching Assistant	College of Health and Human Development Penn State University, University Park, PA
07/95 – 07/98	School Psychologist	Baltimore County Public Schools, Towson, MD
08/94 – 06/95	School Psychology Intern	Williamsburg-James City County Public Schools, Williamsburg, VA

Publications and Presentations:

Ward, S.B., Ward, T.W., Drewer, H., Hatt, C.V., Ruby, S., Mollner, N.R., & Young, D.L. (1994). *The WISC-III: Criterion-related validity and utility with a referred population*. Involved with a symposium presented at the 1994 NASP Convention.

Ward, S.B., Ward, T.W., Hatt, C.V., Young, D.L., Mollner, N.R. (1995). The incidence and utility of the ACID, ACIDS, and SCAD profiles in a referred population. *Psychology in the Schools*, 32, 267-276.

Watkins, M.W., & Coffey, D.Y. (2005). Reading motivation: Multidimensional and indeterminate. *Journal of Educational Psychology*, 96, 110-118.

Professional Affiliations:

National Association of School Psychologists
Maryland School Psychologists Association
Baltimore County School Psychologists Association

Certification:

Nationally Certified School Psychologist (NCSP)
Maryland Certified School Psychologist II