

A Thesis
entitled
The Child and Adolescent Functional Assessment Scale (CAFAS):
A Reliability and Validity Evaluation
by
Megan N. Bodine
Submitted to the Graduate Faculty as partial fulfillment of the
requirements for the Master of Arts in Psychology

Wesley A. Bullock, Ph.D., Committee Chair

Jeanne Brockmyer, Ph.D.,
Committee Member

Kamala London, Ph.D., Committee Member

Patricia Komuniecki, Ph.D., Dean
College of Graduate Studies

The University of Toledo

August 2011

Copyright 2011, Megan N. Bodine. This document is copyrighted material. Under copyright law, no parts of this document may be reproduced without the expressed permission of the author

An Abstract of
The Child and Adolescent Functional Assessment Scale (CAFAS):
A Reliability and Validity Evaluation

by

Megan N. Bodine

Submitted to the Graduate Faculty as partial fulfillment of the
requirements for the Master of Arts in Psychology

The University of Toledo
August 2011

The Child and Adolescent Functional Assessment Scale (CAFAS) is used to measure functioning of children and adolescents in daily life activities. It was designed to assess youth with emotional, behavioral, psychological, or substance use problems. The CAFAS is currently being used in 30 states, typically to shape a child's treatment plan, based on the deficiencies the youth is experiencing. Despite its widespread use, the psychometric properties of the CAFAS are currently unclear in the literature, as is the degree to which the CAFAS is useful with different child and adolescent populations. The current study evaluated the psychometric properties of the CAFAS in a sample of 1121 juvenile justice youth, including the internal reliability of the CAFAS subscale scores to CAFAS Total score, as well as the criterion and convergent validity. These youth were enrolled in a juvenile justice treatment program at a Community Mental Health Center in Southeast Michigan. Enrollment in this particular program is based on the youth's home address, and they must have committed some crime, ranging from status offense to felony offense. Results of the current study indicated that the internal reliability of the CAFAS, assessed by computing coefficient alpha values for the CAFAS

subscale scores to CAFAS Total, was fair to moderate. Convergent validity was examined by comparing CAFAS subscales to comparable scales from the Millon Adolescent Clinical Inventory (MACI). These results indicated limited convergent validity between the CAFAS and the MACI scales. Criterion-related validity was examined by comparing the number and severity of offenses committed prior to intake to the CAFAS scores at intake. Results indicated the current study demonstrated significant gender differences between several of the CAFAS subscales at intake, as well as significant differences between some CAFAS subscales, both of which indicate that the individual CAFAS subscales should be utilized separately, rather than just using the CAFAS Total score as an overall level of functioning. Although the CAFAS Total score is typically used as an overall assessment of level of functioning, the results of the current study suggest that the individual CAFAS scores should be examined separately, with special attention given to gender differences.

Table of Contents

Abstract	iii
Table of Contents	v
List of Tables	vii
I. Introduction	1
II. Literature Review	4
A. History of the Child and Adolescent Functional Assessment Scale (CAFAS)	4
a. North Carolina Functional Assessment Scale (NCFAS)	4
b. Child Assessment Schedule (CAS)	5
c. The Child and Adolescent Functional Assessment Scale (CAFAS)	7
B. Psychometric Properties of the CAFAS	9
a. Reliability	10
b. Validity	11
C. Statement of the Problem	15
D. Purpose of the Present Study	16
E. Research Questions	17
F. Additional Research Questions	17
III. Method	18
A. Participants	18
B. Sample Descriptives	19
C. Design and Procedure	23
D. Measures	24

a. Child and Adolescent Functional Assessment Scale (CAFAS)	24
b. Millon Adolescent Clinical Inventory (MACI)	24
IV. Analyses and Results	27
V. Discussion	42
VI. References	54
VII. Appendices	64
A. Appendix A: Brief Descriptions of Each CAFAS subscale	64
B. Appendix B: MACI Scale Description	65

List of Tables

Table 1.	Participant Age.....	19
Table 2.	Participant Ethnicity.	20
Table 3.	Family Income.....	20
Table 4.	Family Insurance Source	21
Table 5.	Primary DSM-IV TR Diagnosis	22
Table 6.	Participant Full Scale IQ	23
Table 7.	Means and Distributions of CAFAS scores at Intake.....	27
Table 8.	Initial CAFAS Subscale Scores by Gender.....	29
Table 9.	Mean Total CAFAS score for Comparative Youth Samples at Intake and Discharge (when available).....	31
Table 10.	Internal Consistency(Cronbach's Alpha) for CAFAS Total and Subscale Scores.....	33
Table 11.	CAFAS Subscale Correlation.....	34
Table 12.	Severity of Offenses at Intake.....	35
Table 13.	MACI vs. CAFAS Subscale Correlation.....	36
Table 14.	DSM-IV TR Primary Diagnoses, by Group.....	38
Table 15.	DSM-IV TR Primary Diagnosis for females.....	40
Table 16.	DSM-IV TR Primary Diagnoses for males.....	41

Chapter One

Introduction

Knowledge about juvenile delinquency, and the interventions and treatments that come from this knowledge, encompasses a very broad range of competing biological, social, cultural, and psychological theories (Ohlin, 1998). In rural areas, 30-40% of all boys will be arrested before their eighteenth birthday (Zagar, Busch & Hughes, 2009). The treatments already implemented lead to only a 10-40% effectiveness (Zagar, 2009), which means that at least half of those who are arrested and complete treatment still continue to reoffend post treatment. Engaging in delinquent behaviors as an adolescent has been found to later predict adult criminal activity (Greenwood, 2008). In fact, in the 2006 National Report, the United States Department of Justice reported that one-fourth of juvenile delinquents who offended at the ages of 16-17 continued on to offend later in life. These statistics are compelling and prompt many researchers to study the population in attempts to find some solution to curb juvenile offending and recidivism after juvenile justice programs have been completed.

There are hundreds of studies testing theories dealing with prevention of delinquency, predicting recidivism, and evaluating treatment programs (Zagar et al., 2009), but the one conclusion that is clear is that there is no one successful intervention for delinquency. Part of the problem is that offenses can range from basic status offenses, like truancy, to felonies, like murder. Another problematic issue is that juvenile offenders have many different social, educational, and health problems (including mental health problems), and so may need additional services along with delinquency services (McReynolds, et al., 2008). Juvenile offenders are not a homogeneous set of individuals

when broken down by gender, either. Some research (Graves, Frabutt, & Shelton, 2007) has found evidence of more mental health issues in female offenders, since girls are more prone to internalizing mental health problems (i.e. depression or anxiety). These problems may manifest themselves in girls by causing them to react in ways other than typical delinquent behaviors.

Overall, youth involved in juvenile justice systems have more reported mental health disorders than youth in the general population (Jonson-Reid, Williams, & Webster, 2001). In 1991, a study was conducted to look at the prevalence of emotional disorders in a juvenile justice population in Ohio (Davis, Bean, Schumacher & Stringer, 1991). In this population, nearly 20 percent of the 173 participants had an experience with inpatient mental health treatment. In the overall United States population, however, it is estimated that between 40-90% of juvenile offenders also suffer from mental illness (Graves, Frabutt, & Shelton, 2007), and little research has been conducted focusing on this link between mental illness and juvenile delinquency (Rosenblatt, Rosenblatt, & Biggs, 2000). Although this is a potentially huge number of juvenile offenders suffering from mental illness, little else is known about the dual involvement in the juvenile justice and mental health systems. This highlights the importance of youth needing appropriate mental health assessment and treatment in addition to traditional juvenile delinquency treatment programs.

Because of the various theories behind, and treatments for, juvenile delinquency, there are numerous measures available for use with this population. Each measure is designed to assess slightly different concepts, from measuring daily functioning to predicting recidivism. One such measure is the Child and Adolescent Functional

Assessment Scale (CAFAS). The CAFAS is a widely used clinician rating scale that examines the functioning of children and adolescents in daily life activities. Although the CAFAS was designed to assess youth with emotional, behavioral, psychological, and substance use problems, it has not been psychometrically studied for usage within a juvenile justice population. Many juvenile offenders suffer from emotional problems, but since basic psychometric properties of the CAFAS are missing from the current literature, it is unclear if the CAFAS is appropriate for youth involved in a juvenile justice population. The current study will evaluate the psychometric properties of the CAFAS with a new juvenile justice population and will examine the utility of the measure for this population.

Chapter Two

Literature Review

History of the Child and Adolescent Functional Assessment Scale (CAFAS)

One popular functional impairment scale that is currently being used in over 30 states is the Child and Adolescent Functional Assessment Scale (CAFAS) (Bates, 2001). The CAFAS is a multidimensional psychometric assessment instrument that provides scores across a group of descriptive categories used to measure daily functioning of children and adolescents, ages six to 17 (Hodges, Wong & Latessa, 1998). It was designed to assess special populations of children and adolescents, such as those with emotional, behavioral, psychological, or substance use problems and focuses on mental health deficiencies. The CAFAS can be used with any youth who is currently experiencing problems or is at risk for experiencing problems in these areas. The CAFAS was developed using two other measures as models: the North Carolina Functional Assessment Scale and the Child Assessment Schedule (Hodges et al., 1998).

The North Carolina Functional Assessment Scale (NCFAS)

The North Carolina Functional Assessment Scale (NCFAS) is a multidimensional functional impairment scale for adults (Bates, 2001). The measure is completed by the clinician every three months (Walker, Minor-Schork, Bloch & Esinhart, 1996) and has six dimensions: role performance, emotional health, ability to care for basic needs, behavior, thinking, and substance use. The subscale scores are combined to form a total score, and that score is used to determine level of functioning. An individual's total score could range from 0 to 180, and a score above 40 indicates serious dysfunction. The only published study using the NCFAS (Walker et al., 1996) does not include any reliability or

validity information, and there are no other published studies that provide more information.

The Child Assessment Schedule (CAS)

The Child Assessment Schedule (CAS) (Hodges, Kline, Stern, Cytryn & McKnew, 1982) was developed as a diagnostic interview to be used with children to generate diagnoses and other relevant clinical information. The interview begins with a structured section of approximately 75 questions about a child's life. These questions include items about school, family, fears, hobbies and activities, etc. The second part of the interview is clinician rated behavioral observations of the child. There are 53 items that address areas including grooming, cognitive ability, insight, and emotional expression. A total score is found based on the responses to the structured questions, as well as the clinician's observations, which indicate the presence of problems or symptoms. Each area of interest in the structured interview also could be scored (i.e., the school or family section), and the number of items endorsing a problem in that area is the score. Questions dispersed throughout the structured interview could also lead to a DSM diagnosis, and a score for this diagnosis is determined by the number of items endorsed for a particular diagnosis.

Some of the psychometric properties of the CAS were described by Hodges et al. (1982). The CAS interviews were videotaped in order to assess interrater agreement. The raters consisted of the two CAS interviewers and two first year graduate students, all of whom had training experience with the CAS. Interrater reliability was then calculated by comparing whether the raters agreed on the presence or absence of a symptom (Hodges et al., 1982). The majority of the symptom ratings demonstrated a Kappa value

of .60 or greater, which indicates moderate agreement between the raters (Landis & Koch, 1977). In order to test the reliability of the scale, correlation coefficients were calculated for each pair of the raters and their total CAS scores, as well as the subscale scores and DSM diagnosis scores (Hodges et al., 1982). The CAS total score demonstrated a mean correlation of .90, which indicates a high reliability for overall pathology (Myers & Winter, 2002). The other CAS score correlations ranged from .56 to .84, with the majority reaching moderate reliability.

Concurrent validity of the CAS was examined by using comparisons between contrast groups of inpatient, outpatient, and control groups. There were 87 total youth, and their ages ranged from 7 to 14 years. The mean ages for each group were 10.3, 10.2, and 10.0 years, respectively. The inpatient group consisted of 18 youth, the outpatient group consisted of 32 youth, and the control group consisted of 37 youth. Comparisons between the groups were made for the CAS total score, as well as agreement between maternal report (mothers completing the Child Behavior Checklist (CBCL) (Achenbach, 1978; Achenbach & Edelbrock, 1978) and CAS scores, and also agreement between child report (children completed the Child Depression Inventory (CDI) (Kovacs, 2003) and the State-Trait Anxiety Inventory for Children (STAIC) (Spielberger, 1973) and CAS scores. There were significant differences found between the three groups for the total CAS score, as well as for 9 out of 11 of the subscale scores, and for 8 out of 9 of the DSM diagnosis categories. In comparisons between the CAS and the CBCL, significant differences were again found between the groups that closely matched the differences found using the CAS. Using the child reported CDI scores, significant differences were found between the comparison groups for depression, such that the control group scored

lower than both the outpatient and the inpatient groups. The same differences were also found between the comparison groups for anxiety using the STAIC scores, such that the control group scored lower than both the outpatient and inpatient groups.

The Child and Adolescent Functional Assessment Scale (CAFAS)

As noted above, the CAFAS was developed as a combination of the NFCAS, a functional impairment rating scale, and the CAS, a semi-structured diagnostic interview. In the original version of the CAFAS, Hodges (1989) does not provide specific information regarding how the specific items for the CAFAS were chosen, although Bates (2001) has noted that the CAFAS is not reportedly based on any particular theory of child psychopathology. Bates (2006) researched the individual CAFAS items and found that the CAFAS may not have a clear structure and may need some additional editing. Hodges (1989) modeled the rating scale format after that of the NFCAS. The original version of the CAFAS contained 164 items designed to measure impairment in five domains: role performance, behaviors toward others/self, moods and emotions, self-harmful behavior, substance use, and thinking processes. These domains of impairment are cumulative and give an overall sense of how the youth is functioning on a day-to-day basis. The total CAFAS score is then calculated and used to determine a youth's overall functioning.

In addition to the youth rating scales, there are two additional scales used to rate the youth's caregivers: the Material Needs and the Family/Social Support scales. Both are designed to rate the extent that the youth's impairment is due to lack of resources in each category.

The second version of the CAFAS (Hodges, 2000) contained 315 items and eight core subscales, as well as the two parent/caregiver subscales. The original subscale of role performance (Hodges, 1989) was separated into three new subscales: school/work, home, and community.

For each of the eight impairment scales, as well as the two caregiver scales, a rater (such as a therapist) determines the level of impairment that seems most closely related to the youth's daily functioning at the time of assessment (Hodges, 2003). The four severity levels are as follows: Minimal or No Impairment (i.e., no disruption of functioning); Mild (i.e., significant problems or distress); Moderate (i.e., persistent disruption or major occasional disruption); and Severe (i.e., severe disruption or incapacitation) (Hodges, 2003). For each scale and each impairment level, there is a set of behaviors the rater must review. The rater must begin in the most severe list of behaviors, and if any behavior is found that matches the youth's functioning, the "severe" level is given. If no behavior matches in the severe level, the rater continues down through the levels until a match is found. The ratings for behaviors are based on what has been reported by the youth or parent, or what the rater has directly observed or read in the youth's case notes. It is important to note that any source of information a rater has can be used to determine the youth's functioning.

Each level of impairment has a corresponding numerical value: Severe, 30; Moderate, 20; Mild, 10; Minimal or No Impairment, 0 (Hodges, 2003). The values for each of the eight functioning scales are summed, giving a total impairment score, ranging from zero to 240, while the caregiver scale scores are examined separately. The higher the score, the worse the youth's functioning is, and the greater the impairment.

There is a standardized training protocol for every practitioner using the CAFAS (Hodges, 2003). Training begins with learning about the CAFAS and familiarizing each practitioner with the behaviors listed and the rules for scoring. After that, a set of six training vignettes are used to give practitioners case examples to practice rating. They are then given the answers to compare to what they chose and individuals are given the opportunity to discuss any discrepancies between their rating and the actual correct score. After these training vignettes, there are 10 more vignettes that are to be completed individually and scored by the trainer.

Psychometric Properties of the CAFAS

Although the CAFAS is currently being used in over 30 states around the United States, Bates (2001) has asserted that the psychometric properties of the CAFAS are not clear. Hodges (2003) reported that validity and reliability information has been established for the CAFAS, “using large data sets generated by two evaluation studies” (Hodges, 2004). The first was the Fort Bragg Evaluation Project (Hodges, 2003; Breda, 1996). The youth recruited for the study were Army dependents, and mostly middle class, from families with generous mental health benefits. The youth were referred for mental health problems. The age ranged from five years to 17 years, with a mean age of 11 at intake. The second study was the Center for Mental Health Service (CMHS)-funded Evaluation Project. The youth from this study had serious mental health problems and were generally from impoverished families, involved in multiple services from multiple agencies. The age ranged from four years to 23 years.

Basic descriptive analyses need to be examined for the CAFAS. Hodges and Wong (1996) stated that no significant differences have been found for the CAFAS based

on gender or race/ethnicity, which is a controversy noted by Bates (2001) in his review of the CAFAS. In addition, this asserted lack of differences as a function of gender or ethnicity is not consistent with other research looking at mental health dimensions and diagnoses across similar demographic variables, such as the Behavior Assessment System for Children (Reynolds & Kamphaus, 2004), and the Child Behavior Checklist (Achenbach, 1991).

Reliability: Internal Consistency and Inter-rater Reliability

One of the most widely used methods of reliability is coefficient alpha. Coefficient alpha values were used to demonstrate internal consistency within the different scales of the CAFAS for both of the evaluation studies, cited by Hodges (2004) as the basis for the reliability of the CAFAS. The Cronbach alpha at intake was .63 (Hodges, 2003) for the Fort Bragg Evaluation Project, and .73 for the CMHS-funded Evaluation Project. These values suggest that the measure has a low to moderate level of internal consistency, which means the extent to which the CAFAS subscales assess the same characteristic is only low to moderate. Hodges (2003) reported that although these are lower alpha values, each scale evaluated is designed to assess different constructs, and so should be more heterogeneous; however, Bates (2001) argues that these low values do not provide support for her claims of good internal consistency.

In contrast to internal reliability, Bates (2001) reported that interrater reliability for the CAFAS is good. Hodges and Wong (1996) conducted Pearson Product Moment correlations between raters' scores and the criterion scores. For the CAFAS total scores, the correlations ranged from .92 - .96, while the individual subscale correlations ranged from .73 - .99. In order to examine how the raters agreed with each other, intraclass

correlations were calculated and ranged from .63 - .96. While the total score correlations were very high, once the total score was broken down into subscales, the values demonstrated only a fair interrater reliability according to reliability values Cicchetti (1994) published as guidelines. His guidelines state that a correlation below .70 is unacceptable, while correlations from .70-.79 are fair; a correlation from .80-.89 is good, and if the measure of consistency reaches .90 and above it is considered excellent. Reliability for raters' agreement was not sufficient, as it ranged from only fair (.73) to high (.99), depending on which group was being compared.

In a later study (Ogles, Davis & Lunnen, 1999), interrater reliability of the CAFAS was re-examined. This study also examined whether level of raters' clinical training and experience had an effect on reliability. There were three rater comparison groups: undergraduates, graduates, and case managers. All groups received the standardized CAFAS training protocol. Each rater received 10 vignettes and 10 actual archival cases. Using the vignettes, raters demonstrated a high level of reliability, with values ranging from .88 - .94. When the actual cases were used, raters demonstrated a poor to fair level of agreement, with values ranging from .55 - .75. There was no significant difference found between comparison groups, based on raters' clinical training and experience.

Validity: Content Validity

Hodges (2003) based the content validity of the CAFAS on information from the original measure she designed, the Child Assessment Scale (CAS) (Hodges & Saunders, 1989), which she reported as having "considerable evidence of reliability and validity to guide the selection of items for the CAFAS." While the CAS had demonstrated adequate

criterion validity, Bates (2001) noted that this is not sufficient information to establish content validity. He further pointed out that it is unknown how items were chosen for the CAFAS, what the factor structure of the CAFAS is, and whether the items chosen for the CAFAS appropriately represent the factors created.

Validity: Structural and Scaling Validity

Until a study in 2006 (Bates, Furlong & Green, 2006), structural and scaling validity information were missing from the CAFAS literature. In the first of their two-part study, Bates et al. had the raters complete a representativeness survey of CAFAS items. They only used the items that ranked at least a 10 in severity, because the items that are below that (items with Minimal or No Impairment rating) do not contribute to the CAFAS score. The raters were asked to put the items into the categories (as given by the subscales on the CAFAS), and also to judge how well that item fit into the category they chose (using a five point Likert scale, with 1 = only a little, 3 = moderately, 5 = very well).

Total agreement among the raters was examined (called agreement), as well as the agreement for how well an item fit into the category (called representativeness). The items that were found to have the greatest level of agreement and representativeness were used for the second part of the experiment. A value of .75 was used for a cutoff for agreement of items, and a value of 4.0 was used as a cutoff for the representativeness of items. Only 77 of the original 132 items were deemed as acceptable for both agreement and representativeness and were kept for the second part of the study.

In the second part of their study, the researchers asked a second group of raters to complete a severity survey, using the 77 items that were found to of acceptable

agreement and representativeness from part one of the study. These raters were asked to judge each item from mild to extreme severity (1 = mild, 4-5 = severe, 8-9 = extreme). A Simplified Successive Interval Scaling analysis was conducted to see if the CAFAS scores were appropriately weighed in ratings of severity. The results of this study found that items with scores of moderate severity significantly overlapped with items from both the mild and severe scores. There also were discrepancies found between the original CAFAS item values and the new values calculated here. Bates et al. (2006) describe the need for more research to further examine the factor structure of the CAFAS and to create empirically grounded items to include on the measure.

Validity: Concurrent Validity

Concurrent validity was also examined to see if the CAFAS was related to other standardized measures that were also designed to measure similar constructs (Hodges & Wong, 1996). The measures used were Child Assessment Schedule, Parent Version (PCAS) (Hodges, Kline, Stern, Cytryn & McKnew, 1982); the Child Behavior Checklist (CBCL) (Achenbach, 1978; Achenbach & Edelbrock, 1979); the Child Assessment Schedule (CAS) (Hodges, Kline, Stern, Cytryn & McKnew, 1982); and the Burden of Care Questionnaire (BCQ) (Bickman et al. 1995). Comparisons of scores were examined at four different time periods, or waves: intake, 6 months post intake, 12 months post intake, and 18 months post intake. Zero-Order correlations were conducted between CAFAS total scores and scores from each of the other measures, and all correlations were found to be in the positive direction and statistically significant for each wave of the study (e.g, .36 to .59 for Wave 1, .42 to .62 for Wave 2, .43 to .58 for Wave 3, and .42 to .63 for Wave 4). This suggests that the CAFAS has high concurrent validity.

Validity: Predictive Validity

Hodges and Wong (1997) also examined the predictive validity of the CAFAS. From the original Fort Bragg evaluation sample, they used the total CAFAS score at intake to predict service utilization at both six and 12 months post intake. They examined the restrictiveness of care, total cost of services, number of bed days (if in residential placement), and number of days of service. The regression analysis conducted demonstrated that worse levels of functioning and higher impairment at intake were significantly related to more service utilization.

Another study (Quist & Matshazi, 2000) examined the relationship between the CAFAS and a juvenile justice population, to see if the CAFAS could predict juvenile recidivism. The researchers used CAFAS total scores and demographic variables (ethnicity, age, sex, number of offenses prior to program participation) as predictors, and examined recidivism as the total number of crimes committed after program completion (felonies, misdemeanors, probation violations). Out of the demographic predictors, only sex was a significant predictor of recidivism, but the CAFAS total score was also a significant predictor, and was an even stronger predictor of recidivism than sex. In a related study, Hodges and Kim (2000) also examined the relationship between the CAFAS and contact with law and poor school attendance. They also found the CAFAS to be a good predictor of contact with the law, as well as with poor school attendance.

Validity: Validity for Use with Target Population

The CAFAS was designed to be used with children and adolescents with emotional and behavioral problems (Hodges, 2003). The original evaluation projects both included youth with these types of problems, but what about populations that are not

exactly what it was designed for? As examined in the previous studies (Quist & Matshazi, 2000), the CAFAS was used as a predictor of contact with the law or recidivism. The CAFAS is being used with the juvenile justice population in many places. In Wayne County, Michigan, for example, it is county mandated that children and adolescents participating in the juvenile justice system be continually assessed using the CAFAS as an indicator of functioning. Although youth involved in juvenile justice may have emotional and/or behavioral problems, it is unclear whether the CAFAS is the correct evaluation tool to assess this unique population's needs, beyond predicting recidivism.

Statement of Problem

Currently it is unclear whether the CAFAS is appropriate for the children and adolescents involved in juvenile justice programs. Not all juvenile offenders have emotional problems, though all do exhibit behavioral problems. The CAFAS has been useful in predicting recidivism in delinquent youth (Quist & Matshazi, 2000), as well as contact with the law (Hodges & Kim, 2000); however, basic psychometric properties of the CAFAS are missing in the general child and adolescent literature, as well as the psychometric properties of the CAFAS with a juvenile justice population, specifically. Gender differences were not examined when conducting the original CAFAS evaluation projects, and while some studies included gender as part of their demographic information collected (Hodges, Doucette-Gates & Kim, 2000; Hodges, Doucette-Gates & Liao, 1999; Quist & Matshazi, 2000; Rosenblatt, Rosenblatt & Biggs, 2000), gender was not examined as an indicator of CAFAS subscale differences. Recent research (Stephens, Petras, Fabian & Walrath, 2009) has found gender differences in overall functional

impairment, as well as differences in service experiences between males and females. In addition, a recent report from the U.S. Department of Justice (Brumbaugh, Walters & Winterfield, 2010) states that assessments used for girls within the juvenile justice system should have gender norms to appropriately account for developmental differences in gender. With so many juvenile offenders suffering from emotional problems, using the CAFAS to assess functioning in a juvenile justice population may be useful, but currently remains empirically untested.

Purpose of the Current Study

Bates (2001) concluded that the psychometric properties and clinical validity of the CAFAS need to be further researched, especially with different child and adolescent populations. Hodges has stated that the CAFAS is “supported by over 20 years of research and 80 published articles” and that it is “a robust, psychometrically sound measure” (Hodges, n.d.). However, taken at face value, this statement is misleading, in that many of the articles Hodges is referring to cite Hodges’ two evaluation studies psychometric results, and there are few that add any relevant and updated reliability and validity information to the literature. The purpose of the current study was to update reliability and validity information on the CAFAS within a juvenile justice population. If adequate reliability and validity of the CAFAS cannot be demonstrated with this population, then changes in the current assessment protocol for agencies serving the juvenile justice system would clearly be warranted.

Research Questions

The following research questions were investigated: 1) What is the reliability of the CAFAS with a juvenile justice population? 2) Is the CAFAS a valid measure to use with a juvenile justice population?

Additional Research Questions

In addition to investigating the two primary research questions, other demographic and psychometric aspects of the CAFAS were explored, including gender differences in CAFAS Total scores and subscale scores, and whether it would be useful to interpret subscale scores rather than the Total score for use in treatment planning.

Chapter Three

Methods

Participants

The current study was archival and was conducted using a computerized, de-identified database of existing CAFAS scores for the children and adolescents previously and currently enrolled in the juvenile justice (JJ) program at a community mental health center located in Southeast Michigan. The agency offers many mental health treatment and developmental programs for the surrounding community, including programs for children, adolescents, adults, and families. The specific program of focus for this study was the JJ program. This program is designed to help children and adolescents who have been court-ordered to participate, as well as those that may be considered at-risk by the legal system.

Since its inception in 2000, the juvenile justice program has served over 1000 participants, ranging from those who have participated in the program to those who are still receiving services. The current JJ population is comprised of 1121 youth. The average age for the overall sample was 15.05 years ($SD = 1.38$), and ranged from ages 10 to 18 years (see Table 1). There were 304 females and 817 males in the current database. This gender proportion is congruent with national juvenile justice statistics, which demonstrate that about 28% of arrested youth are female (McReynolds et al., 2008). For females, the average age was 14.88 years ($SD = 1.24$), and ranged from ages 11 to 18 years. For males, the average age was 15.12 years ($SD = 1.42$), and ranged from ages 10 to 18 years.

Table 1. Age ($N = 1035$)

	Males		Females		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
10	1	0.1	0	0	1	0.1
11	6	0.8	2	0.7	8	0.8
12	23	3.1	9	3.1	32	2.9
13	54	7.2	28	9.8	82	7.3
14	125	16.7	55	19.2	180	16.1
15	199	26.5	105	36.6	304	27.1
16	254	33.9	67	23.3	320	28.5
17	80	10.7	19	6.6	100	8.8
18	7	0.9	2	0.7	9	0.8

Sample Descriptives

Comparative demographic analyses were conducted to further examine the current population, provide updated normative information for the CAFAS, and explore the utility of the CAFAS for different demographic groups. Demographic variables examined included gender, ethnicity, income, insurance funding type (to determine financial resources), and age. The participant's primary DSM-IV TR diagnosis was also included to further examine the population.

Using the categories provided by the agency, ethnicity was separated into eight categories: European American, African-American, Native American Indian/Alaskan Native, Hispanic, Asian/Pacific Islander, Other, Arabic/Chaldean, and Multi-Racial (see Table 2). The majority of the JJ participants were either European American (72.9%) or African-American (16.9%) with much smaller numbers of other ethnicities, which is consistent with national juvenile justice statistics, which state 78% of juvenile offenders are European American, and 16% are African American (Knoll & Sickmund, 2007).

Similarly, Hodges' (1993) study provided data for only two ethnic groups (African American and European American).

Table 2. Ethnicity ($N = 1095$)

	Males		Females		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
European American	592	72.5	225	74.0	817	72.9
African-American	151	18.5	39	12.8	190	16.9
Native American/ Alaskan Native	6	0.7	1	0.3	7	0.6
Hispanic	40	4.9	16	5.3	56	5.0
Asian/ Pacific Islander	0	0.0	1	0.3	1	0.1
Arab/Chaldean	3	0.4	1	0.3	2	0.2
Multi-Racial	2	0.2	1	0.3	32	2.9
Other	19	2.3	13	4.3	4	0.4

Because of the wide ranges of specific family income values provided, income was collapsed into categories to help better organize and clarify the data. Table 3 provides the categories and frequencies of the income demographic. Based on recent national Census data from the United States Department of Health and Human Services, in America, the poverty level for a typical family of four was \$22,050. Because almost two thirds of the sample's income information was missing, another source of financial resources was examined, and that was Insurance Funding Source. (See Table 3.)

Table 3. Family Income ($N = 374$)

	Males		Females		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
1-15000	121	45.3	53	49.5	174	15.5
15001-30000	74	37.7	32	29.9	106	9.5
30001-45000	36	13.5	10	9.3	46	4.1
45001-60000	23	8.6	4	3.7	27	2.4
Above 60000	13	4.9	8	7.5	21	1.9
Missing					747	66.6

In order to provide further information on socioeconomic status of the family and the child provided with services, their insurance funding source (private insurance versus Medicaid) was examined. Approximately half of the current sample received Medicaid funding for services. (See Table 4.)

Table 4. Insurance ($N = 862$)

	Males		Females		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Medicaid	374	59.2	139	60.4	513	45.8
Private	258	40.8	91	39.6	349	31.1
Missing					259	23.1

As part of the initial intake session at the agency, all participants were given a primary DSM-IV TR diagnosis (see Table 5) to help define treatment goals and provide appropriate treatment. In addition, 61% of males had a secondary diagnosis, while 44% of females had a secondary diagnosis.

Table 5. Primary DSM-IV TR Diagnosis ($N = 1067$)

	Male		Female	
	Frequency	Percent	Frequency	Percent
Mood Disorder	100	0.13	80	0.27
Anxiety	11	0.01	5	0.02
Adjustment Disorder	79	0.10	48	0.16
Schizophrenia and other Psychotic Disorder	6	0.01	0	0.00
Substance Use Disorder	117	0.15	39	0.13
Oppositional Defiant Disorder	108	0.14	43	0.15
Conduct Disorder	58	0.07	15	0.05
Attention-Deficit/Hyperactivity Disorder	59	0.08	4	0.01
Impulse Control Disorder	13	0.02	5	0.02
Child/Adolescent Antisocial Behavior	8	0.01	2	0.01
Learning Disorder/MR	5	0.01	3	0.01
Disruptive Behavior Disorder	201	0.26	46	0.16
Abuse of Child	8	0.01	0	0.00
PTSD	0	0.00	1	0.00
Sexual Disorder	2	0.00	0	0.00
Total	775		291	

Participants also complete the Wechsler Intelligence Scale for Children-IV (WISC-IV). The WISC-IV is a measure used to identify cognitive strengths and weaknesses in several different areas of cognitive functioning: Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed (Wechsler, 2003). Scores are calculated for each individual subtest, as well as a score for a participant's Full Scale IQ (FSIQ). The FSIQ is the number used to describe a participant's overall cognitive functioning. For the current population, the average FSIQ score was 86.15 ($SD = 12.38$) (see Table 6).

Table 6. Participants' FSIQ ($N = 866$)

	Males		Females		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
40-54						
Moderate MR	5	0.01	0	0.00	5	0.5
55-69						
Mild MR	54	0.08	14	0.06	68	6.1
70-79						
Borderline	146	0.23	34	0.16	180	16.1
80-89						
Low Average	204	0.32	81	0.37	285	25.4
90-109						
Average	213	0.33	81	0.37	294	26.2
110-119						
High Average	17	0.03	9	0.04	26	2.3
120-129						
Superior	8	0.01	0	0	8	0.7
>130						
Very Superior	0	0	0	0	0	0
Missing					255	22.7

Design and Procedure

Participants were all adjudicated youth who had been court ordered to participate in services through the JJ program at a Community Mental Health Center in Southeast Michigan. Prior to beginning services at The Guidance Center, each youth undergoes an independent evaluation when they are first arrested and taken to the Wayne County Juvenile Detention Facility. This evaluation consists of a clinical battery, given to all youth, prior to service starting at The Guidance Center. For the purpose of the current study, the measures from the Juvenile Justice battery that are included are the Millon Adolescent Clinical Inventory (MACI) (Millon, et al., 2006) and the Wechsler Intelligence Scale for Children (WISC-IV) (Wechsler, 2003). Each youth also receives a DSM-IV diagnosis as part of this assessment. All JJ participants are then given the

CAFAS at the time of initial intake. For the current study, the CAFAS scores at intake were extracted from a larger database of multiple CAFAS scores accumulated over each individual's treatment period.

Measures

The CAFAS (Child and Adolescent Functional Assessment Scale) is used to measure functioning of children and adolescents in daily life activities. The CAFAS measures impairment in eight domains: school and/or work situations, home life, community attitudes, behaviors toward others, moods and emotions, self-harmful behavior, substance use, and thinking processes. There are also two caregiver domains, material needs and family/social support, which were not be examined for the current study. For a detailed description of what each subscale measures, see Appendix A.

The Millon Adolescent Clinical Inventory (MACI) was used to determine convergent validity for the CAFAS. It is a 160-item self report measure that is used to identify personality characteristics and clinical problems in adolescents (Millon, et al., 2006). The MACI was developed to be used in a variety of clinical settings, and the scores generated are used to develop diagnoses and treatment plans. The MACI is comprised of 31 scales which are broken into different types of scales: Personality Patterns (12), Expressed Concerns (8), Clinical Syndromes (7), and Modifying Indices (4). The Personality Patterns scales are based on Millon's classic theory of personality (Millon, 1969) and are based on combinations of three domains: pleasure-pain, active-passive, and self-other. There are eight scales that focus on "areas of life that troubled adolescents often find problematic" (Millon, Millon, Davis, & Grossman, 2006, p. 20),

and seven scales that are based on clinical problems that a therapist can help with, such as anxiety or eating disorders.

In order to establish convergent validity of the MACI scales, scales were chosen from the 31 scales to compare to comparable CAFAS subscales. For instance, the CAFAS Substance Use subscale was correlated with the Clinical Syndrome Scale BB: *Substance-Abuse Proneness*. Overall, seven MACI scales were chosen: Personality Pattern Scale 6A: *Unruly*; Personality Pattern Scale 8A: *Oppositional*; Expressed Concerns Scale G: *Family Discord*; Clinical Syndromes Scale BB: *Substance-Abuse Proneness*; Clinical Syndromes Scale CC: *Delinquent Predisposition*; Clinical Syndromes Scale FF: *Depressive Affect*; Clinical Syndromes Scale GG: *Suicidal Tendency*. (See Appendix B for specific MACI scale descriptions.)

The reliability and validity of the MACI has been established using a sample of more than 1,000 adolescents from the United States and Canada (Millon et al., 2006). Reliability was established by examining internal consistency and test-retest reliability. Coefficient alpha values ranged from .73 to .90, indicating that some scales had fair reliability values, while others had good reliability values. Test-retest reliability coefficients were computed within a range of three to seven days, and ranged from .57 to .92, again indicating that some scales were more reliable than others.

Validity for the MACI was established by correlating MACI scale scores with clinician judgment and other similar measures (convergent validity). To gain a clinical perspective, a clinician was asked to interview a youth and rate what personality characteristics (the Personality Patterns scales) seemed to best fit the youth. Correlations ranged from zero to .52, indicating there was some problem with identifying correct

personality characteristics. Millon et al. (2006) hypothesize that the lower correlations are because the clinician only has a limited time to make a clinical judgment about the youth, without having known them for any length of time. Convergent validity was established by correlating the MACI scale scores with measures assessing similar constructs. For instance, the MACI Eating Dysfunction scale was compared to the Drive for Thinness and Body Dissatisfaction measures of the Eating Behaviors Inventory—2. The correlations were both very high for both those comparisons, with a value of .75 computed between the MACI Eating Dysfunction scale and the Drive for Thinness measure, and a value of .88 computed between the MACI Eating Dysfunction scale and the Body Dissatisfaction measure. Millon et al. (2006) reported that when the construct examined were very similar, correlation coefficients were very high.

Chapter Four

Analyses and Results

Descriptive Statistics

An exploratory analysis of the data was conducted to examine the quality of the distributions of CAFAS scores (both individual subscale scores and Total score) across the combined genders. For the intake scores, some skewness on the individual subscales in the negative direction was to be expected. Negative skewness indicates a frequency of higher scores, and it was expected that the participants would score higher at intake (high scores equal more impairment). This was the case with the School/Work Role Performance, Community Role Performance, and Behavior Toward Others subscales. The skewness score of the Home Role Performance, Moods/Emotions, Substance Use, and Total CAFAS score were all close to zero, indicating a normal distribution of scores. The Self Harm and Think scores were both positively skewed, which indicates a presence of primarily low scores. This would also be expected, as few people in the sample were expressing a desire to harm themselves, or having problems with unusual thinking. See Table 7 for means and distribution of CAFAS scores at intake.

Table 7. Means (SD) and Distribution of CAFAS Scores at Intake ($N = 1121$)

	<i>M</i>	<i>(SD)</i>	Skew	Kurtosis	Mode
Total	111.20	(44.42)	-0.17	-0.59	130.00
School/Work Role Performance	22.11	(11.64)	-1.09	-0.50	30
Home Role Performance	17.50	(12.60)	-0.34	-1.55	30
Community Role Performance	25.11	(5.41)	-0.66	0.61	30
Behavior Toward Others	16.76	(9.72)	-0.48	-0.75	20
Moods/Emotions	11.41	(9.30)	0.16	-1.10	20
Self Harmful Behavior	3.60	(8.33)	2.11	3.01	0
Substance Use	14.20	(13.36)	0.10	-1.77	0
Thinking	1.43	(4.69)	3.61	13.32	0

Gender Difference on the CAFAS Total and Subscale Scores

To examine gender differences in CAFAS scores, both Total scores and each subscale score were separated by gender (see Table 8). Because the sample size of males and females was unequal, Levene's test for equality of variance was run, and where appropriate, adjusted degrees of freedom and significance values for unequal variances was utilized. Because of the multiple comparisons being conducted on the same data with the subscales, a conservative significance level was utilized ($p = .00625$) using a Bonferroni correction for gender comparisons across the eight subscales to protect against Type I error and alpha inflation. Independent samples t tests were performed to see if there were any differences between the genders on any of the CAFAS subscales or on the Total score. These analyses found significant differences on several of the CAFAS subscales, as well as the Total CAFAS score, as a function of gender, so subsequent analyses using the intake scores were conducted separately for the genders.

Table 8. Initial CAFAS Subscale Scores by Gender ($N = 1121$)

	<i>Male</i>		<i>Female</i>				
	Mean	(SD)	Mean	(SD)	<i>t</i>	df	<i>p</i>
CAFAS Total	109.16	(43.65)	116.78	(46.12)	-2.14	787	0.033*
School/Work Role Performance	22.15	(11.63)	22.01	(11.67)	0.18	1067	0.856
Home Role Performance	16.41	(12.61)	20.53	(12.07)	-4.87	521.34	<0.001**
Community Role Performance	25.41	(5.24)	24.26	(5.81)	3.09	1067	0.002**
Behavior Toward Others	16.71	(9.76)	16.9	(9.6)	-0.28	1067	0.780
Moods/Emotions	10.24	(9.28)	14.63	(8.6)	-6.95	1065	<0.001**
Self Harmful Behavior	2.76	(7.48)	5.92	(9.96)	-4.86	404.36	<0.001**
Substance Use	14.69	(13.35)	12.85	(13.32)	1.99	1067	0.047*
Thinking	1.36	(4.52)	1.62	(5.13)	-0.79	1067	0.430

*Significant at a $p < .05$

**Significant at a Bonferroni corrected p value $< .006$.

The CAFAS Total score was significantly different between the genders. This indicates that on average, the females ($M = 116.78$) received significantly higher scores, and therefore were more impaired than the males ($M = 109.16$). Females were also found to have significantly higher (more impaired) scores on the Home Role Performance subscale, Moods/Emotions subscale, and the Self-Harmful Behavior subscale, while the males were found to have significantly higher (more impaired) scores on the Community Role Performance subscale.

Comparative Data

In addition to Hodges' two primary normative studies, Table 9 also includes data from studies that examined different samples of youth participating in other types of programs, such as educational programs (Rosenblatt & Rosenblatt, 1999) or part of

Intensive Mental Health Programs (IMHP) Roy, Roberts, Vernberg & Randall, 2007; Vernberg et al., 2004). Hodges' (1999) sample was from the national CMHS System of Care Initiative, and included several subsamples of youth, such as youth who had been psychiatrically hospitalized, youth with low and average intelligence, and youth with contact with the law versus no contact. Hodges et al.'s (2000) study focused on groups of youth that with Serious Emotional Disturbance (SED) and that were involved with law or had poor school attendance, and Hodges, Xue & Wotring (2004)'s study focused on outcomes for youth with SED. Other SED populations were examined in different states (Walrath, Sharp, Zuber & Leaf, 2001; Williams, 2009), as well as other mental health populations (Kenaley & Williams, 2010; Nakamura, Daleidens & Mueller, 2007). Hodges et al. (1999) sample of youth who had contact with the law had similar CAFAS means at intake ($M = 112.24$) to the current sample of JJ youth ($M = 110.20$, $SD = 44.42$), though Walrath (2001) sample of JJ youth was not so similar ($M = 68.6$, $SD = 27.6$).

Table 9. Mean Total CAFAS score for Comparative Youth Samples at Intake and Discharge (when available)

	Intake		Discharge	
	Mean	(SD)	Mean	(SD)
Bodine (2011) Juvenile Justice (N = 1121)	111.20	(44.42)	52.01	(36.29)
Hodges (1993) Evaluation Project(N = 781)	45.65	(26.47)	31.39	(26.03)
Hodges, Wong & Latessa (1998) CMHC* (N = 179)	60.78	(25.01)	38.60	(25.74)
Rosenblatt & Rosenblatt (1999) Education*(a) (N = 61)	78.4	(19.3)		
Rosenblatt & Rosenblatt (1999) Education(b) (N = 82)	74.9	(26.6)		
Hodges et al. (1999) Nat. CMHS*(a) Psy. Hosp (N = 658)	107.4			
Hodges et al. (1999) Nat CMHS (b) Low IQ (N = 856)	101.13			
Hodges et al. (1999) Nat CMHS (c) Avg IQ (N = 1110)	88.62			
Hodges et al. (1999) Nat CMHS (d) Law* (N = 534)	112.24			
Hodges et al. (1999) Nat CMHS (e) No Law* (N = 1542)	86.71			
Hodges et al. (2000) SED (a) (N = 1629)	91.1	(44.5)		
Hodges et al. (2000) SED (b) (N = 797)	100.1	(45.1)		
Walrath et al. (2001a) (a) EBMHP* (N = 393)	68.04	(45.63)		
Walrath et al. (2001a) (b) FRIENDS*	77.50	(29.86)		
Walrath et al. (2001b) (a) Juvenile Justice	68.6	(27.6)		
Walrath et al. (2001b) (b) School	62.5	(26.4)		
Walrath et al. (2001b) (c) Mental Health(MH)	64.7	(26.5)		
Walrath et al. (2001b) (d) Social Services	57.2	(25.8)		
Walrath et al. (2001b) (e) Family	59.9	(26.1)		
Walrath et al. (2001b) (f) Other	60.9	(26.8)		
Hodges, Xue & Wotring (2004) SED (N = 5638)	89.35	(32.35)	63.14	(38.78)
Vernberg et al. (2004) IMHP*(N = 50)	122.20	(27.65)	62.60	(38.27)
Nakamura, Daleidens & Mueller (2007) MH (N = 121)	111.9	(33.8)		
Roy et al. (2008) IMHP (N = 70)	107.1	(32.9)		
Williams (2009) SED (N = 218)	112.61	(26.34)	73.81	(33.44)
Kenaley & Williams (2010) MH (N = 53)	111.89	(25.11)		

*Community Mental Health Center (CMHC), Educational Program, National Community Mental Health Services (Nat. CMHS), Contact with Law (Law), No Contact with Law (No Law), Serious Emotional Disturbance (SED), East Baltimore Mental

Health Partnership (EBMHP), Families Reaching in Ever New Directions (FRIENDS), Intensive Mental Health Services (IMHP)

Internal Reliability of the CAFAS Total and Subscale Scores

To address internal consistency of the CAFAS, coefficient alpha values and item to total correlations were calculated for the CAFAS Total score as well as separately for the eight subscales. For the CAFAS Total score, internal reliability using Cronbach's alpha was .68 across combined genders, .67 for males, and .70 for females (see Table 12). The item to total correlation values ranged from .20 to .57 among the subscales. George and Mallery (2003) provide the following rules of thumb: less than .5 is unacceptable; .5-.6 is poor; .6-.7 is questionable; .7-.8 is acceptable; .8-.9 is good; greater than .9 is excellent. All item to total correlations were below an acceptable value of reliability, in fact, all were almost unacceptable, and the overall alpha values would not increase if any of the items (or scales, in this case) were deleted. It is important to note that the ~~Item-~~ "Total" statistics presented in the table are the subscale to Total CAFAS scores, not individual items. Hodges' Total score coefficient alpha value at intake was .63 for the Ft. Bragg Evaluation Project sample, using the same statistical methodology.

Table 10. Internal Consistency (Alpha) for CAFAS Total and Subscale Scores

	Total		Males		Females	
	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Total	.68	--	.67	--	.70	--
School/work	.365	.648	.383	.635	.332	.687
Home	.567	.585	.564	.576	.572	.620
Community	.292	.666	.288	.659	.346	.684
Behavior	.533	.605	.510	.601	.602	.622
Mood	.434	.630	.421	.625	.470	.655
Self Harm	.326	.655	.312	.651	.345	.680
Substance						
Use	.284	.681	.274	.677	.348	.689
Think	.203	.677	.214	.669	.172	.705

The Relationship between the CAFAS subscales

In order to examine the relationship between the CAFAS subscales, each subscale was correlated with each other, as well as the CAFAS Total score. Many of the subscales were significantly correlated with each other (see Table 11), although most of these inter-scale correlations were in the small to moderate (.04-.37) range, with the exception of Community and Behavior subscales ($r = .55$).

Table 11. CAFAS Subscale Correlation (N = 1121)

Variable	1	2	3	4	5	6	7	8
1.S/W	---							
2.Home	.33**	---						
3.Com	.13**	.25**	---					
4.Beh	.30**	.55**	.30**	---				
5.MDS	.18**	.37**	.13**	.34**	---			
6.SHB	.10*	.25**	.12**	.20**	.37**	---		
7.Sub	.25**	.23**	.16**	.18**	.14**	.11**	---	
8.Think	.04	.14**	.06	.14**	.14**	.26**	.00	---
9.Total	.60**	.74**	.40**	.66**	.60**	.50**	.55**	.28**

** $p < .01$

* $p < .05$

CAFAS subscales include School/Work Role Performance (S/W), Home Role Performance (Home), Community Role Performance (Com), Behavior Toward Others (Beh), Moods/Emotions (MDS), Self Harmful Behavior (SHB), Substance Use (Sub), and Thinking (Think)

Criterion-Related Validity

In order to evaluate criterion validity of the CAFAS, the CAFAS Total score at intake was correlated with the number of offenses committed prior to intake. For males ($N = 817$), there was a small but significant correlation between CAFAS Total score at intake and the number of offenses committed prior to intake, $r = .24, p = .002$, such that as number of offenses increased, level of impairment as indicated by Total CAFAS score also increased. There was no significant correlation with the female sample ($N = 304$). Across combined genders ($N = 1121$), there was a small but significant correlation between CAFAS total scores at intake and the number of offenses committed prior to intake, $r = .12, p = .01$.

Another method to evaluate the criterion validity of the CAFAS was to compare the CAFAS total score at intake to the severity of the offenses committed prior to intake. The offenses were coded into one of three respective overarching categories: 1) Status

offense, 2) Misdemeanor, or 3) Felony offense. The participant's most severe offense was used for this analysis. Offense data was available for 358 participants in the program (256 males and 102 females) (see Table 12).

Table 12. Severity of Offenses at Intake ($N = 358$)

	<i>Males</i> ($N=256$)		<i>Females</i> ($N=102$)		<i>Total</i> ($N=358$)	
	<i>N</i>	<i>Percent</i>	<i>N</i>	<i>Percent</i>	<i>N</i>	<i>Percent</i>
Status Offense	61	23.8	53	52.0	114	31.8
Misdemeanor	12	48.8	41	40.2	166	46.4
Felony	5	27.3	8	7.8	78	21.8

Across combined genders, no significant difference in CAFAS Total score was found as a function of severity of offense at intake. Likewise, when males were examined separately, there was no significant effect of offense severity on the CAFAS Total score at intake. There was, however, a significant effect of offense severity on the CAFAS Total score at intake for females, $F(2,73) = 5.33, p = .007$. Hochberg's GT2 post hoc analyses were used due to unequal sample sizes. These analyses indicated there was a difference between those female participants who had a status offense versus a felony offense, such that those with a felony ($M = 161.43, SD = 36.71$) received significantly higher scores (more impaired) than those with a status offense ($M = 112.16, SD = 44.42$). The misdemeanor group was not significantly different from either the felony or status offense groups with regard to CAFAS Total score.

Convergent Validity

In order to examine convergent validity, correlations between initial CAFAS subscales and scores from subscales of the Millon Adolescent Clinical Inventory (MACI) were conducted. In order to investigate convergent validity, selected MACI subscales

were correlated with comparable CAFAS subscales (see Table 14). For instance, a participant's score on the MACI Substance Use scale was compared to their score on the CAFAS Substance Use subscale, with the expectation of a high correlation between two scales that were both designed to measure a participant's substance use. Results indicated that the relationship between the MACI and CAFAS substance use scales was in fact the highest correlation found between all comparable MACI and CAFAS subscales yet, even this was still only a medium correlation ($r = .36$). Most of the comparable MACI and CAFAS subscales were found to have only small correlations with each other, although given the large sample size, these correlations were statistically significant (See Table 13).

Table 13. MACI vs. CAFAS Subscale Correlation¹ ($N = 1,121$)

Variable	S/W	Home	Com	Beh	MDS	SHB	Sub	Think	Total
Unruly	.13**	.21**	.05	.20**	.10**	.09**	.16**	.04	.23**
Oppositional	.09**	.18**	.06	.23**	.20**	.18**	.08*	.09**	.23**
Family Discord	.10**	.28**	.02	.27**	.17**	.17**	0.07	.09*	.24**
Substance Use	.14**	.24**	.11**	.25**	.23**	.16**	.36**	.09**	.37**
Delinquent	.13**	.15**	.06	.17**	.04	.02	.21**	.00	.24**
Depressive Affect	.04	.16**	.13**	.18**	.31**	.19**	.00	.12**	.22**
Suicidality	-.03	.15**	.08**	.16**	.31**	.29**	.00	.18**	.21**

** $p < .01$

* $p < .05$

¹*Comparable CAFAS and MACI subscales are bolded and italicized.*

CAFAS subscales include School/Work Role Performance (S/W), Home Role Performance (Home), Community Role Performance (Com), Behavior Toward Others (Beh), Moods/Emotions (MDS), Self Harmful Behavior (SHB), Substance Use (Sub), and Thinking (Think)

The Relationship between the CAFAS and Ethnicity

An independent samples t -test was conducted to examine the relationship between ethnicity and the CAFAS scores at intake for females. Because of the small numbers of

different ethnicities, only the two largest categories of ethnicity (i.e. European American and African American) were used for subsequent analyses of possible ethnic differences. No significant differences were found on the CAFAS Total score between European American females ($M = 116.90$, $SD = 48.86$) and African American females ($M = 112.91$, $SD = 45.65$).

An independent samples t -test was conducted to examine the relationship between ethnicity and the CAFAS scores at intake for males. There was no significant differences found for Total CAFAS score as a function of ethnicity for the male sample; however, there was a significant difference found on the CAFAS Self Harmful Behavior subscale, $t(433.92) = 4.65$, $p < .001$, such that European American male participants ($M = 3.23$, $SD = 8.03$) received significantly higher (more impaired) scores than African American male participants ($M = .96$, $SD = 4.29$). Another significant difference was found such on the Moods/Emotions subscale, $t(713) = 4.01$, $p < .001$, such that European American male participants ($M = 10.95$, $SD = 9.37$) received significantly higher (more impaired) scores than African American male participants ($M = 7.53$, $SD = 8.35$).

The Relationship between the CAFAS and Family Financial Resources

An independent samples t -test was conducted to examine the relationship between insurance funding of the female participants and their CAFAS scores at intake. The funding source was collapsed into two groups: Medicaid and Private Insurance. There was a significant difference found between female participants on the CAFAS Behavior Toward Others subscale, $t(57.27) = -3.24$, $p = .002$, such that female participants with Medicaid ($M = 15.87$, $SD = 9.70$) received lower scores (less impairment) than those with Private Insurance ($M = 20.54$, $SD = 7.80$). There was also a significant difference found

on the CAFAS Substance Use subscale, $t(53.89) = 3.09, p = .003$, such that females with Medicaid ($M = 13.90, SD = 13.44$) received higher scores (more impairment) than females with Private Insurance ($M = 7.30, SD = 11.70$).

An independent samples t-test was conducted to examine the relationship between insurance funding of the male participant and their CAFAS scores at intake. There were no significant differences found between male participants with Medicaid on the CAFAS Total score ($M = 108.07, SD = 42.23$) and those with Private Insurance ($M = 113.54, SD = 43.82$).

The Relationship between the CAFAS and Primary DSM-IV TR Diagnosis

Due to the large number of different DSM-IV TR diagnoses that were given to participants, it was necessary to categorize the diagnoses into fewer categories that were diagnostically and conceptually congruent. Thus, each participant's primary diagnosis was classified into one of three groups: 1) Externalizing behavior disorders (such as ADHD or Conduct Disorder); 2) Internalizing disorders (such as Anxiety or Depression); or 3) Adjustment disorders, since these disorders can include both Internalizing and Externalizing problems, but by definition require an identified event or set of circumstances to which the youth is adjusting (see Table 14).

Table 14. DSM-IV TR Primary Diagnoses, by Group ($N = 1042$)

	<i>Males</i>		<i>Females</i>		<i>Total</i>	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Externalizing Disorders	117	15.5	85	29.6	718	68.9
Internalizing Disorders	564	74.7	154	53.7	202	19.4
Adjustment Disorders	74	9.8	48	16.7	122	11.7

A one-way Analysis of Variance (ANOVA) was conducted to examine differences between CAFAS Total and subscale scores for female participants, as a function of primary DSM-IV-TR diagnosis. Significant differences as a function of primary diagnosis were found for the CAFAS Total score, and many of the CAFAS subscales (see Table 16). There was a main effect of diagnosis found on the CAFAS Total score, $F(2, 197) = 3.45, p = .034$, as well as a main effect of diagnosis found on the CAFAS Home Role Performance subscale, $F(2, 270) = 3.74, p = .025$. There was also a main effect of diagnosis found on the CAFAS Community Role Performance subscale, $F(2, 270) = 4.70, p = .010$, as well as a main effect of diagnosis found on the CAFAS Moods/Emotions subscale, $F(2, 269) = 8.99, p < .001$. The last effect of diagnosis was found on the CAFAS Self-Harmful Behavior subscale, $F(2, 270) = 5.30, p = .006$. Pairwise comparisons were made using either Hochberg's GT2 or Games-Howell post hoc analyses, with respect to unequal variances and/or unequal sample sizes. Overall, those participants with an Internalizing disorder diagnosis were, on average, rated higher (more impairment) than those with an Externalizing behavior diagnosis or those with an Adjustment Disorder.

Table 15. Mean CAFAS score and Primary DSM-IV TR Diagnosis for females

	ID		ED		AD	
	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>
Total	131.40 ^a	(43.69)	115.09 ^b	(44.19)	109.03 ^b	(46.07)*
Home Role Performance	22.72 ^a	(11.18)	20.97	(11.63)	16.81 ^b	(13.53)*
Community Role Performance	25.93 ^a	(5.65)	24.00 ^b	(6.01)	22.98 ^b	(4.62)*
Moods/Emotions	18.13 ^a	(7.65)	13.24 ^b	(8.57)	14.47 ^b	(8.55)**
Self-Harmful Behavior	8.40 ^b	(10.89)	5.86 ^b	(9.97)	2.55 ^a	(7.06)*

* * $p < .001$

* $p < .05$

Significant differences between means are indicated with different superscript letters
Internalizing Disorder (ID), Externalizing Disorder (ED), Adjustment Disorder (AD)

A one-way Analysis of Variance (ANOVA) was conducted to examine differences between CAFAS Total and subscale scores for male participants, as a function of primary DSM-IV-TR diagnosis. Significant differences as a function of primary diagnosis were found for the CAFAS Total score, and many of the CAFAS subscales (see Table 17). There was a main effect of diagnosis found on the CAFAS Total score, $F(2, 538) = 3.11, p = .045$, as well as a main effect of diagnosis found on the CAFAS Home Role Performance subscale, $F(2, 734) = 5.25, p = .005$. There was also a main effect of diagnosis found on the CAFAS Moods/Emotions subscale, $F(2, 733) = 14.88, p < .001$, as well as a main effect of diagnosis on the CAFAS Self Harmful Behavior subscale, $F(2, 734) = 3.57, p = .029$. The last effect of diagnosis was found on the CAFAS Thinking subscale, $F(2, 734) = 3.83, p = .022$. Pairwise comparisons were made using either Hochberg's GT2 or Games-Howell post hoc analyses, with respect to unequal variances and/or unequal sample sizes. Overall, those participants with an Internalizing disorder diagnosis were, on average, rated higher (more impairment) than those with an Externalizing behavior diagnosis or those with an Adjustment Disorder.

Table 16. Mean CAFAS score and Primary DSM-IV TR Diagnosis for males

	ID		ED		AD	
	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>
Total	120.69	(41.42)	109.19	(43.02)	105.00	(42.33)*
Home Role						
Performance	19.56 ^a	(11.76)	16.39 ^b	(12.60)	13.70 ^b	(12.42)*
Moods/Emotions	14.73 ^a	(9.86)	9.56 ^b	(8.99)	10.14 ^b	(9.20)**
Self-Harmful Behavior	4.51	(9.16)	2.61	(7.32)	1.92	(6.38)*
Thinking	2.21 ^b	(5.63)	1.29 ^b	(4.35)	0.41 ^a	(2.60)*

* * $p < .001$

* $p < .05$

Significant differences between means are indicated with different superscript letters
Internalizing Disorder (ID), Externalizing Disorder (ED), Adjustment Disorder (AD)

Discussion

The two aims of the current study were: 1) evaluate the reliability and validity of the CAFAS with a juvenile justice child/adolescent population and, 2) improve the current CAFAS literature by providing normative data for a juvenile justice population. Evaluating the psychometric properties of the CAFAS helps address concerns expressed by Bates (2001), who called for clarification of the limited reliability and validity data currently available for the CAFAS (Hodges, 1993; Hodges & Wong, 1996), but it is also important for another reason. If adequate reliability and validity of the CAFAS cannot be demonstrated with the juvenile justice population, then the current assessment protocol for agencies serving the juvenile justice system would need to be reevaluated.

Gender Differences on the CAFAS Total and Subscale Scores

As an important initial step in the current study, the CAFAS Total and subscale scores were examined to determine whether there were differences as a function of gender. Hodges (1996) stated that there were no differences found between the genders on the CAFAS Total score, but her study looked at a non-forensic sample of youth. Research has shown that there are distinct differences between child and adolescent males and females involved in the juvenile justice system as a function of mental health (Cauffman, Lexcen, Goldweber, Shulman & Grisso, 2007), as well as a function of crime (Tracey, Kempf-Leonard & Abramoske-James, 2009). In fact, Tracy et al. (2009) stated, “Gender unequivocally is the most discriminating factor associated with crime.” In addition, it is still unclear if the same assessments should be used for both males and females involved in the juvenile justice system, as there are developmental and mental health differences between genders (Brumbaugh, Walters & Winterfield, 2010).

Significant differences at intake were found between males and females for a number of subscales, as well as the CAFAS Total score. Females had higher Total scores than males, indicating they had greater impairment, overall. Females were also found to have significantly higher scores (more impairment) on the CAFAS subscales related to mood and emotion, self harming behavior, and their behavior at home, while males were found to have significantly higher scores (more impairment) on the subscale assessing their behavior in the community. This suggests that males and females in the juvenile justice system are very different in presentation at the time of initial intake; specifically, when the CAFAS was used in the current JJ sample, females were seen as having more traditional internalizing problems than males. Also, while males were seen as having more externalizing problems in the community setting, females' externalizing problems were more likely to be viewed as emerging in the context of home life.

Internal Reliability of the CAFAS Total and Subscale Scores

The Cronbach alpha for the CAFAS Total score for the current JJ sample (.68) indicated the CAFAS did not quite reach the commonly acceptable level of internal consistency (.70) for the overall sample (Cortina, 1993). Kline (1999) stated that psychological tests measuring different constructs (i.e. cognitive, ability, or other psychological constructs) have different acceptable alpha values. For instance, when examining a cognitive test, a value of .8 is acceptable, while ability tests use the cut-off point of .7. Psychological tests measuring other constructs may have acceptable alpha levels below .7. Even so, the current alpha is somewhat higher than Hodges' (1996) alpha of .63 with the Ft. Bragg Evaluation Project sample. Although statistically significant with the current sample of 1121, most of the individual CAFAS subscales had

only small to medium (.2-.3), correlations with each other CAFAS subscale, or in the case of the Thinking subscale, non-significant correlations with other subscales due to the minimal variability on the Thinking scale. Hodges (1993) has stated that only moderate correlations would be expected between the CAFAS subscales, because the subscales were designed so that each subscale would measure a different facet of functioning. Cortina (1993) stated that though some people misinterpret internal consistency as homogeneity, even a test designed to measure a specific construct with separate and non homogeneous subscales should still have an acceptable level of internal consistency. Overall, the results of the current study support the conclusion that the CAFAS subscales should be interpreted separately, rather than using a Total score for treatment purposes (Hodges, 2003), as the CAFAS is not homogeneous and does not demonstrate great internal consistency among the subscales. For example, if each subscale is measuring a different facet of functioning, then it does not make sense to plan the same treatment for a youth with a high Total score due to an inflated Self Harmful Behavior score, as for a youth with a high Total score due to an inflated Community Role Performance score.

Concurrent Validity: The Relationship between the CAFAS and Offenses Prior to Intake

In the overall sample with genders combined, there was a small but statistically significant correlation between the number of offenses committed prior to intake and the severity of the participant's CAFAS Total score. The relationship suggests that the more offenses a youth commits prior to intake, the higher their CAFAS Total score, which means higher impairment and worse functioning overall. Even though the relationship is small, it is expected that someone who is in more contact with the law is probably

functioning worse in other areas of their life, as well, and would affect their CAFAS rating.

However, when genders were examined separately, the relationship between number of offenses and CAFAS Total score held up for males, but not for females. In the current sample, females had no significant relationship between number of offenses prior to intake and their CAFAS Total score at intake. This could be explained by the initial gender differences found with this sample. Females tended to receive worse ratings in the areas related to moods and emotions, as well as self harmful behavior, and behavior at home. Problems rated on these subscales are most likely occurring at home, which means less contact with the law. Males seem to receive worse ratings on behaviors in the community, which would provide direct community contact with unlawful behaviors (i.e., criminal offenses for which they could be charged). Another explanation for this finding is that when females are participating in unlawful activities, they are more likely to be sent to inpatient mental treatment facilities, rather than juvenile justice programs (Tracey et al., 2009). Thus, the females may not even be making it into the same programs as males for some of the same behaviors.

In examining the relationship between the severity level of the offenses committed prior to intake and the participant's CAFAS Total score, it was interesting that there was only an effect of offense severity for the females. One reason for the lack of an overall effect of offense severity across both genders may be that 68% of the overall offense history was missing for the sample. The effect of offense severity for the females, however, was such that the eight females who committed a felony (the most severe offense category) had significantly higher CAFAS Total scores than the 53

females who committed a status offense (the lowest severity category). Although the result makes intuitive sense, given that the effect was not present for the males, it is unclear if it is really due to a significant difference between the females who committed the worst crimes versus the females who committed the least offensive crimes, or rather, some other attribute that is unique to the very small sample of females with felony offenses.

Convergent Validity: The Relationship between the CAFAS and the MACI

Significant correlations between the CAFAS subscales and comparable MACI scales would support convergent validity for the CAFAS, as do significant correlations found in Hodges' (1993) study which compared the CAFAS to the Child Behavior Checklist, the Child Assessment Schedule, and the Burden of Care Questionnaire. Although there were many significant correlations found in the current comparison, the magnitude of all of these correlations was quite small, with the exception of the Substance Use scales. In order to provide strong support for convergent validity, the subscales should be highly correlated; however, even such similar scales as the CAFAS Self Harmful Behavior and the MACI Suicidality scale were only found to have a small correlation (.29). Other scales that should have been correlated, but were found to have only a modest relationship at best, were the CAFAS Community Role Performance subscale and the MACI Unruly, Oppositional, and Delinquent scales. As indicated by the small size of these correlations, the shared variance between each of the two comparable scales was quite small, indicating little convergent validity. An explanation for the low correlation values may be that the CAFAS is a level of functioning scale, while the MACI is a diagnostic tool designed to identify pathology. The structure of the two tests

is different, and the MACI has questionable psychometric properties, itself. The difference in reasoning behind the items on both measures (i.e. identifying a substance use problem versus examining how/if a substance is affecting a youth's daily functioning) may be demonstrated in the presentation of the items, and may contribute to the lack of correlation found between the two measures.

The Relationship between the CAFAS and Demographic Variables

The CAFAS Total score, as well as the previously noted significant subscales, were examined as a function of Ethnicity and Family Financial Resources. Ethnicity (European American versus African American) was a significant factor for male participants on the Self Harmful Behavior subscale, with European American males being rated higher (more impaired) than African American males at intake. Historically, European American males have been more likely to use self harming behaviors or commit suicide (Kubrin & Wadsworth, 2009) than African American males. The current study supports this finding. The current study found no differences on the CAFAS based on financial resources of the family. This result is congruent with what Hodges (1993) found in her Ft. Bragg Evaluation Project sample. However, recent research has found that lower income is associated with higher occurrence of mental illness (Preidt, 2011). Hodges (1993) found income may have more influence over the family and caregiver scales that are not part of the current study (and which have no impact on the CAFAS Total score).

The Relationship Between the CAFAS and Primary Diagnosis

In examining participants' differences in CAFAS scores based on their DSM-IV TR primary diagnosis, there were significant differences on almost every subscale. This

suggests that mental health plays a serious role in the overall functioning of the youth. With the diagnoses grouped into internalizing disorders, externalizing disorders, and adjustment disorders, it was evident that youth with primarily internalizing mental health diagnoses (i.e., anxiety or depression) were significantly different from those with either externalizing behavior problems or adjustment disorders, as they were rated higher (more impaired) on many CAFAS subscales at intake. In fact, youth with internalizing disorders were rated significantly higher on the CAFAS Total score, the Home Role Performance subscale, Community Role Performance subscale, Moods/Emotions subscale, and Self Harmful Behavior subscale. These findings make sense, because youth with mood or anxiety disorders tend to have more problems at home, with their moods or emotions, and can also have more self harmful behavior, and therefore, higher scores on the CAFAS. The finding that youth with internalizing disorders were rated higher on the Community Role Performance subscale was unusual, in that we would expect youth with externalizing disorders to demonstrate worse behavior in the community. Youth with internalizing disorders may have diagnoses that are more pervasive in their overall emotional and behavioral functioning.

The differences found based on diagnosis are further evidence that the CAFAS subscales should be interpreted separately for treatment planning. A youth who is rated as having higher scores (more impairment) on the Self Harmful Behavior scale is likely to be very different from a youth with a similar score on the Community Role Performance subscale, and these subscale differences should be taken into consideration when planning individual treatment. It is also important to note that in the current juvenile justice population, 61% of males had a secondary diagnosis, while 44% of

females had a secondary diagnosis. These diagnoses, while not categorized as the youth's primary diagnosis, may well have an effect on their functioning, as well, and should also be taken into consideration when planning treatment for the youth.

Summary and Conclusions

In terms of the original research questions, ~~“What is the reliability of the CAFAS with a juvenile justice population?”~~ and ~~“Is the CAFAS a valid measure to use with a juvenile justice population?”~~ the CAFAS was not found to have overwhelming support for internal reliability, and still may be lacking in convergent validity. In examining reliability, the coefficient alpha values for the scale to Total reliabilities did not reach a conventionally acceptable level, and although the subscales were significantly correlated, these were small correlations. For criterion related validity, there was only a small (yet significant) relationship between number of offenses committed prior to intake, and this result was only found for males. In examining the severity of offenses committed prior to intake, there was only a significant effect between some groups of females (status offense versus felony offense), and no effect of offense severity for males, or for the overall combined sample. Lastly, in terms of convergent validity, the CAFAS scales and comparable MACI scales did not share much variance. Though the scales are not designed to assess the same construct (i.e. level of functioning versus diagnostic tool), there is little to no evidence of convergent validity for scales between the two that were designed to measure similar constructs such as self harmful behavior or unruly behavior.

Although it is currently used as a measure of level of functioning, the CAFAS may have limitations in accurately assessing level of functioning within a juvenile justice population. If, however, it is continued to be used, it is recommended that some

modifications should be made in terms of how it is used. First, CAFAS Total scores should not be interpreted as a total level of functioning or impairment. The subscales were not designed to be correlated, and they do not correlate highly together, suggesting each subscale needs to be rated, interpreted, and used, on its own. This is not currently the way the CAFAS is being used by either clinical or juvenile justice staff (Vignola, personal communication, April 14, 2011; Rykert, personal communication, April 14, 2011) at the site the current data was gathered.

Gender also played a significant role in these subscale differences. Clearly, the female juvenile youth are different from the male juvenile youth. This was evident by the CAFAS subscales (and behaviors) on which the females were receiving worse scores. Their diagnoses are different, the number and severity of offenses committed prior to intake is different, and the overall fact that there are not as many female juvenile offenders leads to limited information.

Significant differences based on DSM-IV TR primary diagnosis were found as well. Even though this is a juvenile justice program, and the curriculum is focused on unlawful behavior, mental health does seem to play a role in what behaviors some youth are having difficulty with. There are specific subscales which seem to target mental health, while others seem to target behavior, and they should be interpreted separately, so that every youth gets the exact services they need.

Limitations

The current study examined CAFAS Total and subscale scores within a juvenile justice population, while previous studies (Hodges, 1993; Hodges et al., 1998) examined samples of youth involved with mental health treatment programs. Means and standard

deviations from the current juvenile justice population ($M = 111.20$, $SD = 44.42$) at intake were very different than either the Ft. Bragg Evaluation Project ($M = 45.65$, $SD = 26.47$) or the CMHC study ($M = 60.78$, $SD = 25.01$). Recent research (Kenaley & Williams, 2010; Vernberg et al., 2004; Nakamura et al., 2007) has examined other mental health samples, and found their youth's mean CAFAS scores at intake to be more similar to the current study of JJ youth (see Table 11). One of the primary contributions of the current research was to provide normative data for males and females in the juvenile justice population, however, more research needs to be done to address norms and demographic information for other samples of forensic populations where the CAFAS is being used.

Another limitation of the current study regards the amount of missing data for some outcome or grouping variables. For instance, about 66% of the income data was missing for the families of the participants in the JJ program, so that demographic variable could not be examined further. Offense history was also incomplete, so it was difficult to make broad inferences where there was limited data available. In addition, the primary diagnosis of youth was examined, but not the secondary diagnosis, which could have an effect on the youth's overall functioning or specific subscales they are receiving worse ratings on.

Another limitation was related to the CAFAS scoring protocol itself and the nature of the data set that was available, in that the database available for the current study included scores for each subscale, as well as Total score, but not raw responses to the individual items that comprised each subscale score. In this regard, while the current analyses are based on the same level of analysis as that provided by Hodges (1993) in previously reported reliability studies of the CAFAS subscales and Total scores, the

current study supports the conclusion of Bates (2001) that additional analyses of the individual items comprising each subscale of the CAFAS is warranted.

In addition, the internal consistency for the CAFAS was established by Hodges (1993) and replicated in the current study by using the coefficient alpha test. Though it appears to be the most appropriate statistic to evaluate the internal consistency of the CAFAS, there is some concern about only having eight subscale ~~items~~ available for analysis. Cortina (1993) reported there was a direct link between the alpha level and the number of items on a scale, but also stated that ~~alpha~~ is a sound measure of proportion of error variance regardless of test length” (p.101). Similar to the current study, coefficient alpha was used by Boone (1998) to establish internal consistency values of the Personality Assessment Inventory (PAI) for all the scales as well as subscales, and the internal consistency of the MMPI was also established using coefficient alpha (Butcher et al., 1989).

Future Considerations

Future research should further examine the reliability, validity, and utility of individual CAFAS subscales for the juvenile justice population, rather than just CAFAS Total scores. These individual subscales could be interpreted for specific interventions, but at the present time some agencies only utilize the CAFAS Total score. Furthermore, although the juvenile justice staff from the agency the current sample data is collected from is all trained on the CAFAS, they are typically not clinically trained, and may not be completing the CAFAS in a reliable and consistent way compared to mental health staff (Rykert, personal communication, 2011). This may be more of a programmatic issue that should be examined on an individual basis.

Other suggestions for future research include looking at how the CAFAS performs over time as an outcome measure. Examining program variables such as the average length of time a participant is in the program, or the length of time in the program required to reach the discharge CAFAS score, or looking more closely at those participants who dropout or are terminated from the program prior to a successful conclusion, are all areas of the CAFAS literature that warrant further research.

References

- Achenbach, T. M. (1978). Psychopathology of childhood: Research problems and issues. *Journal of Consulting and Clinical Psychology, 46*, 759-776.
doi:10.1037/0022-006X.46.4.759
- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T.M., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*, 1275-1301. doi:10.1037/0033-2909.85.6.1275
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, D.C.: Author.
- Bates, M. P. (2001). The Child and Adolescent Functional Assessment Scale (CAFAS): Review and current status. *Clinical Child and Family Psychology Review, 4*, 63-84. doi:10.1023/A:1009528727345
- Bates, M.P., Furlong, M.J., & Green J.G. (2006). Are CAFAS subscales and item weights valid? A preliminary investigation of the Child and Adolescent Functional Assessment Scale. *Administration and Policy in Mental Health, 33*, 682-695. doi:10.1007/s10488-006-0052-4
- Bickman L., Guthrie P. R., Foster E. M., Lambert E. W., Summerfelt W. T., Breda C. S., Heflinger, C. A. (1995). *Evaluating Managed Mental Health Services: The Fort Bragg Experiment*. New York: Plenum Press.

- Boone, D. (1998). Internal Consistency Reliability of the Personality Assessment Inventory with Psychiatric Inpatients. *Journal of Clinical Psychology*, 54, 839-843. doi:10.1016/j.psychres.2008.11.001
- Breda, C.S. (1996). Methodological issues in evaluating mental health outcomes of a children's mental health managed care demonstration. *The Journal of Mental Health Administration*, 23, 40-50. doi: 10.1007/BF02518642
- Brumbaugh, S., Hardison Walters, J.L., & Winterfield, L.A. (2010). Girls Study Group: Understanding and responding to girls' delinquency. U.S. Department of Justice, Office of Justice Programs, April 2010.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring. University of Minnesota Press: Minneapolis, MN.
- Cauffman, E., Lexcen, F. J., Goldweber, A., Shulman, E.P., & Grisso, T. (2007). Gender differences in mental health symptoms among delinquent and community youth. *Youth Violence and Juvenile Justice*, 5, 287-307. doi:10.1177/1541204007301292
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284- 290. doi:10.1037/1040-3590.6.4.284
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104. doi:10.1037/00219010.78.1.98

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334. doi:10.1007/BF02310555
- Davis, D. L., Bean, G. J., Schumacher, J. E., & Stringer, T. L. (1991). Prevalence of emotional disorders in a juvenile justice population. *American Journal of Forensic Psychology*, 9, 5-17. Retrieved from <http://rave.ohiolink.edu/databases/record/psyz/1991-18705-001>.
- Fite, P.J., Wynn, P., & Pardini, D.A. (2009). Explaining discrepancies in arrest rates between Black and European American male juveniles. *Journal of Counseling and Clinical Psychology*, 77, 916-927. doi:10.1037/a0016626
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Graves, K. N., Frabutt, J. M., & Shelton, T. L. (2007). Factors associated with mental health and juvenile justice involvement among children with severe emotional disturbance. *Youth Violence and Juvenile Justice*, 5, 147-167. doi:10.1177/1541204006292870
- Greenwood, P. (2008). Prevention and intervention programs for juvenile offenders. *Future of Children*, 18, 185-210. Retrieved from http://www.princeton.edu/futureofchildren/publications/docs/18_02_09.pdf
- Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston: Allyn and Bacon.
- Heffer, R.W., Barry, T.D., & Garland, B.H. (2009). History, overview, and trends in child and adolescent psychological assessment. In J.L. Matson, F. Andrasik, &

- M.L. Matson (Eds.). Assessing childhood psychopathology and developmental disabilities (pp. 3-29). New York: Springer.
- Hodges, K. (n.d.) Child and Adolescent Functional Assessment Scale (CAFAS) Overview of Reliability and Validity. Retrieved from:
http://www.fasoutcomes.com/RadControls/Editor/FileManager/Document/FAS611_CAFAS%20Reliability%20and%20Validity%20Rev10.pdf
- Hodges, K. (1989). Child and Adolescent Functional Assessment Scale. Ypsilanti: Eastern Michigan University.
- Hodges, K. (1993). Structured interviews for assessing children. *Journal of Child Psychology and Psychiatry*, 34, 49-68. doi: 10.1111/j.1469-7610.1993.tb00967.x
- Hodges, K. (2000a) Child and Adolescent Functional Assessment Scale. Ypsilanti, MI: Eastern Michigan University.
- Hodges, K. (2000b) Child and Adolescent Functional Assessment Scale Self-Training Manual. Ypsilanti, MI: Eastern Michigan University.
- Hodges, K. (2003). *CAFAS manual for training coordinators, clinical administrators, and data managers*. Ann Arbor, MI: Author.
- Hodges, K. (2004). Child and Adolescent Functional Assessment Scale (CAFAS). In M.E. Maruish (Ed), *The use of psychological testing for treatment planning and outcome assessment*. (3rd ed., pp. 405-441). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hodges, K., Doucette-Gates, A., & Kim, C.S. (2000). Predicting service utilization with the Child and Adolescent Functional Assessment Scale in a sample of youths with serious emotional disturbance served by Center for Mental Health Services-

- funded demonstrations. *The Journal of Behavioral Health Services & Research*, 27, 47-59. doi:10.1007/BF02287803
- Hodges , K., Doucette-Gates, A., & Liao, Q. (1999). The relationship between the Child and Adolescent Functional Assessment Scale (CAFAS) and indicators of functioning. *Journal of Child and Family Studies*, 8, 109-122. doi:10.1023/A:1022902812761
- Hodges, K. & Kim, C.S. (2000). Psychometric study of the Child and Adolescent Functional Assessment Scale: Prediction of contact with the law and poor school attendance. *Journal of Child and Family Studies*, 5, 445-467. doi:10.1023/A:1005100521818
- Hodges, K., Kline, J., Stern, L., Cytryn, L., and McKnew, D.(1982). The development of a child assessment interview for research and clinical use. *Journal of Abnormal Child Psychology*, 10, 173-189. doi:10.1007/BF00915939
- Hodges, K. & Saunders, W. (1989). Internal consistency of a diagnostic interview for children: The Child Assessment Schedule. *Journal of Abnormal Child Psychology*, 17, 691-701. doi:10.1007/BF00917731
- Hodges, K. & Wong, M. M. (1996). Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale. *Journal of Child and Family Studies*, 5, 445-467. doi:10.1007/BF02233865
- Hodges, K. & Wong, M. M. (1997). Use of the Child and Adolescent Functional Assessment Scale to predict service utilization and cost. *Journal of Mental Health Administration*, 24, 278-290. doi:10.1007/BF02832662

- Hodges, K., Wong, M. M., & Latessa, M. (1998). Use of the Child and Adolescent Functional Assessment Scale (CAFAS) as an outcome measure in clinical settings. *Journal of Behavioral Health Services and Research*, 25, 325-336.
doi:10.1007/BF02287471
- Hodges, K., Xue, Y., & Wotring, J. (2004). Use of the CAFAS to evaluate outcome for youths with severe emotional disturbance served by public mental health. *Journal of Child and Family Studies*, 13, 325-339.
doi:10.1023/B:JCFS.0000022038.62940.a3
- Jonson-Reid, M., Williams, J. H., & Webster, D. (2001). Severe emotional disturbance and violent offending among incarcerated adolescents. *Social Work Research*, 25, 213-222. Retrieved from
<http://web.ebscohost.com.proxy.ohiolink.edu:9099/ehost/pdfviewer/pdfviewer?sid=942adfab-df7b-4a2e-86a9-cfa575aa3f45%40sessionmgr104&vid=2&hid=110>
- Kenaley, B. L. D. & Williams, N.J. (2010). A preliminary evaluation: Demographic and clinical profiles and changes in functioning in children receiving psychosocial rehabilitation. *Children and Youth Services Review*, 33, 301–307.
doi:10.1016/j.childyouth.2010.09.013
- Knoll, C., & Sickmund, M. (2010). Delinquency Cases in Juvenile Court, 2007. U.S. Department of Justice, Office of Justice Programs, June 2010.
- Kovacs, M. (2003). *Children's Depression Inventory: Technical Manual*. North Tonawanda, NY: Multi-Health Systems, Inc.
- Kubrin, C.E., & Wadsworth, T. (2009). Explaining suicide among Blacks and European

- Americans: How socioeconomic factors and gun availability affect race-specific suicide rates. *Social Science Quarterly*, 90, 1203-1227. doi:10.1111/j.1540-6237.2009.00654.x
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. doi:10.2307/2529310
- McReynolds, L.S., Wasserman, G. A., DeComo, R. E., John, R., Keating, J. M., & Nolen, S. (2008). Psychiatric disorder in a juvenile assessment center. *Crime & Delinquency*, 54, 313-334. doi:10.1177/0011128707301629
- Millon, T. (1969). Modern psychopathology: A biosocial approach to maladaptive learning and functioning. Philadelphia: W. B. Saunders. (Reprinted 1985 by Waveland Press, Prospect Heights, IL).
- Millon, T., Millon, C., Davis, R., & Grossman, S. (2006). MCMI-III Manual (3rd ed.). Minneapolis: NCS Pearson.
- Myers, K., & Winters, N. C. (2002). Ten-year review of rating scales. II: Scales for internalizing disorders. *Journal of American Academic Child and Adolescent Psychiatry*, 41, 634-659. doi:10.1097/00004583-200206000-00004
- Nakamura, B.J., Daleiden, E.L., & Mueller, C.W. (2007). Validity of treatment target progress ratings as indicators of youth improvement. *Journal of Child and Family Studies*, 16, 729-741. doi: 10.1007/s10826-006-9119-9
- Ogles, B. M., Davis, D. C., & Lunnen, K. M. (1999). Inter-rater reliability of four measures of youth functioning. In C. Liberton, K. Kutash, & R. Friedman (Eds.), *The 11th Annual Research Conference Proceedings, A System of Care for Children's Mental Health: Expanding the Research Base* (pp. 321-326). Tampa,

FL: University of South Florida, Florida Mental Health Institute, Research and Training Center for Children's Mental Health.

Ohlin, L. E. (1998). The future of juvenile justice policy and research. *Crime & Delinquency*, 44, 143-153. doi:10.1177/001112878302900306

Preidt, R. (04/04/2011). Poverty linked to anxiety, mood disorders. *MedicineNet.com*.

Retrieved from

<http://www.medicinenet.com/script/main/art.asp?articlekey=142682>.

Quist, R. M. & Matshazi, D. G. M. (2000). The Child and Adolescent Functional Assessment Scale (CAFAS): A dynamic predictor of juvenile recidivism. *Adolescence*, 35, 181-192. Retrieved from

http://findarticles.com/p/articles/mi_m2248/is_137_35/ai_62959087/

Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavior Assessment System for Children* (2nd ed.). Minnesota: AGS Publishing.

Rosenblat, J.A., & Rosenblat, A. (1999). Youth functional status and academic achievement in collaborative mental health and education programs: Two California care systems. *Journal of Emotional and Behavioral Disorders*. 7, 21-30. doi:10.1177/106342669900700103

Rosenblatt, J. A., Rosenblatt, A., & Biggs, E. E. (2000). Criminal behavior and emotional disorder: Comparing youth served by the mental health and juvenile justice systems. *Crimes and Mental Health Disorders*, 27, 227-237. doi:10.1007/BF02287315

Roy, K. M., Roberts, M. C., Vernberg, E. M., & Randall, C. J. (2008). Measuring treatment outcomes for children with serious emotional disturbances:

- Discriminant validity and clinical significance of the Child and Adolescent Functioning Assessment Scale. *Journal of Child and Family Studies*, 17, 232-240. doi:10.1007/s10826-007-9162-1
- Spielberger, C. D. (1973). Preliminary Manual for the State-Trait Anxiety Inventory for Children. Palo Alto, California: Consulting Psychologists Press.
- Stephens, R.L., Petras, H., Fabian, A., & Walrath, C.M. (2009). Patterns of functional impairment and their change among youth served in systems of care: An application of latent transition analysis. *Journal of Behavioral Health Sciences & Research*, 37, 491-507. doi:10.1007/s11414-009-9186-y
- Tracy, Paul E., Kimberley Kempf-Leonard and Stephanie Abramoske-James 2009. Gender differences in delinquency and juvenile justice processing: Evidence from national data. *Crime and Delinquency*, 55, 171-215. doi:10.1177/0011128708330628
- United States Department of Justice, Office of Juvenile Justice and Delinquency Prevention, <http://ojjdp.ncjrs.org/faqs.html>, retrieved June 12, 2009. United States Department of Justice, Office of Juvenile Justice and Delinquency Prevention.
- Vernberg, E.M., Jacobs, A.K., Nyre, J.E., Puddy, R.W., & Roberts, M.C. (2004). Innovative treatment for children with serious emotional disturbance: Preliminary outcomes for a school-based intensive mental health program. *Journal of Clinical Child and Adolescent Psychology*, 33, 359-365. doi:10.1207/s15374424jccp3302_17
- Walker, R., Minor-Schork, D., Bloch, R., & Eisenhart, J. (1996). High risk factors for

rehospitalization within six months. *Psychiatric Quarterly*, 67, 235–243.

doi:10.1007/BF02238954

Walrath, C., DosReis, S., Miech, R., Liao, Q., Holden, W., De Carolis, G., Satiago, R., & Leaf, P. (2001). Referral source differences in functional impairment levels for children served in the comprehensive community mental health services for children and their families program. *Journal of Child and Family Studies*, 10, 385-397. doi:10.1023/A:1012533127953

Walrath, C.M., Sharp, M.J., Zuber, M., & Leaf, P.J. (2001). Serving children with SED in urban systems of care: Referral agency differences in child characteristics in Baltimore and the Bronx. *Journal of Emotional and Behavioral Disorders*, 9, 94-105. doi:10.1177/106342660100900204

Wechsler, D. (2003). Wechsler Intelligence Scale for Children-Fourth Edition.

Administration and scoring manual. San Antonio, TX: Harcourt Assessment, Inc.

Williams, N.J. (2009). Preliminary evaluation of children's psychosocial rehabilitation youth with serious emotional disturbance. *Research on Social Work Practice*, 19, 5-18. doi:10.1177/1049731507313996

Zagar, R. J. (2009). Identification, treatment, and prevention of homicide: Fallacies in research, treatment, and policy—A postscript on youth violence. *Psychological Reports*, 104, 363-377. doi:10.2466/PRO.104.1.363-377

Zagar, R. J., Busch, K. G., & Hughes, J. R. (2009). Empirical risk factors for delinquency and best treatments: Where do we go from here? *Psychological Reports*, 104, 279-308. doi:10.2466/PRO.104.1.279-308

Appendix A
Brief Descriptions of Each CAFAS Subscale
(manual)

Subscales Assessing The Youth

School/Work	Ability to function satisfactorily in a group educational environment
Home	Extent to which youth observes reasonable rules and performs age appropriate tasks
Community	Respect for the rights of others and their property and conformity to laws
Behavior Toward Others	Appropriateness of youth's daily behavior
Moods/Emotions	Modulation of the youth's emotional life
Self-Harmful Behavior	Extent to which the youth can cope without resorting to self-harmful behavior or verbalizations
Substance Use	Youth's substance use and the extent to which it is not appropriate or is disruptive
Thinking	Ability of youth to use rational thought processes

Subscales Assessing the Caregiver

Material Needs	Extent to which the youth's functioning is interfered with due to lack of resources, such as food, clothing, housing, medical attention or neighborhood safety
Family/Social Support	Extent to which the youth's functioning is disrupted due to limitations in the family's psychosocial resources relative to the youth's needs

Appendix B

Millon Adolescent Clinical Inventory

Scale Descriptions

The MACI scales assess the major personality characteristics, the primary realms of psychosocial concern, and the most important clinical syndromes among adolescents. The following sections describe these scales and the characteristics they measure. (These are the scales we chose and used for analyses)

Personality Patterns

The MACI test measures 12 Personality Patterns, which are based on the author's theoretical scheme (Millon, 1969, 1981, 1986a, 1986b, 1990) and the most recent versions of the DSM. Because many clients exhibit more than one of these patterns, the degree to which the adolescent characteristically displays each pattern is expressed in the scale profile configuration.

Scale 6A: Unruly. High scorers tend to act out in an antisocial manner, often resisting efforts to make them adhere to socially acceptable standards of behavior. These adolescents may display a pervasively rebellious attitude that could create conflict with parents and school or legal authorities.

Scale 8A: Oppositional. High scorers tend to be discontented, sullen, and passive-aggressive. They often behave unpredictably. They may be outgoing and pleasant one minute and hostile and irritable the next. These individuals often feel confused and contrite about their moodiness, but they seem unable to control their moods for very long.

Expressed Concerns

Eight scales focus on areas of life that troubled adolescents often find problematic. The intensity of the client's concern is reflected in the elevation of the scale score.

Scale G: Family Discord. Those who score high on this scale report that their families are tense and full of conflict. Few sources of support are noted, and there is a general feeling of estrangement from their parents. Depending on the individual's personality, these problems may reflect either parental rejection or adolescent rebellion.

Clinical Syndromes

Seven MACI scales involve areas of clinical significance that call for intervention on the part of the therapist. These diagnostic categories represent problems that are found in a significant proportion of adolescents who are seen by mental health professionals.

Scale BB: Substance-Abuse Proneness. High scorers evince a maladaptive pattern of alcohol and/or drug abuse that has led to significant impairment of their performance and behavior. Many spend an inordinate amount of time obtaining these substances, behave in a socially unacceptable manner, and have continued to use alcohol and/or drugs even though they know they have persistent and deleterious effects on their lives.

Scale CC: Delinquent Predisposition. High scorers evince behavior that has led or likely to lead to situations in which the rights of others are violated. Such inclinations may result in a variety of societal norms or rules being broken. This may include threatening others, using weapons, persistent deceptiveness or lying, stealing, and other antisocial behavior.

Scale FF: Depressive Affect. Those who score high on this scale exhibit a decreased level of activity, clearly distinct from what has been characteristic of them in the past. They exhibit a notable decrease in effectiveness, feelings of guilt, fatigue, a tendency to despair about the future, social withdrawal, loss of confidence, and diminished feelings of adequacy and attractiveness.

Scale GG: Suicidal Tendency. High scorers admit to having suicidal thoughts and plans. They express feelings of worthlessness and purposelessness. A sense that others would be better off without them is a common aspect of this type of thinking. High scores call for professional attention and alertness on the part of family members.