PREDICTING TREATMENT OUTCOMES WITH

CAREGIVER-YOUTH DISCREPANCY REPORTING PATTERNS

IN A MULTIETHNIC PUBLIC MENTAL HEALTH SAMPLE

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PSYCHOLOGY

AUGUST 2023

By

Tommie M. Laba

Dissertation Committee:

Brad Nakamura, Chairperson
Charles Mueller
Emily Blumenthal
David Jackson
Joseph Mobley, University Representative

Keywords: multi-informant reporting, evidence-based practices, informant discrepancies, youth outcomes

Acknowledgements

I would like to express my sincerest gratitude and appreciation to everyone who contributed to this research and supported me along this journey. To my advisor and committee chair, Dr. Brad Nakamura, thank you for providing me with your valuable guidance throughout all stages of this project. Your ongoing support and mentorship has been incredibly meaningful for my growth, both personally and professionally. I would also like to express my deepest appreciation to Dr. Charles Mueller, Dr. Emily Blumenthal, Dr. David Jackson, and Dr. Joseph Mobley for serving as my committee members and providing valuable insight and feedback for the current project. To Dr. Kentaro Hayashi, thank you for being so exceedingly generous with your time and providing feedback on the interpretation of my study results. Moreover, I am very grateful to have had the opportunity to learn from the late Dr. Ronald Heck, who dedicated countless hours supporting the academic dreams of so many students, including myself. Dr. Heck, you are deeply missed. I would also like to thank my labmates for their time assisting with various aspects of the current project, and my colleague, Dr. Andrew Young Choi, for his time and valuable consultation. Furthermore, without the partnership and support of the Hawaiʻi State Department of Health Child and Adolescent Mental Health Division, this work would not have been possible. Thank you to my friends, Dr. Krista Brown, Kalyn Holmes, and Amanda Vincent, for enriching my graduate school experience in many ways. I feel so blessed to have you in my life. To my parents, Pattie and Jared Komori, thank you for your unconditional love and support. I appreciate all that you have done and continue to do for me. Finally, to my husband, Chad Laba, I will never be able to express how meaningful your love and encouragement in life and throughout graduate school has been to me. Thank you for being my rock.

Abstract

Multi-informant reporting has long been recommended as one of several best practices for assessing youth psychopathology. In children's mental health, this approach includes gathering information from the youth as well as others in their ecology. Gathering information from multiple perspectives yields benefits through broadening available data and examining cross-informant reporting patterns that might affect assessment findings and treatment recommendations. However, for many decades it has been acknowledged that in both clinical and research settings, informants often provide diverging perspectives when reporting about youth emotional and behavioral concerns. These diverging perspectives have been studied extensively and are commonly referred to in the literature as "informant discrepancies" (or low cross-informant rater agreement). More recently, researchers have found that such discrepancies at treatment intake have been linked to poor youth outcomes and responses to intervention. In line with this area of work, the current study examined the relation between caregiver-youth informant discrepancies at treatment intake and youth functional impairment at discharge using a multiethnic public mental health sample of 366 caregiver-youth dyads. Prior to investigating the current study's major aim, preliminary confirmatory factor analyses involving the Ohio Scales Problem Severity domain was conducted to inform subsequent analytic strategies. These confirmatory factor analysis results suggested utilizing both the Total Scale and Subscale (i.e., Externalizing, Delinquency, Anxiety, and Depression) scores of the Problem Severity domain for my major aim, which was to examine the extent to which caregiver-youth discrepancy scores at intake into a mental health system predicted youth functional impairment at treatment discharge. Overall, while there were slightly mixed findings across Total Scale and Subscale scores, results generally supported an association between increased levels of youth functional impairment at treatment discharge with both (a) high levels of caregiver-youth agreement (i.e., high caregiver and high youth Problem Severity domain scores) and (b) higher caregiver-reports of problem severity compared to their respective youth at intake. The present findings contribute to the current literature by utilizing the latest statistical approach recommendations for research in this area, incorporating measurement schemes that include identical item content across reporter versions, and including an ethnically-diverse public mental health sample of youths and their caregivers. Implications and suggestions for future research are discussed.

Table of Contents

LIST OF APPENDICES

List of Tables

List of Figures

Predicting Treatment Outcomes with Caregiver-Youth Discrepancy Reporting Patterns in a Multiethnic Public Mental Health Sample

Multi-informant reporting has long been recommended as one of several best practices for assessing children and adolescent (herein after collectively referred to as “youth”) psychopathology (Kraemer et al., 2003). In children’s mental health, this approach includes gathering information from the youth as well as others in their ecology, often including caregivers, teachers, and mental health providers (Hunsley & Mash, 2007). Gathering information from multiple perspectives is often not only essential (e.g., reporting limitations due to a child’s developmental age), it also yields benefits through broadening available data and examining cross-informant reporting patterns that might affect assessment findings and treatment recommendations. This includes, for example, assignment of mental health diagnoses, evidence-based practice recommendations, and ongoing treatment planning (De Los Reyes & Kazdin, 2008; De Los Reyes et al., 2015; Weisz et al., 2005). However, for many decades it has been acknowledged that in both clinical and research settings, informants often provide diverging perspectives when reporting about youth emotional and behavioral concerns. These diverging perspectives are commonly referred to in the literature as “informant discrepancies” (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005).

Informant Discrepancy

Informant discrepancy refers to low-cross informant rater agreement, and early researchers in this area initially conceptualized this phenomenon as a form of measurement error due to psychometric or methodological inaccuracies (e.g., Edgeworth, 1888 [as cited in De Los Reyes et al., 2013]; Roberts & Caspi, 2001). Although an important and contributing factor to informant discrepancies, measurement error alone does not fully account for low cross-informant rater agreement, and a breadth of literature has revealed another side of the argument; namely, that informant discrepancies may instead largely reflect important and meaningful information about the development of youth mental health concerns (De Los Reyes, 2011; De Los Reyes & Ohannessian, 2016).

Informant discrepancies have long been found in both clinical and research settings, with cross-informant rater agreement generally falling in the low to moderate range (Achenbach et al., 1987; Carlston & Ogles, 2006; Chen et al., 2017; De Los Reyes et al., 2015; Verhulst & Ende, 1992). Achenbach et al.'s (1987) seminal meta-analysis, including 119 studies published between 1960 to 1986 examining cross-informant rater agreement for youth with emotional and behavioral concerns, noted mean Pearson product correlation *r* values ranging from .20 to .60, with an average *r* value of .28 for overall internalizing and externalizing concerns. Furthermore, when looking specifically at caregiver-youth dyads, the authors found a mean *r* value of .25. Following this line of research, De Los Reyes et al. (2015) conducted an updated meta-analysis including 341 studies that were published between 1989 and 2014. Consistent with Achenbach et al.'s (1987) findings, their results indicated significant low to moderate Pearson product correlation *r* values, with an average *r* value of .25 for internalizing concerns, .30 for externalizing concerns, and .28 for overall or combined problem areas. Specifically regarding caregiver-youth dyads, the authors found average *r* values of .26 for internalizing concerns and .32 for externalizing concerns (De Los Reyes et al., 2015).

Literature in this area also suggests that the directionality of informant discrepancies may be related to the type of sample involved in a study (e.g., non-clinical versus clinical sample). For instance, some studies involving general population or non-referred participants have found that youths on average reported higher levels of mental health concerns compared to their respective caregivers (Chen et al., 2017; Rescorla et al., 2013; Seiffge-Krenke & Kollmar, 1998; Verhulst & Ende, 1992; Wang et al., 2014). Other studies including clinical samples, however, have found that caregivers on average reported higher levels of mental health concerns compared to their respective youths (Becker-Haimes et al., 2018; Carlston & Ogles, 2009; Humphreys et al., 2017; Kolko & Kazdin, 1993).

There is also a long line of literature investigating the extent to which informant discrepancies are moderated along the three dimensions of informant type, youth demographic background characteristics (e.g., age, gender, ethnicity/race), and the type of problem area assessed. For example, Achenbach et al. (1987) and De Los Reyes et al. (2015) found that significantly higher levels of cross-informant rater agreement have been demonstrated between individuals who play generally similar roles and observe a youth in the same setting (e.g., pairs of caregiver reporters, pairs of teacher reporters), compared to individuals who serve different roles and observe a youth in different settings (e.g., a caregiver and teacher reporter pair). Regarding the last two of the three aforementioned factors (i.e., youth demographic background characteristics, type of problem area assessed), in general, early studies have been notably inconsistent and no clear picture has yet emerged (Achenbach et al., 1987; Baumgartner et al., 2020; Carlston & Ogles, 2006, 2009; Chen et al., 2017; Chun, 2008; De Los Reyes et al., 2015; Edelbrock et al., 1986; Ferdinand et al., 2006; Grills & Ollendick, 2003; Kolko & Kazdin, 1993; Lapouse & Monk, 1958; Rey et al., 1992; Seiffge-Krenke & Kollmar, 1998; Thompson et al., 1993; van der Ende et al., 2020; Verhulst & Ende, 1992; Wang et al., 2014; Yeh & Weisz, 2001). Notably, many of these studies pre-date both De Los Reyes and Kazdin's (2005) and De Los Reyes et al.'s (2013) seminal and theory driven framework (i.e., Attribution Bias Context Model, Operations Triad Model) in this area, which is generally agnostic about the roles of the most heavily studied domains of youth age, gender, ethnicity/race, and problem area.

Theory Driven Framework: The Attribution Bias Context (ABC) Model and Operations Triad Model (OTM)

Despite the multitude of descriptive findings summarized thus far, as with other areas of clinical psychology, research and practice can potentially intertwine in more of a synergistic way when seemingly disparate results are placed within a larger theoretical framework. Indeed, in their seminal paper in this area on informant discrepancies, De Los Reyes and Kazdin (2005) called attention to a lack of a theoretically-driven framework at that time, and formally put to the field their ABC Model (De Los Reyes & Kazdin, 2005).

The ABC Model postulates that informant discrepancies arise due to the interaction amongst three components, including variations in (a) youth and informant attributions for the cause of a behavior, (b) perspectives about the severity of a behavior and it subsequently warranting treatment, and (c) the goals of the clinical assessment. More specifically, the ABC Model posits that non-youth informants often involved in the assessment process (e.g., caregivers and teachers who observe the youth across settings) are likely to make attributions that point to inherent or dispositional causes for a behavior problem (i.e., the problematic behavior is due to the disposition of the youth). Conversely, youths however, are more likely to make attributions that instead point to environmental causes for a behavior problem. Second, and related to the aforementioned point, given that youths' and their informants' attributions initially vary with regard to the causes of a behavior, they often hold different perspectives about a behavioral concern warranting treatment or not. Along these lines, the ABC Model suggests that an informant who attributes the cause of a behavior to a youth's disposition will likely believe that the youth will benefit from treatment services, whereas a youth who attributes the cause of a behavior to the environment is less likely to believe that their behavior warrants treatment. Lastly, the goals of the clinical assessment, which fundamentally involves the identification of emotional and behavioral concerns that warrant treatment, also may or may not align with the youth and informant attributions and perspectives. Taken together, these three components are said to contribute to varying levels of informant discrepancy seen across both research and clinical settings (De Los Reyes & Kazdin, 2005).

Building upon this theory driven framework, De Los Reyes et al. (2013) introduced the OTM, which provides a complementary set of guidelines for the prediction and interpretation of multi-informant assessment results. The OTM conceptualizes multi-informant reporting patterns in three main ways, namely converging operations, diverging operations, and compensating operations. Specifically, under the condition of converging operations, the OTM claims that a youth's concerns manifest consistently across the contexts of which the youth is observed, and as such, agreement is expected between informant reports under this circumstance. Conversely, under the condition of diverging operations, a youth's concerns vary across the contexts of which the youth Is observed, and as such, meaningful variation exists between informant reports under this circumstance. Said another way, diverging operations reflect "contextual variation" or information that is relevant and necessary to have in order to gain a better understanding of a youth's mental health concerns. This type of data is commonly referred to in the literature as "domain-relevant information" (De Los Reyes & Epkins, 2023). Similarly, Achenbach et al. (1987) also put forth the notion of meaningful contextual variation regarding youth mental health concerns but referred to this phenomenon as "situational specificity." Related to this area of research, informant discrepancies have also been associated with other variables outside the scope of the current investigation, including for example, high levels of family stress and conflict, and differing perspectives about family relationships or the home environment (e.g., De Los Reyes et al., 2019; Guo & Slesnick, 2013; Human et al., 2016; Mastrotheodoros et al., 2020). Lastly, under the condition of compensating operations, the OTM posits that variation between informant reports that are unrelated to a youth's concerns instead reflect measurement confounds (e.g., rater biases, random error). To note, however, the informant discrepancy literature thus far generally suggests that the most common type of reporting pattern is that of diverging operations, where discrepant multi-informant reports reflect differences in domain-relevant information pertinent to the assessment of youth mental health concerns (De Los Reyes & Epkins, 2023; De Los Reyes et al., 2022).

Taken together, with regard to how therapeutic services may unfold over time, De Los Reyes and Kazdin's (2005) ABC Model suggests that if the attributions and perspectives about a behavioral problem vary at the start of therapy, there will inherently be differences in how much a treatment plan does or does not align with the beliefs of the youth and their caregivers. These initial differences could then interfere with treatment participation and subsequently, youth and/or caregiver responses to intervention. Along these lines, more recent studies have shifted their focus from investigating the more readily researched parameters previously discussed above (e.g., age, gender, ethnicity/race) to exploring the extent to which informant discrepancies predict future youth outcomes.

Informant Discrepancies and Youth Outcomes

The literature examining the association between informant discrepancies and mental health outcomes is complicated, varying along numerous domains such as research design, measurement strategies, analytic approaches, and youth problem types. Broadly speaking, however, these types of studies fall into two larger categories: investigations that do, and do not, consider directionality differences when examining the association between caregiver-youth discrepancy scores and outcomes.

Findings That Do Not Consider Directionality Differences in Caregiver-Youth Discrepancy Scores

There have been mixed findings with regard to the general relation (i.e., without considering directionality of discrepancy scores) between baseline informant discrepancies and subsequent youth outcomes or responses to intervention. More specifically, some studies have pointed to poor outcomes with increased levels of informant discrepancy while others have instead indicated no association.

Research Supporting a Relation Between Informant Discrepancies and Poor Youth Outcomes. As an example of an earlier study in this area, Carlston (2004) examined the association between caregiver-youth informant discrepancy scores at treatment intake with changes in mental health symptoms and treatment discharge status for youths receiving outpatient services in California. Overall, their results suggested that higher levels of informant discrepancy regarding internalizing and externalizing concerns at intake were related to significantly fewer positive changes in problem severity at the end of treatment according to caregiver-report. Relatedly, youths identified as treatment "completers" had significantly smaller caregiver-youth discrepancy scores at treatment intake, compared to youths identified as "pre-mature terminators" of treatment (Carlston, 2004). In addition, Humphreys et al.'s (2017) study involving youths recruited for a triple-blind, placebo-controlled, randomized clinical trial found that treatment "non-responders" had higher levels of caregiver-youth informant discrepancy scores regarding youth symptoms of posttraumatic stress disorder prior to receiving treatment.

Research Supporting No Relation Between Informant Discrepancies and Youth Outcomes. Generally speaking, while some literature in this area has provided evidence for an association between informant discrepancies and subsequent youth outcomes, there are a few studies that have instead suggested no such relation. For example, De Los Reyes et al.'s (2010) study including youths receiving services through a controlled clinical trial investigated the extent to which pre-treatment caregiver-youth discrepancy scores, regarding youth social phobia concerns, were related to end of treatment youth responses to intervention. Results indicated that there were no significant differences in pre-treatment caregiver-youth discrepancy scores for youths identified as treatment "responders" versus "non-responders" (De Los Reyes et al., 2010). Moreover, Becker-Haimes et al. (2018) found that caregiver-youth discrepancy scores regarding youth anxiety symptoms prior to receiving treatment were not related to youth outcomes at post-treatment. Their sample included youths diagnosed with generalized, separation, or social anxiety disorders. In particular, the authors found no association between levels of caregiver-youth informant discrepancy scores and (a) severity of anxiety concerns, (b) youth functioning, (c) responses to treatment, or (d) diagnostic remission status after receipt of treatment services (Becker-Haimes et al., 2018).

Although there have been mixed findings in this area, it is important to note that there has been wide variability in analytic approaches utilized within the studies reviewed thus far (e.g., utilization of raw, standardized, or residual difference scores compared to polynomial regression approaches), which has historically made it difficult for our field to draw sufficient conclusions (De Los Reyes & Kazdin, 2004). However, despite this limitation, the literature in this area has continued to grow in many ways and has provided helpful information about the relation between informant discrepancy scores and youth outcomes, including for example, the extent to which the directionality of caregiver-youth informant discrepancy scores are predictive of later outcomes and responses to intervention.

Findings That Consider Directionality Differences in Caregiver-Youth Discrepancy Scores

Prior to delving into research that has found poorer outcomes depending on the direction of discrepancy scores, it is important to note that one study provided evidence for a relation between poor youth outcomes and informant discrepancies when caregivers reported more problems than their youths and vice versa. In particular, Makol et al.'s (2019) study involving youths (and their caregivers) receiving inpatient care from a major, urban, Mid-Atlantic medical center, found that compared to dyads where both a caregiver and their respective youth reported low levels of internalizing concerns (low caregiver and low youth report), youths in dyads characteristic of (a) high caregiver and low youth reports, (b) low caregiver and high youth reports, and (c) high caregiver and high youth reports were more likely to have an anxiety disorder at treatment discharge. Additionally, compared to dyads in the low caregiver and low youth report group, youths in the three aforementioned groups were more likely to be partially hospitalized or receive services in an additional restrictive environment following treatment discharge (Makol et al., 2019). However, despite this broader finding pointing to a relation between poor youth outcomes and informant discrepancies in either direction, it is important to note that a large portion of the literature has pointed to poorer or worse youth outcomes when caregivers in particular reported higher levels of problem severity compared to their respective youth.

Findings Suggesting Poorer Outcomes When Caregivers Reported Higher Levels of Problem Severity Compared to Youth. Over 2 decades ago, Kendall et al. (1997) were one of the first to report that the directionality of caregiver-youth informant discrepancy scores, with regard to the presence or absence of anxiety-related concerns at treatment intake, were related to subsequent youth treatment outcomes. Specifically, the authors noted that significantly more treatment gains were observed for youths classified as “disclosers” versus “nondisclosers” (i.e., youths who reported no anxiety symptoms while their respective caregiver did; Kendall et al.,1997).

Additionally, another study involving a Dutch community sample drawn from a larger epidemiological investigation found that when caregivers (compared to their respective youths) reported higher levels of attention problems, delinquent behavior, and aggressive behavior, youths were more likely to experience job or school expulsion, have contact with the police or judicial system, receive a referral for mental health services, and struggle with drug use (Ferdinand et al., 2004). Similarly, Ferdinand et al.’s (2006) study, which included caregiver-youth dyads referred for outpatient services at a university clinic in Rotterdam, Netherlands, also found that when caregivers (compared to their respective youths) reported higher levels of attention, anxiety, and depressive problems, youths were more likely to experience school-related disciplinary problems (e.g., truancy, disruptive behavior concerns, instances of suspension or expulsion) and have contact with the police or judicial system (Ferdinand et al., 2006).

Furthermore, although the previously described study by Becker-Haimes et al. (2018) found no general association between levels of baseline informant discrepancy and responses to intervention, the authors found evidence for poorer youth outcomes when looking specifically at the direction and magnitude of pre-treatment discrepancy scores within study intervention conditions. In particular, caregiver-youth dyads were grouped in cross-informant agreement categories based on both the direction and magnitude of their discrepancy scores at pre-treatment. The “Parent Reported High” group was composed of dyads where caregivers reported the same or higher levels of symptom severity compared to their respective youth, and the “Youth Reported High” group was composed of dyads where youths reported higher levels of symptom severity compared to their respective caregiver. The authors noted that youths receiving cognitive behavioral therapy in the “Parent Reported High group” were less likely to achieve diagnostic remission at the end of treatment (Becker-Haimes et al., 2018).

Another recent study explored this association with participants presenting with an array of psychosocial-related difficulties, including attention-deficit/hyperactivity, mood, anxiety, obsessive-compulsive, and autism spectrum disorders (Goolsby et al., 2018). Their sample included caregiver-youth dyads receiving youth mental health services through a large private practice in Washington, D.C. Overall, study results indicated that higher levels of caregiver-youth pre-treatment discrepancy scores regarding depression, anxiety, and atypical behaviors were related to poor youth treatment outcomes (i.e., significantly less improvement on youth social skills development, anger control, emotional self-control, negative emotionality, and ability to manage stress and overcome adversities), while higher levels of pre-treatment discrepancy scores regarding hyperactivity concerns was related to better youth treatment outcomes (i.e., significantly more improvement on youth emotional self-control and negative emotionality). With regard to the directionality of caregiver-youth informant discrepancy scores, the most improvement was seen for youths in dyads where both a caregiver and youth similarly reported low levels of symptom severity. These dyads were followed by dyads where both caregiver- and youth-report similarly suggested high levels of youth symptom severity. Little to no improvement (sometimes worsening of symptoms) regarding outcomes, was seen for youths in dyads where (a) caregivers reported high levels of symptom severity while their respective youth reported low levels of symptom severity, and (b) youths reported high levels of symptom severity while their respective caregiver reported low levels of symptom severity; though to a greater degree for the former compared to the latter (Goolsby et al., 2018). At the same time, however, some studies have instead found poorer outcomes when youths in particular reported higher levels of problem severity.

Findings Suggesting Poorer Outcomes When Youths Reported Higher Levels of Problem Severity Compared to Caregiver. For instance, in Ferdinand et al.'s (2004) study, when youths (compared to their respective caregivers) reported higher levels of attention problems and delinquent behavior, they were more likely to receive a referral for mental health services as well as feel a need for mental health support. Similarly, Ferdinand et al. (2006) found that when youths (compared to their respective caregivers) reported higher levels of aggressive behavior, they were more likely to struggle with future drug use.

More recently, Augenstein et al.'s (2022) study, investigating the extent to which caregiver-youth informant discrepancy regarding depressive symptoms were related to youth suicidal ideation, found worse treatment outcomes when youths reported higher levels of problem severity compared to their caregivers. Study participants included caregiver-youth dyads receiving inpatient psychiatric care in a children's hospital located in a major urban area. Notably, results suggested that compared to three other identified informant discrepancy pattern groups (i.e., dyads where a youth and their respective caregiver similarly reported a "subthreshold" level of youth depressive symptoms; dyads where a youth and their caregiver similarly reported "elevated" youth depressive symptoms; and dyads where caregivers reported elevated levels of depressive symptoms compared to their respective youth who reported subthreshold levels of symptoms), when youths compared to their caregivers reported higher levels of depressive symptoms at treatment intake, they experienced significantly higher levels of suicidal ideation at 3-month follow-up (Augenstein et al., 2022).

Additionally, to investigate the association between baseline caregiver-youth informant discrepancy scores and future mental health problems diagnosed in adulthood, van der Ende et al. (2020) utilized two analytic approaches, including (a) multitrait-multimethod – correlated state-correlated method minus 1 (CS-C[M-1]) change model analyses and (b) interaction terms in polynomial regression. Their sample drew from a larger Zuid-Holland longitudinal study that began in 1983. With regard to the CS-C(M-1) change model results, high caregiver-youth informant discrepancy scores for externalizing concerns at baseline, specifically for dyads where caregivers reported higher levels of symptom severity compared to their respective youth, was related to lower likelihood of a youth having an externalizing DSM–IV diagnosis as an adult. However, results also varied within this study depending on analytic approach, with their polynomial regression analysis indicating no significant relation between caregiver-youth discrepancy scores and receiving an internalizing or externalizing mental health diagnosis as an adult (van der Ende et al., 2020).

Thus, while meaningful advancements have been made with the establishment of useful and underlying theoretical frameworks (De Los Reyes & Kazdin, 2005; De Los Reyes et al., 2013), studies have continued to demonstrate mixed findings regarding the association between informant discrepancies and future youth outcomes. Overall, however, it appears that larger initial caregiver-youth discrepancies are likely related to poor youth outcomes, especially when caregivers endorse higher levels of problem severity compared to their respective youths. It also seems that poor youth outcomes can be associated with increased levels of caregiver-youth agreement, particularly when both a caregiver and youth similarly report high levels of youth mental health concerns. Additionally, out of the aforementioned studies above, only five (i.e., Becker-Haimes et al., 2018; Goolsby et al., 2018; Humphreys et al., 2017; Makol et al., 2019; van der Ende et al., 2020) have utilized a currently recommended statistical approach (i.e., interaction terms in polynomial regression analysis; Laird & De Los Reyes, 2013) or other modern analytic strategies (i.e., latent class analysis, CS-C[M-1] change model approach). Furthermore, to date, most studies have included youth- and caregiver-report measurement tools that slightly vary across reporter versions (e.g., comparing Child Behavior Checklist and Youth-Self-Report [CBCL and YSR; Achenbach & Rescorla, 2001] or Behavior Assessment System for Children [BASC; Reynolds & Kamphaus, 2015] caregiver and youth-report scores), which may inadvertently contribute to levels of discrepancy observed between informant reports (Carlston & Ogles, 2006). Finally, literature in this area has mainly involved participants from less diverse populations, especially in comparison to multiethnic youths and caregivers frequently seen in large community mental health systems. As such, the existing literature on informant discrepancies would likely benefit from further research examining the relation between informant discrepancies and youth outcomes, using the latest statistical approach recommendations, measurement schemes that include identical item content across both youth- and caregiver-report versions, and ethnically-diverse youths and caregivers.

The Current Study

The current study examined the relation between caregiver-youth informant discrepancies at treatment intake and youth functional impairment at discharge using a large, multiethnic public mental health sample in Hawai‘i. Prior to investigating the current study’s major aim, preliminary analyses were conducted to examine the fit of a hierarchical factor structure for the Ohio Scales Short Form, Problem Severity domain (Ogles et al., 1999) to determine if a second-order factor (i.e., Total Scale) existed among the four first-order factors (i.e., Externalizing, Delinquency, Anxiety, and Depression) identified in Laba et al. (2022). This was done to inform subsequent analytic strategies involving caregiver-youth informant discrepancy scores on the Problem Severity domain. Following this preliminary step, the extent to which caregiver-youth informant discrepancy scores at treatment intake into a mental health system predicted youth functional impairment at discharge was examined. To note, I had two hypotheses for the current study. Consistent with previous research in this area (Goolsby et al., 2018; Makol et al., 2019), I first hypothesized that when both a caregiver and their respective youth similarly reported high levels of problem severity, this would predict poorer youth functional impairment at treatment discharge compared to dyads where both a caregiver and their respective youth similarly reported low levels of problem severity. Additionally, as suggested by the ABC model (De Los Reyes & Kazdin, 2005), differences with regard to caregiver and youth perspectives (potentially driven by attributions) about behavioral difficulties may impact informant discrepancy scores. From the start of treatment, these differences could inherently mean less alignment with treatment planning and client goals, which might then interfere with treatment participation and subsequently, youth responses to intervention. Thus, in line with the ABC Model and past research (e.g., Goolsby et al., 2018; Humphreys et al., 2017), I secondarily hypothesized that larger caregiver-youth discrepancy scores at the start of treatment would be significantly predictive of poorer youth functional impairment at treatment discharge, especially when caregivers reported higher levels of problem severity compared to their respective youth.

Methods

Participants

The current study included youths (and their caregivers) receiving public mental health services through the Hawai‘i State Department of Health Child and Adolescent Mental Health Division's (CAMHD's) system of care. CAMHD is the major statewide public sector entity responsible for delivering an array of mental health services to a wide range of youths and families in the state of Hawai‘i, including youths with special needs, youths traditionally classified as having severe emotional and behavioral difficulties, and families eligible for Medicare and Medicaid assistance. Youth services are provided in a variety of forms, ranging from case management to community- and residential-based settings. To be included in the current study, a youth must have (a) been between the ages of 12 to 18, (b) received at least 90 days of treatment within their episode of CAMHD care, and (c) had an intake Problem Severity domain measure completed by both the youth and their respective caregiver within +/-45 days of intake into the CAMHD treatment episode under study. A total of 366[1] youths and their caregivers were included in the current investigation after the aforementioned inclusionary criteria were applied. Although, the current sample had characteristics similar to other large service sector studies involving CAMHD youth (e.g., McLennan et al., 2022), there were slightly

[1] Determining an adequate sample size to conduct a study is dependent on a wide variety of factors, including model size, the distribution of variables, missing data, reliability of the data, and the extent to which variables are related (Muthén & Muthén, 2002). Relatedly, Muthén and Muthén (2002) recommend using a Monte Carlo study to determine sufficient sample size in a study. Along these lines, the statistical expert on my dissertation committee, the late Dr. Ronald Heck, ran preliminary Monte Carlo studies including 133 individuals to determine power for the current investigation (cf., Heck & Thomas, 2015). Extensive consultation with Dr. Heck indicated that the current study will be adequately powered with a sample size of at least 133 youth- and 133 corresponding caregiver-reports to detect relatively small effects (R. Heck, personal communication, September 3, 2021).

more female identified youths as well as participants diagnosed with depressive concerns included in the current study. More specifically, as shown in Table 1, youth participants were 52.2% female (*n* = 191), between the ages of 12.04 and 18.02 (*M* = 15.03, *SD* = 1.58) at the start of their treatment episode, and largely diagnosed with depressive (22.1%), disruptive behavior (21.6%), adjustment (10.1%), and attention (8.5%) concerns. Furthermore, although a large portion of the current sample did not disclose information about their ethnic background (*n* = 115; 31.4%), ethnicities reported were: Multiethnic (*n* = 146; 39.9%), Native Hawaiian or Pacific Islander (*n* = 42; 11.5%), Asian (*n* = 37; 10.1%), White (*n* = 25; 6.8%), and Black or African American (*n* = 1; 0.3%). Youth participants' average treatment length for the episode under study was 218.10 days (*SD* = 130.11).

Measures

The Child and Adolescent Functional Assessment Scale (CAFAS; Hodges, 1998; see Appendix A)

The CAFAS is a widely used and well-validated instrument for measuring levels of functional impairment in youths with emotional, behavioral, and/or substance use difficulties. This measure can be used with youths ages 5-19 years old and can be applied in clinical as well as research settings for both treatment planning and tracking treatment progress. The CAFAS is a 200-item clinician-rated measure that includes eight subscales used to rate the youth. These subscales include, School/Work Role Performance, Home Role Performance, Community Role Performance, Behavior Toward Others, Mood/Emotions, Self-Harmful Behavior, Substance Use, and Thinking. Subscales are rated using a four-level scoring approach (i.e., severe impairment = 30, moderate impairment = 20, mild impairment = 10, minimal or no impairment = 0) and inform the degree of impairment or dysfunction in each domain. Moreover, subscale ratings are summed for a Total Scale score, ranging from 0 to 240, with higher scores indicating higher levels of youth functional impairment (Hodges, 1995).

To note, studies thus far examining the psychometric properties of the CAFAS have suggested good reliability and validity. For instance, Hodges and Wong (1996) found large and positive correlations between criterion- and trainee-rater scores, providing evidence of interrater reliability at the total and subscale levels. Moreover, the authors found fair to good intraclass correlation coefficient reliability estimates within rater samples ranging from .54 to .96, at both the total and subscale levels. To note, however, the same study found questionable levels of internal consistency estimates for CAFAS Total Scale scores at intake, 6-, 12-, and 18-month follow-up ($.63 \leq \alpha \leq .68$). Evidence of this measures convergent validity was also demonstrated through significant and expected correlations between CAFAS Total Scale scores and other instruments measuring broadband psychopathology. Furthermore, youths with high CAFAS Total Scale scores have been found, more likely than youths with lower CAFAS Total Scale scores, to struggle with personal relationships, be physically aggressive towards others, experience thoughts of harming oneself, have involvement with the juvenile justice system, hold negative views towards school, perform poorly in school-related tasks, and engage in truancy and/or other disruptive behaviors that lead to discipline or suspension from school. In addition, CAFAS Total Scale scores have been able to discriminate between clients receiving different levels of care, including inpatient, alternative, or outpatient care (Hodges & Wong, 1996; Jackson et al., 2017).

The Ohio Youth Problem, Functioning, and Satisfaction Scales – Short Form (Ohio Scales Short Form; Ogles et al., 1999; see Appendix B)

The Ohio Scales Short Form is a measure designed to assess treatment outcomes of mental health services for youths ages 5 to 18 years old. Three versions are available (i.e., youth, caregiver, and agency worker forms) to allow data collection from multiple informants and provide a comprehensive overview of a youth's level of functioning and severity of problems as a whole. The Ohio Scales Short Form is relatively brief, cost-effective, contains user-friendly language, and requires minimal training for both scoring and interpretation of results. The instrument consists of four separate domains: (a) Problem Severity, (b) Functioning, (c) Hopefulness, and (d) Satisfaction with Behavioral Health Services, which can be applied independent of one another (Ogles et al., 2004, p. 281). This allows for flexibility, including the option to utilize domains applicable for research and clinical goals. Notably, the current investigation focused solely on the Problem Severity domain youth- and caregiver-report versions. To note, the Problem Severity domain consists of 20 individual items that are identical in language across informant versions. Responses are endorsed based on the frequency and severity of problems experienced by a youth over the past 30 days (prior to the date of administration). Answers are given on a six-point Likert scale, ranging from 0 "Not at all" to 5 "All the time." Problem Severity domain Total Scale scores are calculated by summing all 20 individual item scores, with higher scores suggesting more severe and frequent problems. No items are reverse scored on this measure (Ogles et al., 2000). In the current study, Cronbach's Alpha coefficients were calculated separately for youth- and caregiver-reports. According to George and Mallery (2013), in the current study, caregiver scores fell in the "excellent" range for Total Scale (α = .92) and Externalizing Subscale (α = .91) scores; "good" range for Depression (α = .84) and Anxiety (α = .81) Subscale scores; and "acceptable" range for Delinquency Subscale (α = .78) scores. Additionally, youth scores fell in the "good" range for Total Scale (α = .89) and Externalizing and Depression (α = .85) Subscale scores; and "acceptable" range for Anxiety (α = .77) and Delinquency (α = .76) Subscale scores.

The Problem Severity domain has performed as a reliable and valid measure across an array of studies including youths and caregivers. Foremost, while there have been slightly mixed findings regarding its factor structure, Laba et al. (2022) provided evidence for a four-factor model (i.e., Externalizing, Delinquency, Anxiety, and Depression) using both youth- and caregiver-reports from the existing archival dataset utilized in the current study. Additionally, good to excellent internal consistency estimates for Problem Severity Total Scale scores have been found for both youth- ($.88 \leq \alpha \leq .93$) and caregiver-report ($.88 \leq \alpha \leq .95$) versions (Bonadio & Tompsett, 2016; Dowel & Ogles, 2008; Moore, 2008; Ogles et al., 2000, 2004; Riley et al., 2009; Warnick et al., 2009). Problem Severity Total Scale scores have also demonstrated acceptable one-week test-retest reliability in a combined clinical and comparison youth sample ($r = .72$, $p < .001$; Dowel & Ogles, 2008), as well as acceptable and good two-week test-retest reliability estimates for youth- ($r = .72$, p not reported) and caregiver-report ($r = .88$, p not reported), respectively (Ogles et al., 2004). Problem Severity domain scores have been associated with related indices of youth functional impairment (Laba et al., 2022) and measures used to assess youth behavior problems (Bonadio & Tompsett, 2016). Moreover, the Problem Severity domain has demonstrated the ability to differentiate between clinical and non-clinical samples (Carlston & Ogles, 2006; Dowel & Ogles, 2008), and discriminate between youths with and without pertinent mental health diagnoses (Laba et al., 2022). Lastly, studies have found evidence for the Problem Severity domain's ability to effectively detect treatment effects from pre- to post-treatment (Dowel & Ogles, 2008; Laba et al., 2022).

Procedure

The CAMHD began utilizing the Ohio Scales Short Form (i.e., Problem Severity and Hopefulness domains) with youths and caregivers beginning in 2014. More specifically, youth mental health case managers (who are referred to as "care coordinators" in the CAMHD system of care) are tasked with administering the Ohio Scales Short Form to their caseload youth and/or their caregivers once per month. In addition to administering the Ohio Scales Short Form, CAMHD care coordinators also complete the CAFAS measure for all caseload youth once every 3 months. The current study analyzed data from an existing archival data-limited dataset (Laba et al., 2022), which was electronically extracted from the Child and Adolescent Mental Health Management Information System (CAMHMIS). The current dataset included archival youth- and caregiver-rated Ohio Scales Problem Severity domain and care coordinator-rated CAFAS data collected between January 2014 to December 2018. To note, all policies and procedures for the current study were approved through the University of Hawaiʻi at Mānoa's Institutional Review Board.

Analytic Strategy

Data Integrity

All data were examined in a variety of ways prior to conducting any analyses, including initial evaluation of univariate and bivariate results (Fife, 2020). More specifically, in order to gain a preliminary understanding of the current dataset, distributional properties of the data, such as, normality, standard deviations (*SD*s), skewness, and kurtosis were investigated. Skewness values between -2.0 to 2.0 and kurtosis values between -7.0 and 7.0 were considered “acceptable” (George & Mallery, 2013; Kim, 2013). Ranges of all measures were examined to identify any impossible values within the dataset and outliers were detected by computing and examining standard Problem Severity domain scores (Warner, 2013). Additionally, all data were inspected for missing data and caregiver-youth dyads were excluded from the current study if either informant was missing 85% or more of their data (cf. Laba et al., 2022). To note, a Missing Value Analysis for the two predictor variables (a) youth-reported Problem Severity domain scores (b) caregiver-reported Problem Severity domain scores, and criterion variable (i.e., youth discharge CAFAS scores) was conducted to assess if these data were missing completely at random (MCAR). In particular, statistically non-significant Little’s test (Little, 1988) results indicate that data are MCAR. If missing data are found to be MCAR, multiple imputation approaches including other relevant variables (i.e., youth age, sex, ethnicity/race, and mental health diagnoses) will be utilized to address missing values. If missing data are not found to be MCAR, full information maximum likelihood with auxiliary variables will be utilized to treat data that are not missing at random (Newman, 2014). With regard to multicollinearity for my primary analysis, associations between potential predictor variables were examined to identify any highly correlated values. Regression model variance inflation factors (VIFs) were evaluated as an additional step to assess for multicollinearity. In particular, correlation r values above .80 and VIF values above 10.0 suggests potential issues with multicollinearity (Field, 2009).

Descriptive Statistics

Demographic information, including youth age, sex, ethnicity/race, and mental health diagnoses, were electronically extracted from the CAMHMIS in order to analyze and describe the current sample’s characteristics. In addition, other descriptive statistics, including minimum, maximum, mean, and *SD* values were examined using data gathered with the Problem Severity domain and CAFAS measures.

Preliminary Analysis: Hierarchical Factor Model

Prior to conducting the current study's major aim, confirmatory factor analyses (CFA) in Mplus Version 7.31 (Muthén & Muthén, 2015) were conducted to examine the fit of a hierarchical factor structure of the Problem Severity domain by caregiver- and youth-report. The hierarchical factor structure of the Problem Severity domain was examined to determine if a second-order factor (i.e., Total Scale)[2] existed among four first-order factors (i.e., Externalizing, Delinquency, Anxiety, and Depression) identified in Laba et al. (2022). This preliminary step was completed to inform subsequent analytic strategies. In particular, if the hierarchical factor model of the Problem Severity domain demonstrated good model fit (Hu & Bentler, 1999), the second-order factor (i.e., Problem Severity Total Scale) would be utilized in the current study. However, if instead the hierarchical factor model of the Problem Severity domain demonstrated poor model fit, the four first-order factors (i.e., Externalizing, Delinquency, Anxiety, and Depression Subscales) identified in Laba et al. (2022) would be utilized.

As such, two separate hierarchical factor analyses were conducted to examine the item-to-factor relations of the Problem Severity domain by caregiver- and youth-report. Maximum likelihood estimation with robust standard errors (MLR) was applied, given that this method

[2] To my knowledge, there are currently no studies that have investigated a hierarchical factor model to examine the fit of a second-order factor structure for the Problem Severity domain (i.e., Problem Severity Total Scale broken down into subcomponent scales). However, a multitude of studies have provided evidence of good to excellent internal consistency estimates of Problem Severity Total Scale scores, and as noted above in the Measures section, the Total Scale has been utilized across a variety of studies with youths and their caregivers (Bonadio & Tompsett, 2016; Dowel & Ogles, 2008; Moore, 2008; Ogles et al., 2000, 2004; Riley et al., 2009; Warnick et al., 2009).

yields standard errors that are robust to data deviating from normality. Moreover, factor loadings were assessed to confirm that all items adequately and significantly loaded onto the hypothesized factors. To note, item loadings with z-scores outside the -1.96 to 1.96 range, calculated by [Estimate / Standard Error] at the 95% confidence level ($\alpha = .05$), were considered adequate and significant. Furthermore, while the chi-square (χ^2) test is traditionally considered a gold standard when assessing model fit (Bonadio & Tompsett, 2016), it is important to note that this test statistic is influenced by sample size. Thus, while χ^2 values are reported in Table 2, these results were not formally considered in the current study. Instead, using a holistic judgement approach, four other test statistics were used to examine model fit. Specifically, Comparative Fit Index (CFI; Bentler, 1990) and Tucker-Lewis Index (TLI; Tucker & Lewis, 1973) values close to .90 were indicative of "acceptable" model fit, while values >.95 suggested "excellent" model fit; Root Mean Square Error of Approximation (RMSEA; Steiger, 1990) values <.08 were indicative of "acceptable" mode fit, while values <.06 suggested "excellent" model fit; and Standardized Root Mean Square Residual (SRMR; Steiger, 1990) values <.09 were indicative of "acceptable" model fit, while values <.08 suggested "excellent" model fit (Hu & Bentler, 1999; Mueller & Hancock, 2019). CFA modification indices were also examined to identify any potential sources of model misfit.

Primary Analysis: Polynomial Regression with Response Surface Analysis

Polynomial regression, using the Statistical Package for the Social Sciences v29.0 program, with a subsequent response surface analysis (Edwards & Parry, 1993) was utilized to examine the extent to which caregiver-youth informant discrepancy scores at intake into a mental health system predicted youth functional impairment at treatment discharge. Said another way and as will be described further below, my dependent variable was youth discharge CAFAS scores. As an initial step, bivariate relations between predictor and criterion variables were investigated to inform subsequent analytic strategies. Additionally, available and/or theoretically relevant variables, such as youth age, sex, ethnicity/race, problem area, intake CAFAS score, and treatment length were evaluated as potential covariates in the polynomial regression analysis further described below. Variables were omitted from the analysis if they did not meet statistical significance, which was determined against an alpha of .05.

Polynomial Regression. Polynomial regression utilizes interaction terms to examine if the association between one informant's report (e.g., caregiver-reported Problem Severity scores) and an outcome (e.g., youth discharge CAFAS scores) are moderated by a corresponding informant's report (e.g., youth-reported Problem Severity scores). Specifically, interaction terms indicate the degree to which, within a dyad, a high (or low) informant score is related to a dependent variable when a corresponding informant endorses a high (or low) score. This technique is noteworthy, given that it provides information above and beyond the main effects of one individual's report (cf. Becker-Haimes et al., 2018; Laird & De Los Reyes, 2013).

Following Shanock et al.'s (2010) recommendations, the current study conducted polynomial regression first, followed by subsequent response surface analyses. First, I examined the base rate of informant discrepancy scores for the current sample by standardizing my two predictor variables (i.e., caregiver- and youth-report Problem Severity scores) to identify how many caregiver-youth dyads have scores that are characterized as "discrepant" (i.e., caregiver/youth standardized scores that are half a *SD* above or below the standardized score of their respective caregiver/youth; cf., Shanock et al., 2010). To note, prior to conducting this analysis, Shanock et al. (2010) recommends having approximately 10% or more caregiver-youth dyads with "discrepant" scores to warrant examination of the relation between discrepancies and an outcome. Then, I centered the predictor variables around the midpoint of the Problem Severity domain's scale. This was done to both aid interpretation of results and minimize issues related to potential multicollinearity (Aiken & West, 1991; Edwards, 1994). Next, three additional variables were computed, including the (a) square of the centered caregiver score, (b) cross product of the centered caregiver and youth scores, and (c) square of the centered youth score. To note, the formula outlined below is a general polynomial regression equation that was utilized in the current study: Z represents the dependent variable (i.e., youth discharge CAFAS scores), X represents one predictor variable (i.e., caregiver-reported Problem Severity scores), and Y represents another predictor variable (i.e., youth-reported Problem Severity scores). Following the equation below, I regressed my dependent variable on the two centered predictor variables, square of the centered caregiver score, cross product of the centered caregiver and youth scores, and square of the centered youth score. Significance was determined against an alpha of .05.

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e.$$

Response Surface Analysis. Instead of directly evaluating findings from the polynomial regression analysis, result coefficients are graphed in a three-dimensional space using subsequent response surface analysis (Edwards & Parry, 1993). During this step, I examined the extent to which caregiver-youth dyad scores characterized as having minimal to no discrepancy (i.e., both a caregiver and their corresponding youth similarly had high Problem Severity scores; both a caregiver and their corresponding youth similarly had low Problem Severity scores) compared to large discrepancy (i.e., a caregiver had high Problem Severity scores while their corresponding youth had low Problem Severity scores; a caregiver had low Problem Severity scores while their corresponding youth had high Problem Severity scores) predicted youth functional impairment at treatment discharge. Interaction terms were examined by plotting predicted values and calculating slopes at ±1 *SD* of the moderator (cf. Goolsby et al., 2018; Laird & De Los Reyes, 2013).

To note, results of the response surface analysis are interpreted in two major ways. First, the significance level as well as the direction (i.e., negative or positive) of four “surface test values,” denoted in the literature as a_1, a_2, a_3, and a_4, are evaluated (Shanock et al., 2010). Figure 1 (Rodrigues, 2021, p. 11) provides a helpful description of how to interpret these surface test values and the questions that these values answer. Second, a three-dimensional graph is produced to aid interpretation of results and provide a nuanced perspective of the potential complex relations between predictor variables and a dependent variable, which would have been otherwise minimized using other traditional approaches (e.g., difference scores, moderated regression) that generally explore these relations using a two-dimensional approach (Edwards, 2002; Shanock et al., 2010). For example, as shown in Figure 2, on a response surface graph, the two predictor variables (X and Y) are placed on the bottom two axes, while the dependent variable (Z) is located on the vertical axis. This three-dimensional space allows for interpretation of the response surface pattern as it relates to variety of X and Y combinations and their respective Z values. The response surface analysis also provides a visual representation of a line of congruence (LOC; denoted by a blue solid line in Figure 2) and a line of incongruence (LOIC; denoted by a red dashed line in Figure 2), which are utilized during interpretation of results. According to Rodrigues (2021), the LOC symbolizes all possible combinations where X = Y. When considered with the outcome variable Z, the LOC slope provides information about the extent to which levels of congruence (hereinafter referred to as “agreement”) between X and Y are related to Z. Said another way, and as it relates to the current investigation, the LOC slope demonstrates how varying levels of agreement between caregiver and youth Problem Severity scores are related to youth discharge CAFAS scores. To investigate my first hypothesis, that when both a caregiver and their respective youth similarly reported high levels of problem severity it would be significantly predictive of poorer youth functional impairment at treatment discharge, I examined the aforementioned surface test values a_1 and a_2. Support for my first hypothesis would be established if results indicated a significant and positive a_1 value, and a non-significant a_2 value. Additionally, the LOIC, which is perpendicular to the LOC, symbolizes all possible combinations where $Y = -X$. When considered with the outcome variable Z, the LOIC slope provides information about the extent to which levels of discrepancy (hereinafter referred to as "disagreement") between X and Y are related to Z (Rodrigues, 2021). As it relates to the current study, the LOIC slope demonstrates how varying levels of disagreement between caregiver and youth Problem Severity scores are related youth discharge CAFAS scores. To investigate my second hypothesis, that larger caregiver-youth discrepancy scores at the start of treatment would be significantly predictive of poorer youth functional impairment at treatment discharge, especially when caregivers reported higher levels of problem severity than their respective youth, I examined the surface test values a_3 and a_4. Support for my second hypothesis would be established if results indicated significant and positive a_3 and a_4 values. Thus, and as will be demonstrated below, the response surface analysis allows for a comprehensive overview of the association between a dependent variable and varying levels of both (a) agreement and (b) disagreement between predictor variables.

Results

Data Integrity

As previously mentioned, prior to conducting the current study's major aim, preliminary analysis of the Problem Severity domain was conducted to inform subsequent analytic strategies.

As the reader will see in the next section, analyses were conducted for both the Total Scale and Subscale scores of the Problem Severity domain by caregiver- and youth-report, and as such, data integrity related to both Total Scale and Subscale scores are presented in this section.

Caregiver and youth Problem Severity Total Scale and Subscale skewness and kurtosis values fell within the acceptable range. Moreover, concerning caregiver-report data, there were two statistical outlier scores identified for the Depression Subscale. There were no outliers for Total Scale and all other Subscale scores. Regarding youth-report data, there were two statistical outlier scores identified for the Total Scale, as well as one, eight, zero, and two statistical outlier scores identified for the Externalizing, Delinquency, Anxiety, and Depression subscales, respectively. To note, prior to omitting outliers, Fife (2020) recommends evaluating results with and without the identified outliers in order to determine the extent to which findings change. Thus, after evaluation of the current data, which did not subsequently result in different findings with and without outliers, the identified cases were retained for subsequent analyses.

Additionally, as previously mentioned, caregiver-youth dyads were excluded from the current study if either informant was missing 85% or more of their data. Due to this criteria, one caregiver-youth dyad was excluded, leaving 366 caregivers and 366 youths contributing to their respective analyses. Regarding missing data, and specifically concerning caregiver-report Problem Severity Total Scale values, 19 caregiver participants were missing one item, four were missing two items, two were missing three items, one was missing five items, one was missing 10 items, one was missing 11 items, and one was missing 12 items. With regard to youth-report Problem Severity Total Scale values, 20 youth participants were missing one item, one was missing two items, and two were missing three items. Additionally, 153 youth participants were missing discharge CAFAS Total Scale data. To note, results from a Missing Value Analysis revealed that data were missing completely at random. In particular, Little's MCAR test indicated non-significant results for caregiver-reported Problem Severity Total Scale scores (χ^2 = 387.07, *df* = 376, *p* = .34), youth-reported Problem Severity Total Scale scores (χ^2 = 232.24, *df* = 261, *p* = .90), and youth discharge CAFAS Total Scale scores (χ^2 = 1.47, *df* = 2, *p* = .48). Thus, following recommendations from Newman (2014), multiple imputation approaches including other relevant and available variables (i.e., youth age, sex, ethnicity/race, and mental health diagnoses) were utilized to address missing values. With regard to collinearity diagnostic information, correlation *r* and VIF values fell below the recommended thresholds of .80 and 10.0, respectively; thus, suggesting no likely issues related to multicollinearity between caregiver and youth predictor variables for my primary analysis.

Preliminary Analysis: Hierarchical Factor Model

Prior to conducting the current study's major aim, two CFAs were conducted to examine the fit of a hierarchical factor structure for the Problem Severity domain separately by caregiver- and youth-report. Results of the CFAs broken down by informant report are presented in Table 2.

Caregiver-Report

Fit indices for the second-order factor model included χ^2 (167) = 507.31, RMSEA = .08, CFI = .88, TLI = .86, and SRMR = .10. To note, while RMSEA and CFI values fell in the acceptable range, TLI and SRMR values approached acceptable model fit. Also, while CFA modification indices were initially examined to identify any potential sources of model misfit for the hierarchical factor structure of the Problem Severity domain, modifications only slightly improved model fit and lacked theoretical justification (cf. Mueller & Handcock, 2007). As such, additional modifications to the caregiver-report model did not appear warranted and were not further included in the current study. Furthermore, given the slightly mixed hierarchical factor structure results, an additional CFA involving the four first-order factors (i.e., Externalizing, Delinquency, Anxiety, and Depression) identified in Laba et al. (2022) was conducted. In comparison to the second-order factor model, the four-factor model demonstrated slightly improved, yet similar model fit. Specifically, model fit indices included χ^2 (164) = 423.21, RMSEA = .07, CFI = .91, TLI = .89, and SRMR = .06. Across both CFAs, all items adequately and significantly loaded onto the hypothesized factors. See Tables 3 and 4 for the second-order and four-factor model caregiver factor loading results, respectively.

Youth-Report

Fit indices for the second-order factor model included χ^2 (167) = 438.23, RMSEA = .07, CFI = .87, TLI = .86, and SRMR = .15. To note, while RMSEA fell in the acceptable range, the CFI, TLI, and SRMR values approached acceptable model fit. Similar to the aforementioned caregiver-report CFA, modification indices were examined to identify any potential sources of model misfit for the hierarchical factor structure of the Problem Severity domain by youth-report. However, modifications only slightly improved model fit and also lacked theoretical justification (cf. Mueller & Handcock, 2007). As such, additional modifications to the youth-report model did not appear warranted and were not further included in the current study. Also similar to the aforementioned caregiver-report findings, given slightly mixed hierarchical factor structure results, an additional CFA involving the four first-order factors identified in Laba et al. (2022) was conducted. In comparison to the second-order factor model, the four-factor model demonstrated slightly improved model fit. For instance, model fit indices included χ^2 (164) = 337.98, RMSEA = .05, CFI = .92, TLI = .91, and SRMR = .06. Across both CFAs, all items adequately and significantly loaded onto the hypothesized factors. See Tables 5 and 6 for the second-order and four-factor model youth factor loading results, respectively.

Using a holistic judgment approach, the second-order and four first-order factor models similarly demonstrated overall acceptable model fit across both caregiver- and youth-report. Given these findings as well as the clinical utility of Problem Severity Total Scale scores in the CAMHD system of care, study results from here on out focus on *both* the Total Scale and Subscale (i.e., Externalizing, Delinquency, Anxiety, and Depression) scores of the Problem Severity domain.

Primary Analysis: Polynomial Regression with Response Surface Analysis

Zero-order bivariate Pearson product correlations between predictor and criterion variables are presented in Table 7. Correlation *r* values between caregiver and youth Problem Severity domain scores were significant on the Total Scale (.36, p <.001) as well as the Externalizing (.46, p <.001), Delinquency (.48, p <.001), Anxiety (.28, p <.001), and Depression (.35, p <.001) Subscales. Base rates of informant discrepancy scores for caregiver- and youth-report Problem Severity Total Scale and Subscale scores are provided in Table 8. While caregiver-report was generally higher than youth-report in the overall sample, youth-report was higher on average specifically for Anxiety Subscale scores. Additionally, across Total Scale and Subscale scores, the minimum percent (i.e., approximately 10% or more; Shanock et al., 2010) criteria for caregiver-youth dyads with "discrepant" scores was met. Thus, examining the extent to which caregiver-youth informant discrepancies were related to youth functional impairment at treatment discharge was warranted in the current study. As previously mentioned, instead of directly evaluating findings from the initial polynomial regression analysis (see upper half of Table 9), result coefficients are graphed in a three-dimensional space using response surface analysis (Edwards & Parry, 1993). This subsequent step provided a nuanced perspective of the association between caregiver-youth informant discrepancy scores and youth functional impairment at treatment discharge. The reader is reminded that support for my (a) first hypothesis would be established if results indicated a significant and positive a_1 value, and a non-significant a_2 value, and (b) second hypothesis would be established if results indicated significant and positive a_3 and a_4 values. See the lower half of Table 9 for response surface analysis results broken down by Total Scale and Subscale scores.

Total Scale

Polynomial Regression. Intake CAFAS score, having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders, and treatment length were identified as significant covariates. More specifically, higher discharge CAFAS scores were significantly associated with (a) higher intake CAFAS scores (β = .269, *SE* = .050, *t* = 5.358, *p* <.001), (b) having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders (β = 10.606, *SE* = 4.368, *t* = 2.428, *p* = .016), and (c) shorter treatment length in days (β = -.030, *SE* = .014, *t* = -2.166, *p* = .031).

Response Surface Analysis.

Agreement. Contrary to my first hypothesis, a positive and significant a_2 value (2.21, *p* = .027) pointed to a curvilinear slope with regard to the LOC. Thus, it was found that youth discharge CAFAS Total Scale scores are higher when caregiver and youth scores match at more extreme values compared to when dyads match at midrange values.

Disagreement. Consistent with my second hypothesis, a positive and approaching significance a_3 value (1.90, *p* = .059) suggested that youth discharge CAFAS Total Scale scores are potentially higher when caregivers report higher levels of Problem Severity Total Scale scores compared to their respective youth. A three-dimensional response surface graph, involving caregiver and youth Total Scale scores, is provided in Figure 2.

Externalizing Subscale

Polynomial Regression. Intake CAFAS score, having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders, and treatment length were identified as significant covariates. In particular, higher discharge CAFAS scores were significantly associated with (a) higher intake CAFAS scores (β = .301, *SE* = .049, *t* = 6.130, *p* <.001), (b) having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders (β = 8.965, *SE* = 4.384, *t* = 2.045, *p* = .042), and (c) shorter treatment length in days (β = -.028, *SE* = .014, *t* = -2.026, *p* = .043).

Response Surface Analysis.

Agreement. Contrary to my first hypothesis, no significant associations emerged between levels of caregiver-youth agreement on the Externalizing Subscale and youth discharge CAFAS Total Scale scores.

Disagreement. In line with my second hypothesis, a positive and significant a_3 value (2.30, *p* = .022) indicated that youth discharge CAFAS Total Scale scores are higher when caregivers report higher levels of Externalizing Subscale scores compared to their respective youth. Additionally, and in support of my second hypothesis, a positive and approaching significance a_4 value (1.91, *p* = .058) suggested a potential relation between increased caregiver-youth Externalizing Subscale discrepancy scores and higher youth discharge CAFAS Total Scale scores. A three-dimensional response surface graph, involving caregiver and youth Externalizing Subscale scores, is provided in Figure 3.

Delinquency Subscale

Polynomial Regression. Intake CAFAS score and treatment length were identified as significant covariates. Specifically, higher discharge CAFAS scores were significantly associated with (a) higher intake CAFAS scores (β = .327, *SE* = .051, *t* = 6.367, *p* <.001) and (b) shorter treatment length in days (β = -.025, *SE* = .014, *t* = -1.765, *p* = .078).

Response Surface Analysis.

Agreement. Consistent with my first hypothesis, a positive and significant a_1 value (4.32, *p* < .001) indicated that youth discharge CAFAS Total Scale scores are higher when caregiver and youth Delinquency Subscale scores match at higher versus lower levels.

Disagreement. Contrary to my second hypothesis, however, no significant association emerged between levels of Delinquency Subscale caregiver-youth discrepancy scores and youth discharge CAFAS Total Scale scores. A three-dimensional response surface graph, involving caregiver and youth Delinquency Subscale scores, is provided in Figure 4.

Anxiety Subscale

Polynomial Regression. Intake CAFAS score, having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders, and treatment length were identified as significant covariates. Notably, higher discharge CAFAS scores were significantly associated with (a) higher intake CAFAS scores (β = .363, *SE* = .051, *t* = 7.139, *p* <.001), (b) having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders (β = 14.058, *SE* = 4.634, *t* = 3.034, *p* = .003), and (c) shorter treatment length in days (β = -.038, *SE* = .015, *t* = -2.619, *p* = .009).

Response Surface Analysis.

Agreement. In line with my first hypothesis, a positive and significant a_1 value (2.25, *p* = .025) indicated that youth discharge CAFAS Total Scale scores are higher when caregiver and youth Anxiety Subscale scores match at higher versus lower levels.

Disagreement. Contrary to my second hypothesis, however, no significant relation emerged between levels of Anxiety Subscale caregiver-youth discrepancy scores and youth discharge CAFAS Total Scale scores. A three-dimensional response surface graph, involving caregiver and youth Anxiety Subscale scores, is provided in Figure 5.

Depression Subscale

Polynomial Regression. Intake CAFAS score, having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders, and treatment length were identified as significant covariates. In particular, higher discharge CAFAS scores were significantly associated with (a) higher intake CAFAS scores (β = .350, *SE* = .052, *t* = 6.715, *p* <.001), (b) having a principal diagnosis related to Disruptive, Impulse Control, and Conduct Disorders (β = 15.989, *SE* = 4.618, *t* = 3.462, *p* <.001), and (c) shorter treatment length in days (β = -.038, *SE* = .014, *t* = -2.660, *p* = .008).

Response Surface Analysis.

Agreement. Consistent with my first hypothesis, a positive and significant a_1 value (2.18, *p* = .030) indicated that youth discharge CAFAS Total Scale scores are higher when caregiver and youth Depression Subscale scores match at higher versus lower values.

Disagreement. As predicted in my second hypothesis, a positive and significant a_3 value (2.56, *p* = .011) indicated that youth discharge CAFAS Total Scale scores are higher when caregivers report higher levels of Depression Subscale scores compared to their respective youth. A three-dimensional response surface graph, involving caregiver and youth Depression Subscale scores, is provided in Figure 6.

Discussion

Summary of Findings

This is the first study to thoroughly examine the association between caregiver-youth informant discrepancy scores at treatment intake and youth functional impairment at discharge using a large public mental health sample and newly-recommended statistical analyses. With regard to my preliminary analyses, I found similar results for the second-order (i.e., Total Scale) and four first-order factors (i.e., Externalizing, Delinquency, Anxiety, and Depression Subscales) of the Problem Severity domain by caregiver- and youth-report. More specifically, all CFA model fit indices fell at or near the acceptable model fit range, with the four first-order factors demonstrating slightly better model fit in comparison to the second-order factor. Given these results, as well as the clinical utility of the Problem Severity domain Total Scale scores in the CAMHD system of care, both Total Scale and Subscale scores were utilized in subsequent primary analyses.

The primary aim of the current study examined the extent to which caregiver-youth informant discrepancy scores at intake into a mental health system predicted youth functional impairment at treatment discharge. Related to this primary aim, it is noteworthy to mention that in the current sample, correlation *r* values between caregiver- and youth-reports were positive and relatively larger in comparison to previous seminal investigations looking at relations between various informant scores (Achenbach et al., 1987; De Los Reyes et al., 2015). In particular, for caregiver-youth dyads, Achenbach et al. (1987) found a mean correlation *r* value of .25 for combined problem areas, and De Los Reyes et al.'s (2015) results suggested average correlation *r* values of .26 for internalizing and .32 for externalizing concerns. In the present study, however, correlation *r* values were .36 for combined problem areas, and ranged from .28

to .35 for internalizing and .46 to .48 for externalizing concerns. This may likely, in part, be due to the utilization of measures with identical item content across reporter versions (Carlston & Ogles, 2006), as most studies to date have included youth- and caregiver-report measurement tools that slightly vary across reporter versions (e.g., comparing CBCL and YSR [Achenbach & Rescorla, 2001] or BASC caregiver and youth-report scores [Reynolds & Kamphaus, 2015]).

Additionally, in the overall sample, caregivers generally reported higher levels of problem severity compared to youth, with the exception of Anxiety Subscale scores. This overall finding is consistent with other studies, which found that caregivers on average reported higher levels of mental health concerns compared to their respective youth across a variety of clinical settings. This includes for example, samples from public mental health, outpatient, inpatient, and clinical trial settings (Becker-Haimes et al., 2018; Carlston & Ogles, 2009; Humphreys et al., 2017; Kolko & Kazdin, 1993). Overall, while there were slightly mixed findings across Total Scale and Subscale primary analyses, the results described below generally supported an association between poor outcomes and (a) caregiver-youth agreement at high levels (i.e., high caregiver and high youth Problem Severity domain scores; my first hypothesis) and (b) higher caregiver-reports of problem severity compared to their respective youth (my second hypothesis).

Hypothesis One, Agreement: Response Surface Values a_1 and a_2

Consistent with my first hypothesis and past research (Goolsby et al., 2018; Makol et al., 2019), a significant and positive response surface analysis a_1 value pointed to higher levels of youth functional impairment at treatment discharge (i.e., higher discharge CAFAS Total Scale scores) when caregiver and youth problem severity scores matched (agreed) at higher versus lower values on the Delinquency, Anxiety, and Depression Subscales. It is important to highlight that research looking at this specific association (i.e., relation between levels of informant agreement and youth outcomes) is sparse, with most investigations focusing their efforts on levels of informant *disagreement*. Specifically, it seems that there are only two studies investigating this particular association, of which notably differ in key aspects when compared to the current investigation (e.g., different sample demographics, level of care, analytic approach). For instance, Goolsby et al.'s (2018) study took place in a large private practice in Washington, D.C. providing manualized group psychotherapy focused on the development of social skills to primarily White identified youth. Makol et al.'s (2019) study involved majority Black and White identified youth receiving inpatient psychiatric care following predominantly acute crises. Moreover, unlike the current investigation, their study focused specifically on caregiver- and youth-reports for only internalizing concerns (Makol et al., 2019). With regard to variation in analytic approach, although polynomial regression was utilized in one of these studies (i.e., Goolsby et al., 2018), their analytic strategy did not involve subsequent response surface analysis. The present study, however, involved an ethnically-diverse public mental health sample receiving services in a variety of forms (e.g., ranging from case management to community- and residential-based care) and utilized polynomial regression with subsequent response surface analysis. Notably, despite these key differences, the present findings related to the Delinquency, Anxiety, and Depression Subscales were consistent with overall results from the two aforementioned studies; thus, potentially providing support for a rather robust association between poor outcomes and caregiver-youth informant scores that match at high versus low values.

Further expanding on findings for my first hypothesis; while response surface analysis a_1 results regarding the Total Scale and Externalizing Subscale scores were inconsistent with my predictions (i.e., non-significant a_1 values), some care should be taken when interpreting the entire constellation of these results. On the one hand, findings involving the Total Scale scores may be questionable more broadly, given the slightly mediocre results from my preliminary hierarchical CFA, which suggested that (a) two out of four and (b) three out of four fit indices fell outside, although near the acceptable model fit range for caregiver- and youth-reports, respectively. Moreover, the reader is reminded that the model fit for the second-order (i.e., Total Scale) factor was slightly worse in comparison to the four first-order factors. At the same time, however, the less than desirable hierarchical CFA results may not fully explain the lack of significant and positive response surface a_1 values for the Externalizing Subscale. Instead, perhaps there is something unique regarding Externalizing Subscale scores, which inherently contributes to Total Scale scores.

For instance, results from my initial polynomial regression analysis could help to illuminate qualities of caregiver and youth scores that likely contributed to the non-significant response surface analysis a_1 findings on the Externalizing Subscale. As shown in Table 9, caregiver- (β = 1.495, *SE* = .520, *t* = 2.874) and youth-reports (β = -.635, *SE* = .604, *t* = -1.052) are opposite in direction on the Externalizing Subscale. This suggests that caregiver- and youth-reports regarding externalizing concerns are not symmetrical (not congruent) in the current sample. Consistent with the ABC Model framework (De Los Reyes & Kazdin, 2005), it seems that caregivers generally held perspectives that youth externalizing concerns were present, while youths instead held the opposite perspective (i.e., that the identified behavioral difficulties were not present). As such, this non-significant result may likely be attributable to a lack of overall agreement between caregiver and youth scores regarding externalizing concerns. This lack of agreement (i.e., non-significant findings along the LOC) in turn indicates that strong disagreement is likely occurring between caregiver and youth Externalizing Subscale scores (K. Hayashi, personal communication, June 10, 2023). As such, one would expect significant findings along the LOIC. Indeed, and as will be expanded upon below, a significant response surface analysis a_3 value was found for the Externalizing Subscale scores, suggesting that the direction of caregiver and youth discrepancy (*disagreement*) scores were significantly related to youth functional impairment at treatment discharge.

Related to the explanation above, the reader will also notice that just like the Externalizing Subscale results, caregiver- and youth-reports are opposite in direction on the Depression Subscale. However, unlike the non-significant a_1 finding for the Externalizing Subscale, the Depression Subscale achieved statistical significance for a_1. Based off of the logic described in the paragraph above, one might initially expect the Depression Subscale results to be statistically non-significant as well. However, further examination of the initial polynomial regression results (see Table 9) helps to illuminate why this was the case for the Externalizing Subscale, but not the Depression Subscale. In particular, the reader will see that for the Depression Subscale, the caregiver-report beta value (β = 2.028, *SE* = .629, *t* = 3.226) is noticeably larger than youth-report (β = -.399, *SE* = .578, *t* = -.690), and also that the magnitude of this difference is much larger when compared to all other Total Scale and Subscale polynomial regression findings (see Table 9). This suggests a dominant caregiver effect for Depression Subscale scores. Said another way, high caregiver reports of depressive concerns appear to be a strong predictor of youth discharge CAFAS scores. This aligns with the significant response surface analysis a_1 and a_3 findings for the Depression Subscale. That is, high caregiver-report was related to poorer outcome when youth-report was similarly high, as well as lower than caregiver-report (K. Hayashi, personal communication, June 10, 2023).

Also with regard to my first hypothesis and contrary to expectation, a significant and positive response surface analysis a_2 value indicated a curvilinear slope (Rodrigues, 2021), which pointed to higher levels of youth functional impairment when caregiver and youth scores matched at more extreme (i.e., high caregiver- and high youth-report, *and* low caregiver- and low youth-report) versus midrange values on the Total Scale. While, as previously mentioned, results from the Total Scale should be interpreted with some caution, it is interesting that caregiver-youth scores that matched at more extreme values, particularly low caregiver- and low youth-report, were related to worse outcomes. One might intuitively suspect that agreement at low levels would be related to better outcomes given that both a caregiver and their respective youth are reporting low levels of problem severity. It is possible, however, that while agreement at high values (i.e., high caregiver- and high youth-report) at the Total Scale level could reflect overall significant problems in a youth's life, agreement at low values (i.e., low caregiver- and low youth-report) might instead reflect a lack of awareness or belief from both informants that an emotional or behavioral difficulty warrants treatment.

Indeed, youths and families within the CAMHD system of care are sometimes referred for services by someone other than the youth or caregiver (e.g., child welfare services, juvenile justice system, other Department of Health divisions; CAMHD, 2018). Related to De Los Reyes and Kazdin's (2005) ABC model, if informant attributions and perspectives about a behavioral problem vary at the start of therapy (including situations where both a caregiver and youth report that there are no problems to address, while a referral source believes there are), there will most likely be inherent differences in how much a treatment plan does or does not align with the beliefs of the youth and/or their caregivers. These initial differences could then interfere with treatment participation and subsequently, youth and/or caregiver responses to intervention (De

Los Reyes & Kazdin, 2005). However, future research is warranted in this area, as the current database did not include information on the extent to which youths were or were not required to participant in treatment.

Hypothesis Two, Disagreement: Response Surface Values a_3 and a_4

Consistent with predictions for my secondary hypothesis, a significant and positive response surface analysis a_3 value suggested increased youth functional impairment at treatment discharge when caregivers reported higher levels of problem severity compared to their respective youth on the Externalizing and Depression Subscales. Also consistent with predictions, although only approaching statistical significance (i.e., $p = .059$), this relation nearly emerged for Total Scale scores. Regarding previous research in this area, while there have been some mixed findings regarding the directionality of informant discrepancy scores, most studies have suggested poorer outcomes when caregivers in particular endorsed higher levels of problem severity compared to their respective youth. Moreover, this association has been demonstrated across a variety of problem areas (e.g., both internalizing and externalizing difficulties) and settings, including for example outpatient, private practice, university clinic, clinical trial, and non-clinical areas (Becker-Haimes et al., 2018; Ferdinand et al., 2004, 2006; Goolsby et al., 2018; Kendall et al., 1997). Although not conclusive, these findings are also in line with De Los Reyes and Kazdin's (2005) ABC model, which posits that non-youth informants (e.g., caregivers) are more likely to make attributions that point to inherent or dispositional causes for a behavior problem (i.e., the problematic behavior is due to the disposition of the youth), while youths however, are more likely to make attributions that instead point to environmental causes for a behavior problem. These reporter differences frequently manifest within the context of youth mental health services, where caregivers often report higher levels of problem severity compared to their respective youths (Becker-Haimes et al., 2018; Carlston & Ogles, 2009; Humphreys et al., 2017; Kolko & Kazdin, 1993) and frequently are the informant to initiate mental health treatment services (Hunsley & Lee, 2014; Kazdin, 1989, 2000).

With regard to further discussion of my secondary hypothesis findings; while significant response surface analysis a_3 values were not found specifically for the Delinquency and Anxiety Subscale scores, results from the initial polynomial regression analysis could also provide clues as to the qualities of caregiver and youth scores that likely contributed to these non-significant findings. In particular, the reader is reminded that caregiver- and youth-reports that are opposite in direction suggests that those reports are not symmetrical (not congruent). Likewise, and as it relates to the non-significant a_3 findings on the Delinquency and Anxiety Subscales, caregiver- and youth-reports that are in the same direction suggests that those reports are symmetrical (are congruent). As shown in Table 9, caregiver- and youth-reports are congruent on both the Delinquency and Anxiety Subscales. As such, the non-significant a_3 results on the Delinquency and Anxiety Subscales are likely attributable to a lack of overall disagreement between caregiver and youth scores regarding delinquency- and anxiety-related concerns. This lack of disagreement (i.e., non-significant findings along the LOIC) in turn indicates that strong agreement is likely occurring between caregiver and youth Delinquency and Anxiety Subscale scores in the current sample (K. Hayashi, personal communication, June 10, 2023). As such, one would expect (a) significant findings along the LOC as well as (b) non-significant findings along the LOIC. Consistent with this point, the previously described results noted significant findings along the LOC (i.e., statistically significant response surface analysis a_1 values) for both Delinquency and Anxiety Subscale scores. Also consistent with the aforementioned point, is the statistically non-significant findings along the LOIC for both Delinquency and Anxiety Subscale scores.

Finally, contrary to my secondary hypothesis, only one response surface analysis a_4 value indicated a trend towards an association between higher levels of youth functional impairment at discharge and increased levels of disagreement between caregiver and youth scores at intake. In particular, this relation only approached statistical significance (p = .058) for Externalizing Subscale scores, and all other Total Scale and Subscale response surface analysis a_4 results were statistically non-significant. Interestingly, the reader will also notice that compared to the other scales, the Delinquency and Anxiety Subscales demonstrated larger quadratic and interaction term *SE*s, despite these two Subscales having a smaller range of summed scores as a result of their smaller number of items (see Table 9). Larger quadratic and interaction term *SE*s suggest that the estimates of these terms are less reliable. In typical regression analysis, when such terms are non-significant, they are generally removed from the model while retaining only the linear terms (K. Hayashi, personal communication, June 10, 2023). Thus, as it relates to the current investigation, it is likely that Delinquency and Anxiety Subscale scores can only be expressed as linear terms of caregiver- and youth-report. Taken together, the statistically significant youth-report on the Delinquency Subscale (β = 2.901, SE = .804, t = 3.608, p <.001) suggests that youth-report of delinquent behaviors is a strong predictor of outcomes. Likewise, the statistically significant caregiver-report on the Anxiety Subscale (β = 1.839, SE = .778, t = 2.365, p = .019) indicates that caregiver-report of anxiety concerns is a strong predictor of youth functional impairment at treatment discharge in the current study.

As a whole, there seems to be a lack of strong evidence for an association between increased levels of general informant discrepancy (i.e., without consideration of informant discrepancy direction) and poor youth functional impairment at treatment discharge. While this did not comport with my second hypothesis, it is important to note that these results are somewhat consistent with recent research pointing to mixed findings regarding the general relation (i.e., without consideration of informant discrepancy direction) between baseline informant discrepancies and subsequent youth outcomes or responses to intervention (Becker-Haimes et al., 2018; Carlston, 2004; De Los Reyes et al., 2010; Goolsby et al., 2018; Humphreys et al., 2017). Notably, Becker-Haimes et al.'s (2018) primary results initially suggested no association between levels of pre-treatment caregiver-youth informant discrepancy scores and youth treatment outcomes. However, in their secondary analyses, the authors found that decreased youth-caregiver discrepancy scores over the course of treatment, specifically in the previously described Parent Reported High group (i.e., dyads where caregivers reported the same or higher levels of symptom severity compared to their respective youth), were related to better youth outcomes post-treatment (i.e., improvement in severity of anxiety concerns, youth functioning, response to treatment, and diagnostic remission status; Becker-Haimes et al., 2018). Indeed, this is a promising area for future research.

Limitations

Despite a variety of notable strengths of the current investigation, a few caveats are in order. First, my sample was limited to CAMHD youth with data available during the intake and discharge study time periods. While the current sample had characteristics similar to other large service sector studies involving CAMHD youth (e.g., McLennan et al., 2022); as compared to youth typically enrolled for CAMHD services more generally (CAMHD, 2019), my sample had a higher preponderance of females and youth diagnosed with depressive concerns. While this may limit the generalizability of the current findings to the larger CAMHD system of care, more female representation and youth with depressive concerns potentially align better with other service settings. Furthermore, the Ohio Scales youth form was created for youth between the ages of 12 to 18, and as such, the present study included participants who met this criterion. Related to this point then, it is important to note that the current study's findings with regard to informant discrepancies and youth functional impairment at treatment discharge may be restricted to youth in this age range. Despite these limitations, however, the current study's unique demographic background is a potential strength in comparison to other research in this area, given that existing studies have mainly included less diverse populations, such as youths identified for example as White and struggling specifically with anxiety (Becker-Haimes et al., 2018) or trauma- and stressor-related disorders (Humphreys et al., 2017).

Secondly, as previously shared above, some of the hierarchical CFA model fit indices fell outside of, although near acceptable range benchmarks. While this circumstance warrants careful consideration within the context of interpreting study results, it is also important to note that Problem Severity domain Total Scale score internal reliabilities fell in the good and excellent ranges for youth- and caregiver-report, respectively. Furthermore, Problem Severity domain Total Scale scores are routinely utilized for clinical decision making and outcome monitoring for youths served through the CAMHD system of care. Thus, although potentially far from ideal, there seemed to be merit in the approach utilized in the current study whereby all primary analyses were conducted with both Total and Subscale scores.

Third, the CAFAS measure (utilized as the outcome variable in the current study) was completed by care coordinators, and as such, confounds both reporter and construct with regard to interpretation of study results. That is, interrater discrepancy scores at intake were examined for the two informant types of youth and caregiver, and for the constructs of overall youth mental health problems as well as depression, anxiety, delinquency, and externalizing difficulties. Outcomes, however, were assessed by a care coordinator-reported instrument, and for an entirely different construct. Thus, a cleaner analytic design would have been to include additional treatment progress indicators, including for example, more objective measures of treatment success as well as outcomes as reported by caregivers and/or youth participants. Fortunately, previous investigations have provided some support for the CAFAS as a valid measure of treatment progress (Jackson et al., 2017) and it is also routinely utilized as a helpful decision-making tool in the CAMHD system of care, thus supporting its use in the present investigation.

A fourth potential limitation is related to any unknown factors that might have impacted participant data completion. For example, while CAMHD care coordinators are tasked with the administration of the Problem Severity domain on a monthly basis to their caseload youth and/or caregivers, data collection efforts of the Ohio Scales have reportedly varied over time, and within and across various counties for the state (Milette-Winfree et al., 2019). Administration methods are non-standardized in a variety of regards and occur in different settings (e.g., in the community, at a youth's home, in residential-based settings) and formats (e.g., in-person, over the telephone). Thus, the extent to which such variability with data collection (or the effects from any other unavailable variables) may have influenced the current findings remain unknown. Altogether, the study limitations mentioned above may be considered within the context of ideas for future research.

Implications and Future Directions

The present investigation provides some evidence that caregiver-youth informant discrepancy scores at intake into a public mental health system are predictive of youth functional impairment at treatment discharge. Related to this finding, certain patterns of caregiver and youth discrepancy scores at intake can potentially help to inform discussions involved in treatment planning. In particular, the overall findings regarding (a) caregiver-youth agreement at high levels (i.e., higher functional impairment at treatment discharge when caregivers and youths similarly report high levels of problem severity at intake) and (b) the directionality of informant discrepancy scores (i.e., worse functional impairment at treatment discharge when caregivers report higher levels of problem severity compared to their respective youth at intake) are related to De Los Reyes et al.'s (2022) recent introduction of the Needs-to-Goals Gap framework. The Needs-to-Goals Gap framework serves as an extension of the previously described OTM (De Los Reyes et al., 2013). It includes two main components, namely, (a) a needs assessment followed by (b) collaborative goal setting with the youth, family, and treatment team to individually tailor services to fit the needs of each client's unique emotional/behavioral contingencies and circumstances. These two components can be used to identify the focus of a youth's treatment and initiate a process of shared decision-making between individuals in the youth's ecology. Notably, this process also encourages taking proactive steps to identify and address potential discrepant perspectives between informants ahead of time. As such, this framework promotes using informant discrepancies as a helpful tool to inform, rather than serve as a barrier, to the delivery of effective youth mental health services (De Los Reyes et al., 2022). Given this study's specific findings, consideration of informant discrepancies at treatment intake may be especially considered in circumstances in which caregivers and youths demonstrate informant agreement at high levels (i.e., high caregiver- and high youth-report of problem severity) and caregiver-reports of problem severity are higher compared to their respective youth, as youths in these dyads may be at higher risk for poorer functional impairment at treatment discharge.

Furthermore, given the paucity of caregiver-youth informant discrepancy research utilizing polynomial regression with subsequent response surface analysis, it would be beneficial to include this approach in future studies, especially within the context of potentially complex and messy datasets derived within public mental health settings. As demonstrated in the present study, this technique is noteworthy, given that it provides information above and beyond the main effects of one individual's report (cf. Becker-Haimes et al., 2018; Laird & De Los Reyes, 2013). As such, future studies may benefit from leveraging this technique's unique ability to simultaneously examine patterns between a dependent variable and levels of *agreement* and *disagreement* between independent variables.

Additionally, future studies may aim to examine the relation between caregiver-youth informant discrepancy scores and outcomes using other types of predictors/measurement instruments and outcome indicators (e.g., youth clinical discharge status, treatment progress ratings, return to services, decreases in problem severity, youth and/or caregiver treatment engagement). Forthcoming research may also choose to expand upon the current literature by using the latest statistical approach recommendations to examine other pairs of informants important in youth ecologies (Hunsley & Mash, 2007). For instance, and relative to the CAMHD system of care, three versions of the Ohio Scales are available (i.e., youth, caregiver, and agency worker forms), which are identical in item content across reporter versions. Thus, future studies in CAMHD and other large service sector settings (e.g., schools, other public mental health systems) could examine caregiver-caregiver, caregiver-agency worker, and youth-agency worker reporter pairs.

Future research in this area might also wish to investigate the association between longitudinal changes in informant discrepancy scores and youth treatment outcomes. Indeed, some researchers have already begun to look at this relation, although note some limitations to their study analytic strategies, including use of raw difference scores (Becker-Haimes et al., 2018). Despite this limitation, there are clinical implications around understanding how changes in informant discrepancy scores over the course of treatment are related to youth outcomes. This includes efforts around ongoing treatment progress monitoring and measurement feedback systems, especially within large systems of care. For instance, consideration of informant discrepancy scores at different points throughout a youth's episode of care (e.g., on a quarterly or monthly basis) may help to determine at what point during a treatment episode is it important to focus efforts on increasing levels of concordance between caregiver and youth perspectives to improve outcomes. Given the present investigation's results, this type of discussion may be particularly important for dyads characteristic of high caregiver and low youth report of problem severity.

Despite these limitations and need for future research, the present findings contribute to the current literature by providing some support for the association between initial caregiver-youth informant discrepancy scores and youth functional impairment at discharge. Furthermore, the present study builds upon work in this area by utilizing the latest statistical approach recommendations (Laird & De Los Reyes, 2013), measurement schemes that include identical item content across both youth- and caregiver-report versions, and a large ethnically-diverse public mental health sample of youths and their caregivers. The current findings may be particularly salient for ongoing treatment planning and progress monitoring, especially as it relates to caregiver and youth perspectives about youth emotional and behavioral concerns in large public mental health settings.

Table 1

Demographic Information for Youth Study Participants

Variable	*n*	Percentage
Age		
Minimum: 12.04	–	–
Maximum: 18.02	–	–
Mean (*SD*): 15.03 (1.58)	–	–
Gender		
Female	191	52.2
Male	175	47.8
Ethnicity		
Multiethnic	146	39.9
Native Hawaiian or other Pacific Islander	42	11.5
Asian	37	10.1
White	25	6.8
Black or African American	1	0.3
Unreported or refused to report	115	31.4
Treatment Length in Days		
Mean (*SD*): 218.10 (130.11)	–	–
Intake CAFAS Total Scale Score		
Mean (*SD*): 101.48 (37.98)	–	–
DSM-5 Principal Diagnostic Categories		
Depressive Disorders	81	22.1
Disruptive, Impulse-Control, and Conduct Disorders	79	21.6
Adjustment Disorder	37	10.1
Attention-Deficit/Hyperactivity Disorder	31	8.5
Posttraumatic Stress Disorder	27	7.4
Anxiety Disorders	14	3.8
Bipolar and Related Disorders	5	1.4
Substance-Related and Addictive Disorders	4	1.1
Other Trauma- and Stressor Related Disorders	3	0.8
Intellectual Disabilities	2	0.5
Schizophrenia Spectrum and Other Psychotic Disorders	1	0.3
Autism Spectrum Disorder	1	0.3
General Medical Conditions or Codes No Longer Used	17	4.6
None	64	17.5

Note. Youth principal diagnoses were categorized using the superordinate DSM-5 classification system. However, while the current study's database identified at least one principal diagnosis, it did not provide information about any additional principal diagnoses. As such, youth with more than one principal diagnosis (e.g., co-principal diagnostic conditions) are likely under identified.

Table 2

Fit Statistics for Confirmatory Factor Analytic Problem Severity Domain

Model	χ^2 (*df*)	RMSEA	CFI	TLI	SRMR
Caregiver-Report (*n* = 366)					
Second-order	507.31 (167)*	.08	.88	.86	.10
Four-factor	423.21 (164)*	.07	.91	.89	.06
Youth-Report (*n* = 366)					
Second-order	438.23 (167)*	.07	.87	.86	.15
Four-factor	337.98 (164)*	.05	.92	.91	.06

Note. Second-order = Total Scale broken down into four Subscales (i.e., Externalizing, Delinquency, Anxiety, and Depression) identified in Laba et al. (2022); Four-factor = four first-order factors (i.e., Externalizing, Delinquency, Anxiety, and Depression Subscales); RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; Standardized Root Mean Square Residual = SRMR.
*$p < .05$

Table 3

Caregiver-Report Factor Loadings for the Confirmatory Factor Analysis of the Short Form, Problem Severity Domain by Subscale (Hierarchical/Second-Order Factor Model)

Item #		Estimate (*S.E.*)	*z*
	Externalizing		
3	Yelling, swearing, or screaming at others	0.88 (0.02)	46.48
5	Refusing to do things teachers or parents ask	0.84 (0.02)	44.05
1	Arguing with others	0.85 (0.02)	43.72
4	Fits of anger	0.86 (0.02)	36.05
2	Getting into fights	0.74 (0.03)	25.53
10	Lying	0.74 (0.03)	23.20
6	Causing trouble for no reason	0.75 (0.03)	22.77
11	Can't seem to sit still, having too much energy	0.57 (0.05)	12.38
	Delinquency		
7	Using drugs or alcohol	0.74 (0.05)	14.55
8	Breaking rules or breaking the law (out past curfew, stealing)	0.77 (0.05)	14.15
9	Skipping school or classes	0.73 (0.05)	13.77
	Anxiety		
16	Feeling anxious or fearful	0.90 (0.02)	37.64
17	Worrying that something bad is going to happen	0.86 (0.03)	31.54
19	Nightmares	0.58 (0.06)	10.23
	Depression		
14	Feeling worthless or useless	0.86 (0.02)	35.67
18	Feeling sad or depressed	0.81 (0.03)	30.28
15	Feeling lonely and having no friends	0.76 (0.04)	19.87
13	Talking or thinking about death	0.73 (0.04)	17.82
12	Hurting self (cutting or scratching self, taking pills)	0.60 (0.05)	12.03
20	Eating problems	0.47 (0.06)	8.09

Note. S.E. = Standard Error

Table 4

Caregiver-Report Factor Loadings for the Confirmatory Factor Analysis of the Short Form, Problem Severity Domain by Subscale (Four-Factor Model)

Item #		Estimate (*S.E.*)	*z*
	Externalizing		
5	Refusing to do things teachers or parents ask	0.82 (0.02)	40.97
3	Yelling, swearing, or screaming at others	0.85 (0.02)	38.68
1	Arguing with others	0.81 (0.03)	30.85
4	Fits of anger	0.82 (0.03)	28.98
2	Getting into fights	0.72 (0.03)	23.38
10	Lying	0.74 (0.03)	22.38
6	Causing trouble for no reason	0.74 (0.03)	21.65
11	Can't seem to sit still, having too much energy	0.55 (0.05)	11.96
	Delinquency		
8	Breaking rules or breaking the law (out past curfew, stealing)	0.89 (0.03)	26.43
7	Using drugs or alcohol	0.64 (0.06)	10.84
9	Skipping school or classes	0.63 (0.06)	10.18
	Anxiety		
16	Feeling anxious or fearful	0.89 (0.03)	34.26
17	Worrying that something bad is going to happen	0.85 (0.03)	31.63
19	Nightmares	0.57 (0.06)	10.21
	Depression		
14	Feeling worthless or useless	0.85 (0.03)	32.40
18	Feeling sad or depressed	0.80 (0.03)	29.99
15	Feeling lonely and having no friends	0.75 (0.04)	19.15
13	Talking or thinking about death	0.72 (0.04)	17.28
12	Hurting self (cutting or scratching self, taking pills)	0.58 (0.05)	11.25
20	Eating problems	0.46 (0.06)	7.91

Note. S.E. = Standard Error

Table 5

Youth-Report Factor Loadings for the Confirmatory Factor Analysis of the Short Form, Problem Severity Domain by Subscale (Hierarchical/Second-Order Factor Model)

Item #		Estimate (*S.E.*)	*z*
	Externalizing		
1	Arguing with others	0.84 (0.02)	49.18
3	Yelling, swearing, or screaming at others	0.81 (0.03)	27.19
5	Refusing to do things teachers or parents ask	0.79 (0.03)	27.08
4	Fits of anger	0.82 (0.03)	26.97
2	Getting into fights	0.70 (0.04)	17.46
6	Causing trouble for no reason	0.68 (0.04)	16.92
10	Lying	0.68 (0.04)	15.58
11	Can't seem to sit still, having too much energy	0.57 (0.05)	11.55
	Delinquency		
8	Breaking rules or breaking the law (out past curfew, stealing)	0.79 (0.05)	15.08
7	Using drugs or alcohol	0.74 (0.06)	13.50
9	Skipping school or classes	0.66 (0.06)	10.71
	Anxiety		
16	Feeling anxious or fearful	0.84 (0.03)	27.67
17	Worrying that something bad is going to happen	0.83 (0.03)	27.26
19	Nightmares	0.65 (0.05)	13.74
	Depression		
14	Feeling worthless or useless	0.90 (0.02)	48.23
18	Feeling sad or depressed	0.85 (0.03)	28.14
13	Talking or thinking about death	0.78 (0.04)	21.23
15	Feeling lonely and having no friends	0.75 (0.04)	19.11
20	Eating problems	0.57 (0.05)	11.13
12	Hurting self (cutting or scratching self, taking pills)	0.53 (0.06)	9.28

Note. S.E. = Standard Error

Table 6

Youth-Report Factor Loadings for the Confirmatory Factor Analysis of the Short Form, Problem Severity Domain by Subscale (Four-Factor Model)

Item #		Estimate (*S.E.*)	*z*
	Externalizing		
5	Refusing to do things teachers or parents ask	0.74 (0.03)	22.05
4	Fits of anger	0.75 (0.04)	19.62
1	Arguing with others	0.71 (0.04)	18.73
3	Yelling, swearing, or screaming at others	0.74 (0.04)	18.13
6	Causing trouble for no reason	0.63 (0.04)	14.66
2	Getting into fights	0.62 (0.04)	14.57
10	Lying	0.62 (0.05)	13.09
11	Can’t seem to sit still, having too much energy	0.47 (0.05)	9.45
	Delinquency		
8	Breaking rules or breaking the law (out past curfew, stealing)	0.82 (0.04)	18.74
7	Using drugs or alcohol	0.70 (0.05)	12.81
9	Skipping school or classes	0.66 (0.06)	11.06
	Anxiety		
16	Feeling anxious or fearful	0.81 (0.03)	23.69
17	Worrying that something bad is going to happen	0.80 (0.03)	23.32
19	Nightmares	0.61 (0.05)	12.67
	Depression		
14	Feeling worthless or useless	0.87 (0.02)	40.29
18	Feeling sad or depressed	0.82 (0.03)	26.13
13	Talking or thinking about death	0.74 (0.04)	18.15
15	Feeling lonely and having no friends	0.71 (0.04)	17.27
20	Eating problems	0.52 (0.05)	10.57
12	Hurting self (cutting or scratching self, taking pills)	0.48 (0.06)	8.56

Note. S.E. = Standard Error

Table 7

Problem Severity Domain and CAFAS Bivariate Correlations

	1	2	3	4	5	6	7	8	9	10	11
Caregiver-Report PS											
1. Total Scale	–										
2. Externalizing Subscale	**.88****	–									
3. Delinquency Subscale	**.67****	**.58****	–								
4. Anxiety Subscale	**.68****	**.37****	**.25****	–							
5. Depression Subscale	**.81****	**.51****	**.33****	**.73****	–						
Youth-Report PS											
6. Total Scale	**.36****	**.33****	**.15****	**.24****	**.32****	–					
7. Externalizing Subscale	**.37****	**.46****	**.18****	**.12***	**.20****	**.82****	–				
8. Delinquency Subscale	**.32****	**.32****	**.48****	.07	**.13***	**.54****	**.47****	–			
9. Anxiety Subscale	**.18****	**.09**†	-.02	**.28****	**.26****	**.74****	**.37****	**.12***	–		
10. Depression Subscale	**.19****	.06	-.05	**.26****	**.35****	**.82****	**.41****	**.21****	**.74****	–	
CAFAS											
11. Discharge Total Scale	**.44****	**.44****	**.35****	**.26****	**.27****	**.29****	**.30****	**.34****	**.09**†	**.15****	–

Note. PS = Problem Severity domain; CAFAS = Child and Adolescent Functional Assessment Scale.
†$p < .10$; *$p < .05$; **$p < .01$

Table 8

Frequencies of Caregiver-Report Over, Under, and In-Agreement with Youth-Report

	Frequency	Percentage	Mean Caregiver (*SD*)	Mean Youth (*SD*)
		Total Scale (20 items)		
Caregiver more than youth	109	29.8	2.11 (.78)	.85 (.56)
In agreement	159	43.4	1.05 (.85)	.98 (.74)
Caregiver less than youth	98	26.8	.82 (.62)	1.89 (.74)
Overall sample	–	–	1.30 (.94)	1.19 (.81)
		Externalizing Subscale (8 items)		
Caregiver more than youth	95	26.0	2.76 (.92)	1.02 (.63)
In agreement	162	44.3	1.31 (1.09)	1.11 (.87)
Caregiver less than youth	109	29.8	1.17 (1.03)	2.13 (.95)
Overall sample	–	–	1.64 (1.22)	1.39 (.97)
		Delinquency Subscale (3 items)		
Caregiver more than youth	80	21.9	2.57 (1.22)	.56 (.79)
In agreement	217	59.3	.50 (.86)	.36 (.67)
Caregiver less than youth	69	18.9	.84 (1.28)	2.11 (1.29)
Overall sample	–	–	1.02 (1.33)	.74 (1.07)
		Anxiety Subscale (3 items)		
Caregiver more than youth	96	26.2	2.41 (1.20)	.93 (.99)
In agreement	165	45.1	.72 (.82)	.86 (.98)
Caregiver less than youth	105	28.7	.61 (.76)	2.66 (1.16)
Overall sample	–	–	1.13 (1.20)	1.40 (1.31)
		Depression Subscale (6 items)		
Caregiver more than youth	97	26.5	2.12 (.98)	.65 (.78)
In agreement	168	45.9	.70 (.74)	.64 (.76)
Caregiver less than youth	101	27.6	.69 (.77)	2.08 (1.06)
Overall sample	–	–	1.07 (1.03)	1.04 (1.07)

Note. n = 366 caregiver-youth dyads. Percentage of Problem Severity domain caregiver- and youth-reports that are in agreement as well as in disagreement in either direction (i.e., caregiver-report higher than youth-report, youth-report higher than caregiver-report). The above values represent average item endorsement per scale score.

Table 9

Caregiver-Youth Total Scale and Subscale Scores as Predictors of Treatment Outcome

	Total Scale		Externalizing Subscale		Delinquency Subscale		Anxiety Subscale		Depression Subscale	
	Polynomial Regression									
Variable	β (*S.E.*)	*t*	β (*S.E.*)	*t*	β (*S.E.*)	*t*	β (*S.E.*)	*t*	β (*S.E.*)	*t*
Constant	41.405 (6.898)	**6.003****	39.035 (6.511)	**5.995****	56.830 (6.936)	**8.194****	48.039 (6.876)	**6.987****	45.349 (6.424)	**7.059****
Caregiver-report (X)	.609 (.294)	**2.072***	1.495 (.520)	**2.874****	.918 (.788)	1.165	1.839 (.778)	**2.365***	2.028 (.629)	**3.226****
Youth-report (Y)	-.317 (.323)	-.982	-.635 (.604)	-1.052	2.901 (.804)	**3.608****	.304 (.763)	.398	-.399 (.578)	-.690
X^2	-.002 (.005)	-.465	-.006 (.021)	-.267	.156 (.128)	1.219	-.008 (.129)	-.063	-.094 (.044)	**-2.160***
X*Y	.008 (.006)	1.214	-.020 (.028)	-.701	.164 (.136)	1.206	.295 (.138)	**2.147***	.054 (.048)	1.118
Y^2	.009 (.006)	1.533	.070 (.028)	**2.484***	-.316 (.153)	**-2.060***	-.029 (.115)	-.255	.054 (.041)	1.309
Intake CAFAS	.269 (.050)	**5.358****	.301 (.049)	**6.130****	.327 (.051)	**6.367****	.363 (.051)	**7.139****	.350 (.052)	**6.715****
Disruptive, Impulse Control, and Conduct Disorders	10.606 (4.368)	**2.428***	8.965 (4.384)	**2.045***	-	-	14.058 (4.634)	**3.034****	15.989 (4.618)	**3.462****
Treatment Length in Days	-.030 (.014)	**-2.166***	-.028 (.014)	**-2.026***	-.025 (.014)	**-1.765**†	-.038 (.015)	**-2.619****	-.038 (.014)	**-2.660****
	Response Surface Analysis									
Slope symmetry line (a_1)	.77		1.34		**4.32****		**2.25***		**2.18***	
Curvature symmetry line (a_2)	**2.21***		1.00		0.03		1.34		.33	
Slope asymmetry line (a_3)	**1.90**†		**2.30***		-1.50		1.27		**2.56***	
Curvature asymmetry line (a_4)	-.08		**1.91**†		-1.13		-1.35		-.95	

Note. S.E. = Standard Error; CAFAS = Child and Adolescent Functional Assessment Scale. Polynomial regression with response surface analysis results examining caregiver-youth Problem Severity Total Scale and Subscale scores as predictors of youth discharge CAFAS Total Scale scores.
†$p < .10$; *$p < .05$; **$p < .01$

Figure 1

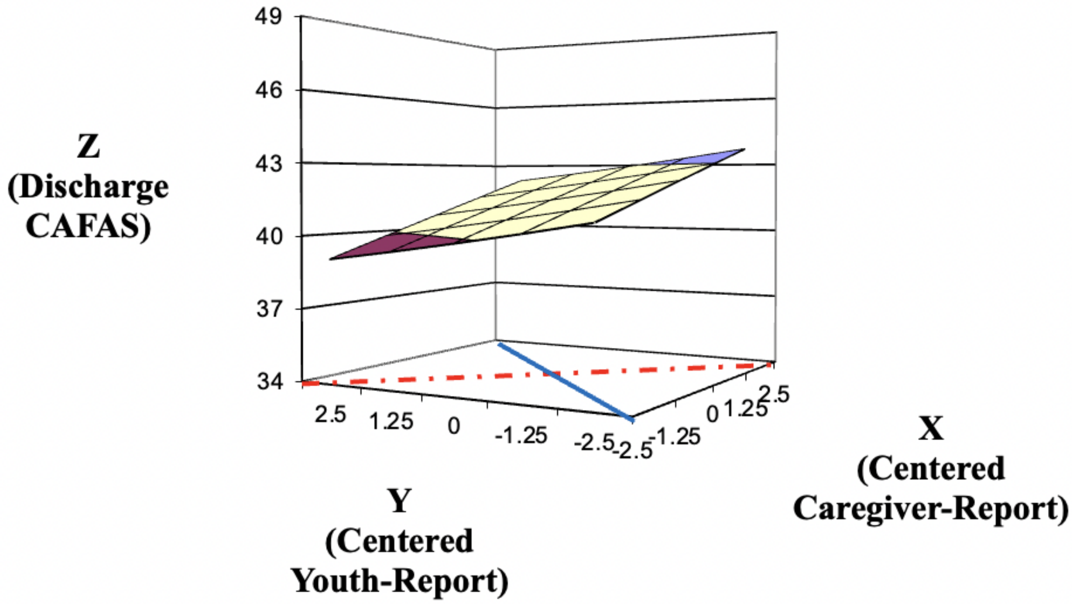
Four Response Surface Analysis Coefficients and the Questions They Answer

Line of congruence: How do matches matter?			
Slope of the line of congruence: a_1		Curvature of the line of congruence: a_2	
Do matches at high values have different outcomes than matches at low values?		Do matches at extreme values have different outcomes than matches at less extreme values?	
Positive a_1	Negative a_1	Positive a_2	Negative a_2
The outcome is higher when X and Y match at higher levels than at lower levels	The outcome is higher when X and Y match at lower levels than at higher levels	The outcome is higher when X and Y match at more extreme levels than at midrange levels	The outcome is higher when X and Y match at midrange levels than at more extreme levels
Line of incongruence: How do mismatches matter?			
Slope along the line of incongruence: a_3		Curvature of the line of incongruence: a_4	
Is one mismatch (X > Y) better or worse than the other (X < Y)?		Are matches better or worse than mismatches?	
Positive a_3	Negative a_3	Positive a_4	Negative a_4
The outcome is higher when X is higher than Y than when Y is higher than X	The outcome is higher when Y is higher than X than when X is higher than Y	The outcome is higher the more X and Y deviate from one another	The outcome is higher the more X and Y match one another

Note. Reprinted from "Response Surface Analysis: A Tutorial for Examining Linear and Curvilinear Effects," by Rodrigues, A. C., 2021, *Journal of Contemporary Administration*, *25*, p. 11. Copyright 2021 by Revista de Administração Contemporânea.

Figure 2

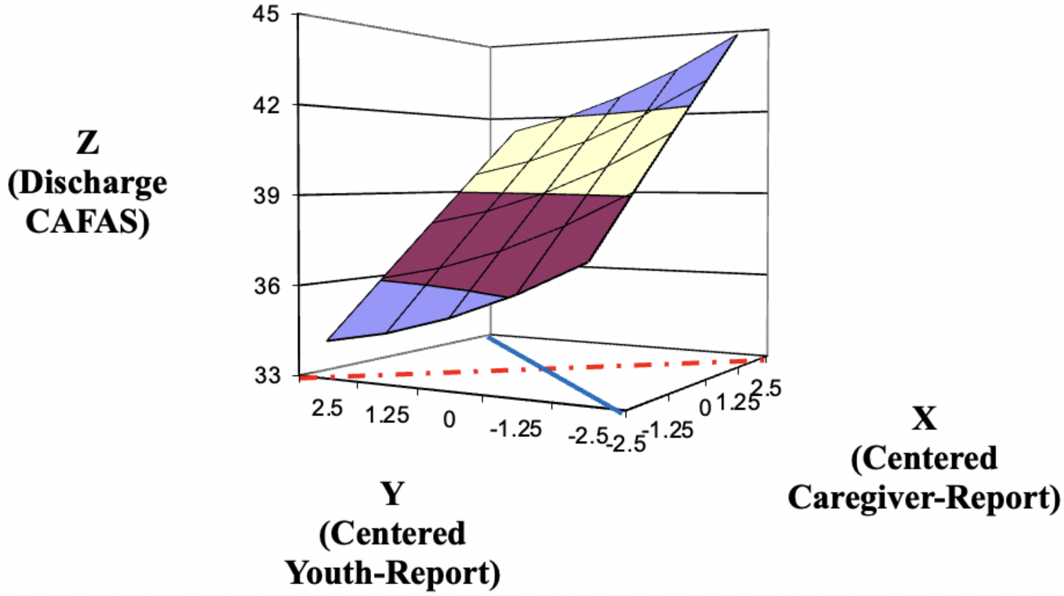
Response Surface: Discharge CAFAS as Predicted by Caregiver and Youth Total Scale Scores



Note. Adapted from "Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores," by Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D., 2010, *Journal of Business and Psychology, 25*, p. 553. Copyright 2010 by Springer Science+Business Media, LLC. Predictor variables X (caregiver-report) and Y (youth-report) are centered around the midpoint of the Problem Severity domain's scale. The blue solid line reflects the line of congruence. The red dashed line reflects the line of incongruence.

Figure 3

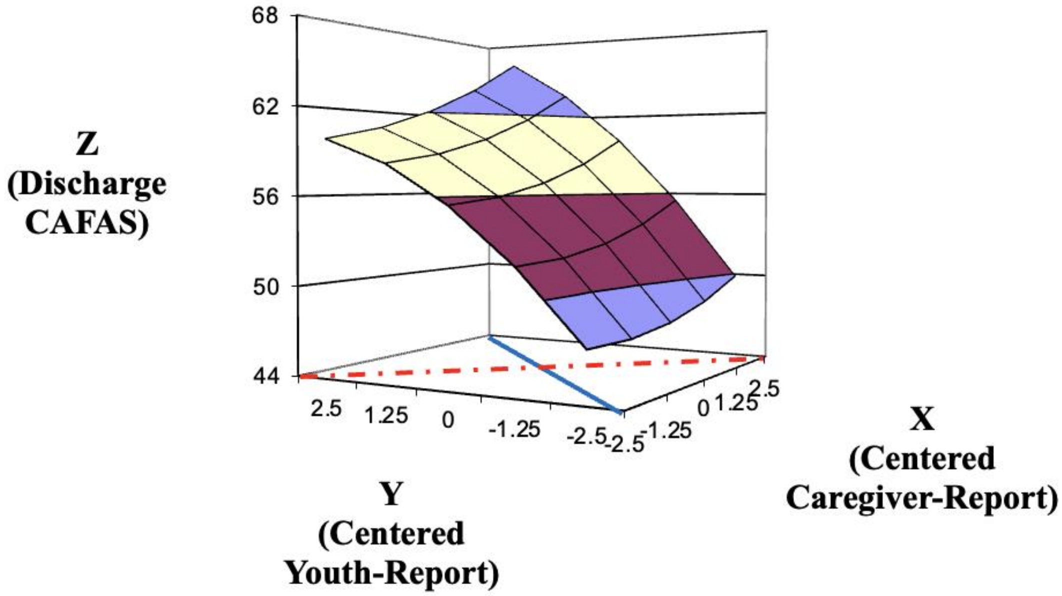
Response Surface: Discharge CAFAS as Predicted by Caregiver and Youth Externalizing Subscale Scores



Note. Adapted from "Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores," by Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D., 2010, *Journal of Business and Psychology, 25*, p. 553. Copyright 2010 by Springer Science+Business Media, LLC. Predictor variables X (caregiver-report) and Y (youth-report) are centered around the midpoint of the Problem Severity domain's scale. The blue solid line reflects the line of congruence. The red dashed line reflects the line of incongruence.

Figure 4

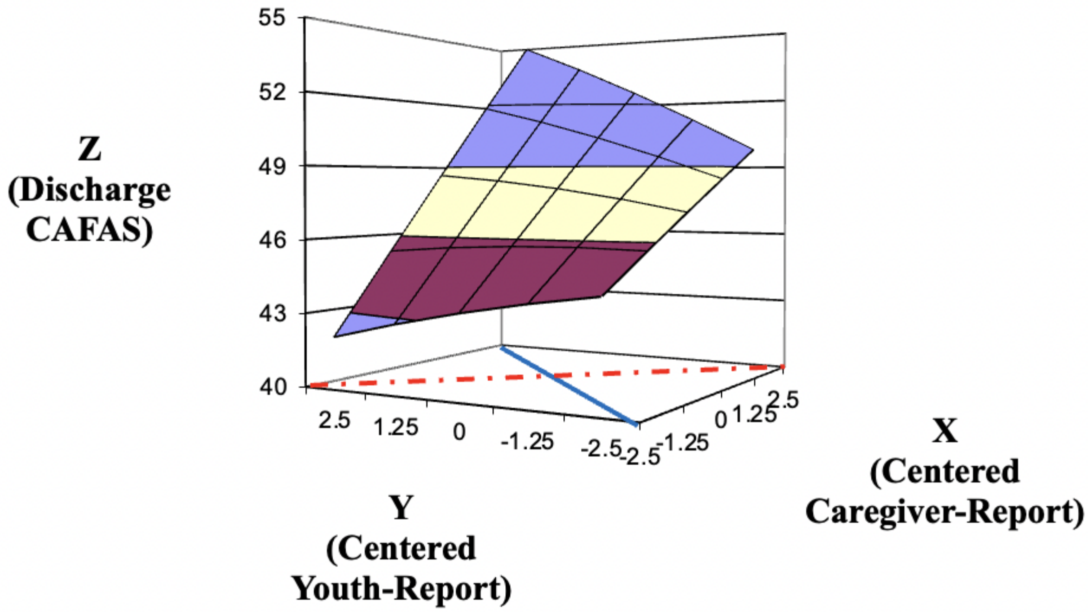
Response Surface: Discharge CAFAS as Predicted by Caregiver and Youth Delinquency Subscale Scores



Note. Adapted from "Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores," by Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D., 2010, *Journal of Business and Psychology, 25*, p. 553. Copyright 2010 by Springer Science+Business Media, LLC. Predictor variables X (caregiver-report) and Y (youth-report) are centered around the midpoint of the Problem Severity domain's scale. The blue solid line reflects the line of congruence. The red dashed line reflects the line of incongruence.

Figure 5

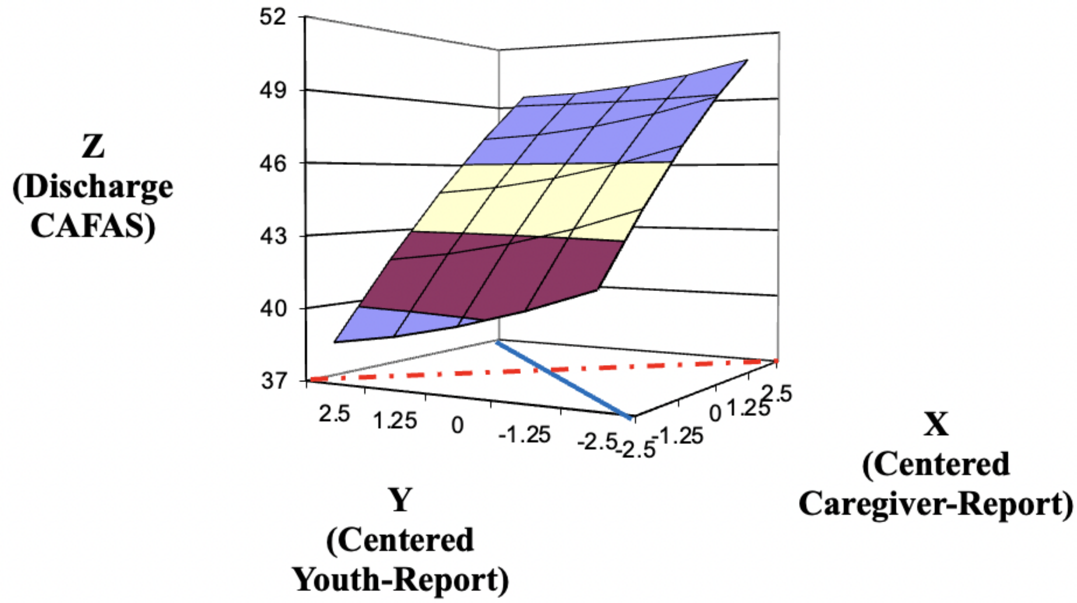
Response Surface: Discharge CAFAS as Predicted by Caregiver and Youth Anxiety Subscale Scores



Note. Adapted from "Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores," by Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D., 2010, *Journal of Business and Psychology, 25*, p. 553. Copyright 2010 by Springer Science+Business Media, LLC. Predictor variables X (caregiver-report) and Y (youth-report) are centered around the midpoint of the Problem Severity domain's scale. The blue solid line reflects the line of congruence. The red dashed line reflects the line of incongruence.

Figure 6

Response Surface: Discharge CAFAS as Predicted by Caregiver and Youth Depression Subscale Scores



Note. Adapted from "Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores," by Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D., 2010, *Journal of Business and Psychology, 25,* p. 553. Copyright 2010 by Springer Science+Business Media, LLC. Predictor variables X (caregiver-report) and Y (youth-report) are centered around the midpoint of the Problem Severity domain's scale. The blue solid line reflects the line of congruence. The red dashed line reflects the line of incongruence.

Appendix A: The Child and Adolescent Functional Assessment Scale

CAFAS™ PROFILE : YOUTH'S FUNCTIONING

Youth's Name ______________________ ID# ______________________ Rater ______________________ Date ____/____/____ Site ____________

Level of Impairment	Role Performance: School/Work	Role Performance: Home	Role Performance: Community	Behavior Toward Others	Moods/ Self-Harm: Moods/ Emotions	Moods/ Self-Harm: Self-Harmful Behavior	Substance Use	Thinking
SEVERE 30	1 2 3 4 5 6 7 8 9 10 11 ○	41 42 43 44 45 46 47 48 49 50 ○	66 67 68 69 70 71 72 ○	88 89 90 91 92 ○	116 117 118 119 120 ○	142 143 144 145 ○	154 155 156 157 158 159 160 161 162 163 164 ○	182 183 184 185 186 ○
MODERATE 20	12 13 14 15 16 17 18 19 20 21 ○	51 52 53 54 55 56 ○	73 74 75 76 77 78 79 ○	93 94 95 96 97 98 99 100 101 102 ○	121 122 123 124 125 126 127 ○	146 147 148 ○	165 166 167 168 169 170 171 ○	187 188 189 190 191 192 ○
MILD 10	22 23 24 25 26 27 ○	57 58 59 60 61 ○	80 81 82 83 ○	103 104 105 106 107 108 109 110 ○	128 129 130 131 132 133 134 135 ○	149 150 ○	172 173 174 175 ○	193 194 195 196 197 ○
MINIMAL/NO 0	28 29 30 31 32 33 34 35 36 37 38 39 ○	62 63 64 ○	84 85 86 ○	111 112 113 114 ○	136 137 138 139 140 ○	151 152 ○	176 177 178 179 180 ○	198 199 ○
COULD NOT SCORE	40 ○	65 ○	87 ○	115 ○	141 ○	153 ○	181 ○	200 ○

For each scale: (1) mark the item number(s) which corresponds to those marked on the CAFAS™ form, (2) fill in the circle indicating severity level, (3) connect the circles.

Appendix B: The Ohio Youth Problem Severity, Functioning, and Satisfaction Scales

Ohio Mental Health Consumer Outcomes System
Ohio Youth Problem, Functioning, and Satisfaction Scales

P

Parent Rating – Short Form

Child's Name: ____________________ Date: __________ Child's Grade: ___ ID#: ________________ Completed by Agency

Child's Date of Birth: ______________ Child's Sex: ☐ Male ☐ Female Child's Race: ____________

Form Completed By: ☐ Mother ☐ Father ☐ Step-mother ☐ Step-father ☐ Other: ______________

Instructions: Please rate the degree to which your child has experienced the following problems in the past 30 days.	Not at All	Once or Twice	Several Times	Often	Most of the Time	All of the Time
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

(Add ratings together) Total ________

Instructions: Please circle your response to each question.

1. Overall, how satisfied are you with your relationship with your child right now?
   1. Extremely satisfied
   2. Moderately satisfied
   3. Somewhat satisfied
   4. Somewhat dissatisfied
   5. Moderately dissatisfied
   6. Extremely dissatisfied
2. How capable of dealing with your child's problems do you feel right now?
   1. Extremely capable
   2. Moderately capable
   3. Somewhat capable
   4. Somewhat incapable
   5. Moderately incapable
   6. Extremely incapable
3. How much stress or pressure is in your life right now?
   1. Very little
   2. Some
   3. Quite a bit
   4. A moderate amount
   5. A great deal
   6. Unbearable amounts
4. How optimistic are you about your child's future right now?
   1. The future looks very bright
   2. The future looks somewhat bright
   3. The future looks OK
   4. The future looks both good and bad
   5. The future looks bad
   6. The future looks very bad

Total: ________

Instructions: Please circle your response to each question.

1. How satisfied are you with the mental health services your child has received so far?
   1. Extremely satisfied
   2. Moderately satisfied
   3. Somewhat satisfied
   4. Somewhat dissatisfied
   5. Moderately dissatisfied
   6. Extremely dissatisfied
2. To what degree have you been included in the treatment planning process for your child?
   1. A great deal
   2. Moderately
   3. Quite a bit
   4. Somewhat
   5. A little
   6. Not at all
3. Mental health workers involved in my case listen to and value my ideas about treatment planning for my child.
   1. A great deal
   2. Moderately
   3. Quite a bit
   4. Somewhat
   5. A little
   6. Not at all
4. To what extent does your child's treatment plan include your ideas about your child's treatment needs?
   1. A great deal
   2. Moderately
   3. Quite a bit
   4. Somewhat
   5. A little
   6. Not at all

Total: ________

Instructions: Please rate the degree to which your child's problems affect his or her current ability in everyday activities. Consider your child's current level of functioning.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4

(Add ratings together) Total ________

Ohio Mental Health Consumer Outcomes System
Ohio Youth Problem, Functioning, and Satisfaction Scales

Youth Rating – Short Form (Ages 12-18)

Y

Name: ____________________________ Date: _____________ Grade: ______ ID#: __________________ Completed by Agency

Date of Birth: ______________________ Sex: ☐ Male ☐ Female Race: __________________

Instructions: Please rate the degree to which you have experienced the following problems in the past 30 days.	**Not at All**	**Once or Twice**	**Several Times**	**Often**	**Most of the Time**	**All of the Time**
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

(Add ratings together) Total ________

Instructions: Please circle your response to each question.

1. Overall, how satisfied are you with your life right now?
    1. Extremely satisfied
    2. Moderately satisfied
    3. Somewhat satisfied
    4. Somewhat dissatisfied
    5. Moderately dissatisfied
    6. Extremely dissatisfied
2. How energetic and healthy do you feel right now?
    1. Extremely healthy
    2. Moderately healthy
    3. Somewhat healthy
    4. Somewhat unhealthy
    5. Moderately unhealthy
    6. Extremely unhealthy
3. How much stress or pressure is in your life right now?
    1. Very little stress
    2. Some stress
    3. Quite a bit of stress
    4. A moderate amount of stress
    5. A great deal of stress
    6. Unbearable amounts of stress
4. How optimistic are you about the future?
    1. The future looks very bright
    2. The future looks somewhat bright
    3. The future looks OK
    4. The future looks both good and bad
    5. The future looks bad
    6. The future looks very bad

Total: ________

Instructions: Please circle your response to each question.

1. How satisfied are you with the mental health services you have received so far?
    1. Extremely satisfied
    2. Moderately satisfied
    3. Somewhat satisfied
    4. Somewhat dissatisfied
    5. Moderately dissatisfied
    6. Extremely dissatisfied
2. How much are you included in deciding your treatment?
    1. A great deal
    2. Moderately
    3. Quite a bit
    4. Somewhat
    5. A little
    6. Not at all
3. Mental health workers involved in my case listen to me and know what I want.
    1. A great deal
    2. Moderately
    3. Quite a bit
    4. Somewhat
    5. A little
    6. Not at all
4. I have a lot of say about what happens in my treatment.
    1. A great deal
    2. Moderately
    3. Quite a bit
    4. Somewhat
    5. A little
    6. Not at all

Total: ________

Instructions: Below are some ways your problems might get in the way of your ability to do everyday activities. Read each item and circle the number that best describes your current situation.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4

(Add ratings together) Total ________

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles.* University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232. https://doi.org/10.1037/0033-2909.101.2.213

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. SAGE Publications.

Augenstein, T. M., Visser, K. F., Gallagher, K., De Los Reyes, A., D'Angelo, E. J., & Nock, M. K. (2022). Multi-informant reports of depressive symptoms and suicidal ideation among adolescent inpatients. *Suicide and Life-Threatening Behavior, 52*, 99-109. https://doi.org/10.1111/sltb.12803

Baumgartner, N., Häberling, I., Emery, S., Strumberger, M., Nalani, K., Erb, S., Bachmann, S., Wöckel, L., Müller-Knapp, U., Rhiner, B., Contin-Waldvogel, B., Schmeck, K., Walitza, S., & Berger, G. (2020). When parents and children disagree: Informant discrepancies in reports of depressive symptoms in clinical interviews. *Journal of affective disorders, 272*, 223–230. https://doi.org/10.1016/j.jad.2020.04.008

Becker-Haimes, E. M., Jensen-Doss, A., Birmaher, B., Kendall, P. C., & Ginsburg, G. S. (2018). Parent-youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry, 23*, 42-56. https://doi.org/10.1177/1359104516689586

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. https://doi.org/10.1037/0033-2909.107.2.238

Bonadio, F. T., & Tompsett, C. (2016). The factor structure of the Ohio Scales: A practical measure of psychological symptoms in youth. *Journal of Child and Family Studies*, *26*, 101-117. https://doi.org/10.1007/s10826-016-0544-0

Carlston, D. (2004). *Demographic and clinical factors associated with parent-child discrepancy from multiple statistical approaches* [Unpublished doctoral dissertation]. Ohio University.

Carlston, D., & Ogles, B. (2006). The impact of items and anchors on parent-child reports of problem behavior. *Child and Adolescent Social Work Journal*, *23*, 24-37. https://doi.org/10.1007/s10560-005-0028-3

Carlston, D. L., & Ogles, B. M. (2009). Age, gender, and ethnicity effects on parent-child discrepancy using identical item measures. *Journal of Child and Family Studies*, *18*, 125-135. https://doi.org/10.1007/s10826-008-9213-2

Chen, Y. Y., Ho, S. Y., Lee, P. C., Wu, C. K., & Gau, S. S. (2017). Parent-child discrepancies in the report of adolescent emotional and behavioral problems in Taiwan. *PLoS ONE, 12*(6), 1-12. https://doi.org/10.1371/journal.pone.0178863

Child and Adolescent Mental Health Division, State of Hawai‘i, Department of Health. (2018). *Child and Adolescent Mental Health Performance Standard.* Retrieved from: https://health.hawaii.gov/CAMHD/files/2018/06/Teal-Book-2018.pdf

Child and Adolescent Mental Health Division, State of Hawai‘i, Department of Health. (2019). *Child and Adolescent Mental Health Annual Evaluation Summary: Fiscal Year 2018.* Retrieved from: https://health.hawaii.gov/camhd/files/2019/05/CAMHD-Annual-Eval-FY2018-Prez-POST-2019_05_16.pdf

Chun, D. (2008). *Parent-child discrepancy: A comparison of U.S. and South Korean clinical samples* [Unpublished doctoral dissertation]. Ohio University.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *40*(1), 1-9. https://doi.org/10.1080/15374416.2011.533405

De Los Reyes, A., Alfano, C. A., & Beidel, D. C. (2010). The relations among measurements of informant discrepancies within a multisite trial of treatments for childhood social phobia. *Journal of Abnormal Child Psychology, 38*, 395-404. https://doi.org/10.1007/s10802-009-9373-6

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*, 858-900. https://doi.org/10.1037/a0038498

De Los Reyes, A., & Epkins, C. (2023). Introduction to a special issue. A dozen year of demonstrating that informant discrepancies are more than measurement error: Toward guidelines for integrating data from multi-informant assessment of youth mental health. *Journal of Clinical Child & Adolescent Psychology, 52*, 1-18. https://doi.org/10.1080/15374416.2022.2158843

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*, 330-334. https://doi.org/10.1037/1040-3590.16.3.330

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483-509. https://doi.org/10.1037/0033-2909.131.4.483

De Los Reyes, A., & Kazdin, A. E. (2008). When the evidence says, "yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science, 17*, 47-51. https://doi.org/10.1111/j.1467-8721.2008.00546.x

De Los Reyes, A., & Ohannessian, C. M. (2016). Introduction to the special issue: Discrepancies in adolescent–parent perceptions of the family and adolescent adjustment. *Journal of Youth and Adolescence*, *45*, 1957-1972. https://doi.org/10.1007/s10964-016-0533-z

De Los Reyes, A., Ohannessian, C. M., & Racz, S. J. (2019). Discrepancies between adolescent and parent reports about family relationships. *Child Development Perspectives, 13*, 53-58. https://doi.org/10.1111/cdep.12306

De Los Reyes, A., Talbott, E., Power, T. J., Michel, J. J., Cook, C. R., Racz, S. J., & Fitzpatrick, O. (2022). The needs-to-goals gap: How informant discrepancies in youth mental health assessment impact service delivery. *Clinical Psychology Review, 92*, 1-17. https://doi.org/10.1016/j.cpr.2021.102114

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123-149. http://dx.doi.org/10.1146/annurev-clinpsy-050212-185617

Edelbrock, C., Costello, A. J., Dulcan, M. K., Conover, N. C., & Kala, R. (1986). Parent-child agreement on child psychiatric symptoms assessed via structured interview. *Journal of Child Psychology and Psychiatry*, *27*, 181-190.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51-100. https://dx.doi.org/10.1006/obhd.1994.1029

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*, 1577-1613.

Ferdinand R. F., van der Ende, J., & Verhulst, F. C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology*, *113*(2), 198-206. http://dx.doi.org/10.1037/0021-843X.113.2.198

Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2006). Prognostic value of parent-adolescent disagreement in a referred sample. *European Child & Adolescent Psychiatry*, *15*, 156-162. http://dx.doi.org/10.1007/s00787-005-0518-z

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). London: SAGE Publications Ltd.

Fife, D. (2020). The eight steps of data analysis: A graphical framework to promote sound statistical analysis. *Perspectives on Psychological Science*, *15*(4), 1054-1075. https://doi.org/10.1177/1745691620917333

George, D., & Mallery, P. (2013). *SPSS for windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.

Goolsby, J., Rich, B. A., Hinnant, B., Habayeb, S., Berghorst, L., De Los Reyes, A., & Alvord, M. K. (2018). Parent-child informant discrepancy is associated with poorer treatment outcome. *Journal of Child and Family Studies, 27*, 1228-1241. https://doi.org/10.1007/s10826-017-0946-7

Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 30-40. https://doi.org/10.1097/00004583-200301000-00008

Guo, X., & Slesnick, N. (2013). Family versus individual therapy: Impact on discrepancies between parents' and adolescents' perceptions over time. *Journal of Marital and Family Therapy, 39*, 182-194. http://dx.doi.org/10.1111/j.1752-0606.2012.00301.x

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (3rd ed.). Routledge.

Hodges, K. (1995). *CAFAS: Self-training manual*. Functional Assessment Systems.

Hodges, K. (1998). *Child and Adolescent Functional Assessment Scale (CAFAS)*. Functional Assessment Systems.

Hodges, K., & Wong, M. M. (1996). Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale (CAFAS). *Journal of Child and Family Studies, 5*(4), 445-467. https://doi.org/10.1007/BF02233865

Hu, L. T., & Bentler, P. M. (1999). Cuttoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. https://doi.org/10.1080/10705519909540118

Human, L. J., Dirks, M. A., DeLongis, A., & Chen, E. (2016). Congruence and incongruence in adolescents' and parents' perceptions of the family: Using response surface analysis to examine links with adolescents' psychological adjustment. *Journal of Youth and Adolescence*, *45*, 2022-2035. https://doi.org/10.1007/s10964-016-0517-z

Humphreys, K. L., Weems, C. F., & Scheeringa, M. S. (2017). The role of anxiety control and treatment implications of informant agreement on child PTSD symptoms. *Journal of Clinical Child & Adolescent Psychology*, *46*, 903-914. https://doi.org/10.1080/15374416.2015.1094739

Hunsley, J., & Lee, C. M. (2014). *Introduction to clinical psychology* (2nd ed.). Wiley.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*, 29-51. http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091419

Jackson, D. S., Keir, S. S., Sender, M., & Mueller, C. W. (2017). Reliable change and outcome trajectories across levels of care in mental health system for youth. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 141-154. https://dx.doi.org/10.1007/s10488-015-0690-5

Kazdin, A. E. (1989). Developmental psychopathology: Current research, issues, and directions. *American Psychologist*, *44*, 180-187. https://doi.org/10.1037/0003-066X.44.2.180

Kazdin, A. E. (2000). *Psychotherapy for children and adolescents: Directions for research and practice*. Oxford University Press. https://doi.org/10.1093/med:psych/9780195126181.001.0001

Kendall, P. C., Panichelli Mindel, S. M., Sugarman, A., & Callahan, S. A. (1997). Exposure to child anxiety: Theory, research, and practice. *Clinical Psychology: Science and Practice*, *4*(1), 29-39. https://doi.org/10.1111/j.1468-2850.1997.tb00096.x

Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*, 52-54. https://doi.org/10.5395/rde.2013.38.1.52

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *The Journal of Child Psychology and Psychiatry*, *34*, 991-1006. http://dx.doi.org/10.1111/j.1469-7610.1993.tb01103.x

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *The American Journal of Psychiatry*, *160*, 1566-1577. http://dx.doi.org/10.1176/appi.ajp.160.9.1566

Laba, T. M., Nakamura, B. J., & Mueller, C. W. (2022). A psychometric investigation of the Ohio Scales, short form, problem severity domain, youth- and caregiver-report versions. *Journal of Psychopathology and Behavioral Assessment*, *44*, 539-559. http://doi.org/10.1007/s10862-022-09970-y

Laird, R. D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology*, *41*(1), 1-14. https://dx.doi.org/10.1007/s10802-012-9659-y

Lapouse, R., & Monk, M. A. (1958). An epidemiologic study of behavior characteristics in children. *American Journal of Public Health*, *48*, 1134-1144. http://dx.doi.org/10.2105/AJPH.48.9.1134

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198-1202. https://doi.org/10.1080/01621459.1988.10478722

Makol, B. A., De Los Reyes, A., Ostrander, R. S., & Reynolds, E. K. (2019). Parent-youth divergence (and convergence) in reports of youth internalizing problems in psychiatric inpatient care. *Journal of Abnormal Child Psychology*, *47*, 1677-1689. https://doi.org/10.1007/s10802-019-00540-7

Mastrotheodoros, S., Van der Graaff, J., Deković, M., Meeus, W. H. J., & Branje, S. (2020). Parent-adolescent conflict across adolescence: Trajectories of informant discrepancies and associations with personality types. *Journal of Youth and Adolescence, 49*, 119-135. https://doi.org/10.1007/s10964-019-01054-7

McLennan, P., Laba, T. M., & Nakamura, B. J. (2022). Effects of modularized workshop trainings on youth clinical progress in a public mental health system. *Journal of Emotional and Behavioral Disorders, 31*, 66-78.

Milette-Winfree, M., Nakamura, B. J., Kotte, A., & Higa-McMillan, C. K. (2019). Multilevel predictors of case managers' assessment administration behavior in a precursor to a measurement feedback system. *Administration and Policy in Mental Health Services Research, 46*(5), 636-648. https://doi.org/10.1007/s10488-019-00941-2

Moore, A. (2008). *The effects of ambiguous loss on behavioral problems in children placed in out-of-home care* [Unpublished doctoral dissertation]. Drexel University.

Mueller, R. O., & Hancock, G. R. (2007). Best practices in structural equation modeling. In J. Osborne (Ed.), *Best Practices in Quantitative Methods* (pp. 488-510). SAGE Publications, Inc. https://doi.org/10.4135/9781412995627

Mueller, R. O., & Hancock, G. R. (2019). Structural equation modeling. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 445–456). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315755649-33

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(4), 599-620. https://doi.org/10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (2015). *Mplus* (Version 7.31). Muthén & Muthén. https://www.statmodel.com/

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, *17*, 372-411. https://doi.org/10.1177/1094428114548590

Ogles, B. M., Dowell, K., Hatfield, D., Melendez, G., & Carlston, D. (2004). The Ohio Scales. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 2, pp. 275-304). Lawrence Earlbaum.

Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (1999). *The Ohio Youth Problem, Functioning, and Satisfaction Scales (short form): User's manual*. Ohio Department of Mental Health.

Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (2000). *The Ohio Youth Problem, Functioning, and Satisfaction Scales: Technical manual.* Ohio Department of Mental Health.

Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., Bilenberg, N., Bird, H., Chahed, M., Dobrean, A., Döpfner, M., Erol, N., Hannesdottir, H., Kanbayashi, Y., Lambert, M. C., Leung, P. W., Minaei, A., Novik, T. S., Oh, K. J., …

Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child and Adolescent Psychology, 42*(2), 262-273. https://doi.org/10.1080/15374416.2012.717870

Rey, J. M., Schrader, E., & Morris-Yates, A. (1992). Parent-child agreement on children's behaviors reported by the child behaviour checklist (CBCL). *Journal of Adolescence, 15*(3), 219-230. https://doi.org/10.1016/0140-1971(92)90026-2

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavioral Assessment System for Children Manual* (3rd ed.). Pearson.

Riley, S. E., Stromberg, A. J., & Clark, J. J. (2009). Relationship between caregiver hopefulness and satisfaction with their children's mental health services. *Community Mental Health Journal, 45*, 307-315. https://doi.org/10.1007/s10597-009-9188-5

Roberts, B. W., & Caspi, A. (2001). Authors' responses - Personality development and the person-situation debate: It's déjà vu all over again. *Psychological Inquiry, 12*(2), 104-109. https://doi.org/10.1207/S15327965PLI1202_04

Rodrigues, A. C. (2021). Response surface analysis: A tutorial for examining linear and curvilinear effects. *Journal of Contemporary Administration, 25*, 1-14. https://doi.org/10.1590/1982-7849rac2021200293.en

Seiffge-Krenke, I., & Kollmar, F. (1998). Discrepancies between mothers' and fathers' perceptions of sons' and daughters' problem behaviour: A longitudinal analysis of parent-adolescent agreement on internalising and externalising problem behaviour. *Journal of Child Psychology and Psychiatry, 39*(5), 687-697.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*(4), 543-554. https://doi.org/10.1007/s10869-010-9183-4

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611. https://doi.org/10.2307/2333709

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180. https://doi.org/10.1207/s15327906mbr2502_4

Thompson, R. J., Jr, Merritt, K. A., Keith, B. R., Murphy, L. B., & Johndrow, D. A. (1993). Mother-child agreement on the child assessment schedule with nonreferred children: A research note. *Journal of Child Psychology and Psychiatry*, *34*(5), 813-820. https://doi.org/10.1111/j.1469-7610.1993.tb01073.x

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1-10. https://doi.org/10.1007/BF02291170

van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2020). Multitrait-multimethod analyses of change of internalizing and externalizing problems in adolescence: Predicting internalizing and externalizing DSM disorders in adulthood. *Journal of abnormal psychology*, *129*(4), 343-354. https://doi.org/10.1037/abn0000510

Verhulst, F. C., & Ende, J. V. D. (1992). Agreement between parents' reports and adolescents' self-reports of problem behavior. *Journal of Child Psychology and Psychiatry*, *33*(6), 1011-1023. https://doi.org/10.1111/j.1469-7610.1992.tb00922.x

Wang, J., Liu, L., Wu, H., Yang, X., Wang, Y., & Wang, L. (2014). Agreement between parents and adolescents on emotional and behavioral problems and its associated factors among

Chinese school adolescents: A cross-sectional study. *BMC Psychiatry*, *14*, 1-8. https://doi.org/10.1186/1471-244X-14-114

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Sage Publications, Inc.

Warnick, E. M., Weersing, V. R., Scahill, L., & Woolston, J. L. (2009). Selecting measures for use in child mental health services: A scorecard approach. *Administration and Policy in Mental Health*, *36*(2), 112-122. https://doi.org/10.1007/s10488-008-0203-x

Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, *56*, 337-363. http://dx.doi.org/10.1146/annurev.psych.55.090902.141449

Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent-child (dis)agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology*, *69*, 1018-1025. https://doi.org/10.1037//0022-006x.69.6.1018