Theses and Dissertations   Graduate School

2023

Using a single measure to assess adherence and differentiation in family therapy for adolescent externalizing problems

Stephanie Violante
Virginia Commonwealth University

Using a single measure to assess adherence and differentiation in family therapy for
adolescent externalizing problems

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor
of Philosophy, at Virginia Commonwealth University

by

Stephanie Violante
Master of Science, Virginia Commonwealth University, 2018
Bachelor of Science, University of Washington, 2013

Director: Bryce D. McLeod, Ph.D.
Professor, Department of Psychology

Virginia Commonwealth University
Richmond, Virginia
April, 2023

Acknowledgement

The completion of this dissertation would not have been possible without the unconditional love and support of my parents, Craig and Nancy Violante. I am deeply grateful for your endless encouragement, generosity, and confidence in me. Thank you to my sister, Marissa, brother, Dennis, and nieces, Adalyn and Maisy, for always supporting me and cheering me on from afar.

I would also like to offer my sincerest gratitude to my mentor and dissertation chair, Dr. Bryce McLeod, for his constant guidance, motivation, support, patience, and kindness throughout the entirety of my graduate career. Thank you to my committee members, Drs. Aaron Hogue, Michael Southam-Gerow, Terri Sullivan, Rosalie Corona, and Joshua Langberg for kindly agreeing to help guide me throughout this process. To Dr. Shannon Dorsey, you sparked my passion for research. Thank you for encouraging me to pursue my goals. Thank you to Alex MacLean, Molly Bobek, and Dr. Nicole Porter for their support and patience in answering my many questions.

To my cohort, lab mates, and friends near and far, thank you endlessly for the love, support, and many laughs. Dr. Ruben Martinez, thank you for your friendship, mentorship, and guidance. Katrina Markowicz, I'm so lucky to have been able to do this alongside you. Jennifer Cecilione, thank you for selflessly dedicating your time to help me with this project. Caitlyn, Rachel, and Lauren, thank you for the joy. Lumi, you're my light!

Finally, I would like to dedicate this dissertation to the memory of my uncle, Paul James Pfeifer, who filled the world with music, laughter, and love. I will continue to pursue a career dedicated to improving mental health services in your honor.

Table of Contents

List of Tables

Abstract

USING A SINGLE MEASURE TO ASSESS ADHERENCE AND DIFFERENTIATION IN FAMILY THERAPY FOR ADOLESCENT EXTERNALIZING PROBLEMS

By Stephanie Violante, M.S.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2023.

Major Director: Bryce D. McLeod, Ph.D.
Professor, Department of Psychology

Family therapy is considered a well-established treatment approach for adolescent externalizing problems; however, research examining its effectiveness when delivered in routine practice settings has produced mixed results. The accurate interpretation of these inconsistent results requires an understanding of what therapeutic techniques were delivered, including those that were prescribed (i.e., adherence) and those that were non-prescribed (i.e., differentiation), suggesting the need for a measure that can concurrently assess treatment adherence and differentiation. The current study examined the extent to which an observational measure of an array of therapeutic techniques for youth emotional and behavioral problems (the Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale; TPOCS-RS) can assess both adherence to and differentiation from family therapy through psychometric assessment. Treatment sessions ($N = 103$) from 42 adolescents with primary externalizing problems treated by 24 therapists in routine practice settings were independently coded using the TPOCS-RS. Treatment sessions were from one of three groups: (a) usual care family therapy; (b) usual care family therapy plus the medication integration protocol; or (c) nonfamily therapy usual care. Adherence was represented by a TPOCS-RS Family Therapy

subscale comprising the TPOCS-RS items considered to be core elements of family therapy for adolescent externalizing problems. Interrater reliability for the TPOCS-RS Family Therapy subscale was ICC = .90 and average interrater reliability for the other TPOCS-RS subscales was ICC = .83. The TPOCS-RS Family Therapy subscale scores demonstrated evidence of convergent and discriminant validity via associations with observer- and therapist-rated measures of adherence to family therapy, observer- and therapist-rated measures of non-family therapy techniques, and a measure of client-therapist alliance. The TPOCS-RS Family Therapy subscale also demonstrated evidence of discriminative validity by identifying expected group differences. Predictive validity was not supported. Results provide preliminary evidence that the TPOCS-RS can serve as a measure of adherence to, and differentiation from, family therapy in a population of adolescents with externalizing problems.

Literature Review

Adolescent Externalizing Problems

Externalizing problems, including conduct problems, delinquency, and substance misuse, are common among adolescents and considered a significant public health concern (Baldwin et al., 2012). Among adolescents, conduct disorder has a 1-year prevalence rate ranging from 5 to 9% (American Psychiatric Association, 2013) and substance use disorder has a prevalence rate of approximately 5%, with more than half of adolescents in the US reporting lifetime alcohol use and 25% reporting exposure to illicit drugs (9.9% reported alcohol use, 6.5% reported marijuana use, and 1.0% reported prescription pain medication misuse within the last month; Center for Behavioral Health Statistics and Quality, 2018; Merikangas & McClair, 2012). Further, there is high comorbidity between conduct problems and substance use, ranging from 17 to 85% among treatment-seeking youth (Hawke et al., 2018; Stanhope et al., 2018; Wu et al., 2011). Adolescents with these problems are more likely to experience family dysfunction, low academic achievement, legal problems (currently 31 million youth have juvenile justice system involvement), and adverse health outcomes including death (Dauria et al., 2018; Grella et al., 2001; Keyes et al., 2015; McLeod et al., 2012; Staff et al., 2008; Wong et al., 2013). These adverse outcomes disproportionately impact youth of color. Black/African American and Latinx/e youth, compared to their White counterparts with comparable behaviors, are less likely to receive appropriate mental and behavioral health treatment and more likely to have justice system involvement (Mizock & Harkins, 2011). Further, among youth with justice system involvement, youth of color are more likely to be transferred to adult courts and receive longer sentences than their White peers (Mizock & Harkins, 2011). When left untreated, adolescent

conduct and substance use problems are likely to persist into adulthood impacting social and occupational success (Rohde et al., 2007).

Family Therapy

The intervention approach with the highest level of empirical support for treating adolescent externalizing problems is family therapy, a treatment modality that intervenes directly with family members to address dysfunctional interactions and challenges within the family system that contribute to the onset and maintenance of adolescent behavior problems (Baldwin et al., 2012; Hogue et al., 2019). The most well-established family therapy interventions for adolescent externalizing problems include Brief Strategic Family Therapy (BSFT; Szapocznik et al., 2003), Functional Family Therapy (FFT; Alexander & Parsons, 1982), Multidimensional Family Therapy (MDFT; Liddle & Hogue, 2001), and Multisystemic Therapy (MST; Henggeler et al., 1998). These interventions have been shown to not only reduce disruptive behavior and substance use, but also to increase treatment engagement and retention and improve academic and peer functioning (Baldwin et al., 2012; Becker & Curry, 2008; Chorpita et al., 2011; Hogue, Henderson et al., 2014; Hogue & Liddle, 2009; Riedinger et al., 2017; Rowe, 2012; Tanner-Smith et al., 2013; Waldron & Turner, 2008). In addition to manualized family therapies, the delivery of core practice elements (i.e., discrete therapeutic techniques common across multiple evidence-based treatments for a given problem or disorder; Chorpita et al., 2005) found in these well-established family therapy models has produced declines in youth-reported internalizing and externalizing symptoms, delinquent behaviors, and substance use, and these clinical gains were maintained over the course of one year (Hogue, Dauber, Henderson et al., 2015; Hogue et al., 2021).

Although family therapy EBPs (i.e., intervention techniques, models, and approaches with empirical support for treating a specific problem; Chorpita & Daleiden, 2009; Embry & Biglan, 2008; Garland et al., 2010; McHugh & Barlow, 2010) have displayed efficacy when delivered under controlled research conditions (e.g., Alexander & Parsons, 1973; Borduin et al., 1995; Henggeler et al., 1986; Santisteban et al., 2003; Waldron et al., 2001), their effectiveness when delivered under routine practice conditions has been less consistent. In a meta-analysis of 52 randomized controlled trials comparing EBPs with usual care in routine practice settings, 18 of which evaluated family therapy for adolescent externalizing problems, Weisz et al. (2013) found a mean effect size of $d = 0.29$, representing a probability of only 58% that a youth receiving an EBP would have better clinical outcomes than a youth receiving usual clinical care.

Implementation Research

To better understand the mixed findings for EBP effectiveness in routine practice settings, the field of implementation science (i.e., the study of the factors and methods involved in translating EBPs into routine practice; Eccles et al., 2009) has emerged. Implementation research may offer several explanations for Weisz et al.'s (2013) findings. For example, it is possible that EBPs are less effective for youth in practice settings than in research settings due to differences in youth characteristics. Youth in practice settings tend to be more racially and ethnically diverse, experience more comorbid mental and behavioral health diagnoses, and are more likely to come from low-income and single-parent families than those in research settings (Ehrenreich-May et al., 2011; Southam-Gerow et al., 2003; Southam-Gerow et al., 2008). Other explanations focus on aspects of treatment delivery. Specifically, there is evidence suggesting that therapists in routine practice settings tend to deliver lower doses of the EBP under investigation than those in research settings (Smith et al., 2017; Weisz et al., 2009); thus, EBP

3

effectiveness in routine practice settings may be diminished due to insufficient treatment delivery. Alternatively, it is possible that usual care delivered in routine practice settings is not as distinct from EBPs as assumed. Research across multiple youth mental and behavioral health problem areas has indicated that usual care therapists often deliver treatment components based in evidence (e.g., Borntrager et al., 2013; Brookman-Frazee et al., 2010; Garland et al., 2014). Therefore, Weisz et al.'s (2013) discovery that the extent to which EBPs outperform usual care in routine practice settings is smaller than expected may not be due to the ineffectiveness of EBPs in these settings, but rather, the similarities between EBPs and usual care. The ability of researchers to accurately interpret results from implementation and effectiveness trials requires a better understanding of what therapeutic techniques are being delivered, and to what extent, across both treatment and usual care groups; the measurement of treatment integrity may be useful for this purpose.

Treatment Integrity

Treatment integrity is broadly defined as the degree to which an intervention is delivered as intended (McLeod et al., 2009; Perepletchikova & Kazdin, 2005). While there is no single agreed upon conceptualization of treatment integrity, it is commonly considered to comprise adherence (i.e., the extent to which prescribed therapeutic techniques are delivered), competence (i.e., the skill and responsiveness with which prescribed techniques are delivered), and differentiation (the delivery of non-prescribed techniques; Bellg et al., 2004; Cox et al., 2019; Perepletchikova et al., 2007). For the purposes of clarifying what, and to what extent, therapeutic techniques are being delivered in implementation and effectiveness trials, researchers should measure both adherence and differentiation.

Adherence

Of the three treatment integrity components, adherence is the most commonly measured in psychotherapy research (Cox et al., 2019; Perepletchikova et al., 2007), and the majority of adherence measures are designed to assess adherence to a specific treatment manual or protocol (i.e., protocol adherence; Carroll & Nuro, 2002; Cox et al., 2019; Perepletchikova et al., 2007; Regan et al., 2019). Research assessing adherence to family therapy has found that increased protocol adherence is associated with a reduction in externalizing symptoms, internalizing symptoms, and substance use (Gillespie et al., 2017; Hogue et al., 2006; Hogue et al., 2004; Huey et al., 2000; Robbins et al., 2011).

More recently, implementation researchers have proposed a shift toward prioritizing the delivery of core practice elements (i.e., discrete therapeutic techniques found across EBPs for a specific problem area rather than a single specific treatment protocol) allowing for more flexible treatment delivery which can be tailored to individual clients (Chorpita et al., 2005; Chorpita & Daleiden, 2009). With this shift there is a need for adherence measures that capture data at the practice elements level (McLeod et al., 2013). Two such measures exist for family therapy—the Therapist Behavior Rating Scale—Competence (TBRS-C; Hogue, Dauber et al., 2008), and the Inventory of Therapy Techniques—Adolescent Behavior Problems (ITT-ABP; Hogue, Dauber et al., 2014; Hogue, Dauber, Lichvar et al., 2015), described below. Research utilizing these measures has revealed that similar to protocol adherence, adherence to core family therapy practice elements predicts improved clinical outcomes. Specifically, in a sample of adolescents receiving treatment for substance use and related behavior problems, Hogue, Henderson et al. (2008) found that stronger adherence to core family therapy practice elements, measured by the TBRS-C, predicted greater reductions in externalizing symptoms, while intermediate levels of

adherence predicted reductions in internalizing symptoms. Further, Henderson et al. (2019) found that among adolescents receiving usual care treatment for conduct and substance use problems, greater levels of adherence to core family therapy practice elements, assessed using the ITT-ABP, predicted decreases in adolescent-reported delinquency, externalizing behaviors, and substance use, as well as parent-reported externalizing behaviors. Overall, there is strong evidence that adherence to manualized family therapy predicts clinical improvements, and initial evidence that adherence to core family therapy practice elements does so as well.

The measurement of adherence can play an important role in understanding the inconsistent effectiveness of EBPs in routine practice settings. For example, if a family therapy EBP failed to outperform usual care in an effectiveness trial, scores of adherence to core family therapy practice elements across both groups would provide information about (a) if levels of adherence in the EBP group were sufficient to produce desired clinical outcomes; and (b) if core family therapy practice elements were delivered in the usual care group, resulting in contamination bias and the potential minimization of group differences in clinical outcomes (Krishna & Surapaneni, 2010).

Differentiation

When measured concurrently with adherence, the assessment of differentiation (i.e., deviation from prescribed EBP techniques; Bellg et al., 2004; Perepletchikova et al., 2007) can aid in the interpretation of implementation and effectiveness research. Implementation and effectiveness research are concerned with the delivery of EBPs in routine practice settings where it is not uncommon for therapists to deliver techniques found outside those specified as evidence-based for the target problem (Southam-Gerow et al., 2010; Weisz et al., 2009). Such techniques may range from simply not prescribed to explicitly proscribed and may positively or negatively

affect clinical outcomes (Garland et al., 2010; Smith et al., 2017). Therefore, to make accurate interpretations regarding the effectiveness, or lack thereof, of EBPs delivered in routine care settings, differentiation should be assessed (McLeod et al., 2021). For example, if a family therapy EBP failed to outperform usual care in an effectiveness trial, differentiation assessment can help determine if (a) non-EBP elements were delivered in the family therapy group that reduced effectiveness or (b) non-EBP elements were delivered in the usual care group that increased effectiveness.

Differentiation assessment in routine practice settings should include a wide array of therapeutic techniques found across multiple treatment orientations (McLeod et al., 2013). Therapists in routine practice settings have diverse education and training backgrounds leading to the delivery of practices from an array of theoretical orientations (e.g., cognitive-behavioral, client-centered), including those that are not considered evidence-based (e.g., psychodynamic; Garland et al., 2010; McLeod & Weisz, 2010). Even when encouraged to deliver EBPs for a specific problem, therapists in these settings may continue to deliver non-prescribed practices, potentially tempering treatment effectiveness (Perepletchikova & Kazdin, 2005; Weisz et al., 2009). When combined with a measure of adherence, a non-problem-specific measure of differentiation that captures a range of therapeutic techniques across treatment orientations would aid in the interpretation of EBP implementation and effectiveness trials.

Combining Adherence and Differentiation

Given that the majority of existing treatment integrity measures assess adherence only, capturing both adherence and differentiation would require the use of multiple measures; however, using a single measure to capture both adherence and differentiation is more conducive to cross-group comparison (e.g., EBT vs. usual clinical care) because item scores could be

directly compared (Drake et al., 2001; Schoenwald et al., 2011). Thus, implementation and effectiveness research would benefit from measures designed to capture a range of therapeutic techniques across multiple treatment modalities that can be used to measure both adherence and differentiation.

There are currently two existing measure that were designed to capture both adherence to family therapy EBPs and differentiation: the Inventory of Therapeutic Techniques for Adolescent Behavior Problems (ITT-ABP; Hogue, Dauber et al., 2014; Hogue, Dauber, Lichvar et al., 2015) and the Therapist Behavior Rating Scale—Competence (TBRS-C; Hogue, Dauber et al., 2008). The ITT-ABP includes a therapist-rated version and an observer-rated version which rate techniques representing four theoretical approaches with empirical support for treating adolescent behavior problems: cognitive-behavioral therapy (CBT), family therapy (FT), motivational interviewing (MI), and drug counseling (DC; Becker & Curry, 2008; Chorpita et al., 2011; Eyberg et al., 2008). Exploratory and confirmatory factor analysis of the ITT-ABP revealed a three-factor solution, resulting in the combination of the CBT and MI items into a single CBT/MI subscale, a FT subscale, and a DC subscale (Hogue, Dauber et al., 2014). The ITT-ABP has demonstrated internal consistency ranging from $\alpha = .79$ to $\alpha = .90$ (Hogue, Dauber et al., 2014) and interrater reliability between independent observers ranging from ICC = .76 (combined CBT/MI subscale) to ICC = .87 (FT subscale; Hogue, Dauber, Lichvar et al., 2015). Agreement between the therapist-rated and observer-rated versions of the ITT-ABP ranged from ICC = .06 (combined CBT/MI subscale) to ICC = .66 (FT subscale), and both versions have also demonstrated convergence with a self-reported measure of therapist allegiance and proficiency in each of the four treatment approaches (CBT, FT, MI, and DC). Finally, the therapist-rated ITT-

ABP produced subscale scores that were able to discriminate between therapists in a clinic that delivered family therapy and therapists in clinics that did not (Hogue, Dauber et al., 2014).

The TBRS-C is an observational measure of adherence to and competence in core practice elements for individual (CBT) and family therapy for adolescent substance abuse that has demonstrated interrater reliability between independent observers ranging from ICC = .52 (FT subscale) to ICC = .74 (CBT subscale) for adherence and ranging from ICC = .55 (FT subscale) to ICC = .56 (CBT subscale) for competence (Hogue, Dauber et al., 2008). Further, the TBRS-C has produced scores that were distinct from ratings of alliance and demonstrated a pattern of inter-item correlations supporting construct validity (Hogue, Dauber et al., 2008).

Although designed to capture adherence to and differentiation from family therapy, the ITT-ABP and TBRS-C include only practice elements with empirical support for treating adolescent externalizing problems, so they are not able to capture the wide range of therapeutic techniques and treatment modalities that may be delivered in routine practice settings. Family therapy effectiveness and implementation research would therefore benefit from a measure that can capture both adherence to core family therapy practice elements as well as a variety of non-problem-specific therapeutic techniques.

Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale

A non-problem-specific measure that captures a wide array of therapeutic techniques and has potential for assessing both differentiation and adherence to family therapy is the Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale (TPOCS-RS; McLeod et al., 2015; McLeod et al., 2021). The TPOCS-RS is an observational measure that assesses how extensively a therapist delivers discrete therapeutic techniques (e.g.,

9

cognitive education, relaxation) found across multiple treatment orientations for youth emotional and behavioral problems (i.e., cognitive, behavioral, family, client-centered, and psychodynamic). Because the TPOCS-RS is not specific to a clinical problem (e.g., anxiety, depression) or treatment approach (e.g., CBT) it may be well-suited to assess differentiation in routine practice settings and, therefore, implementation and effectiveness research (McLeod et al., 2013; McLeod et al., 2015; Mcleod et al., 2021).

The TPOCS-RS was revised from the Therapy Process Observational Coding System for Child Psychotherapy—Strategies scale (TPOCS-S; McLeod & Weisz, 2010). Although originally designed to characterize the therapeutic techniques delivered in usual clinical care (McLeod & Weisz, 2010), the TPOCS-S has shown initial promise as a measure of differentiation in two effectiveness trials which compared cognitive behavioral therapy (CBT) to usual clinical care (Southam-Gerow et al., 2010; Weisz et al., 2009). In both trials, CBT groups scored significantly higher on cognitive-behavioral techniques, and lower on psychodynamic techniques, than usual clinical care groups, indicating that the TPOCS-S has the ability to distinguish between two separate treatment modalities. To more comprehensively capture therapeutic techniques used in youth treatment for emotional and behavioral problems, and thus be more useful as a measure of differentiation, McLeod et al. (2015) revised the TPOCS-S to include 11 additional techniques (e.g., behavioral activation, parenting skills), resulting in the 42-item TPOCS-RS. McLeod et al. (2021) then added an additional five techniques (e.g., crisis management, therapy engagement) based on feedback from coders who had used the 42-item TPOCS-RS to code therapy sessions conducted across both research and clinical settings, resulting in the 47-item TPOCS-RS.

Evidence for Differentiation

Existing evidence supports the use of the TPOCS-RS as a measure of differentiation. First, in a sample of 954 therapy tapes from 89 children diagnosed with a primary anxiety disorder receiving either manual-based CBT (in a research or practice setting) or non-manualized usual care (in a practice setting), McLeod et al. (2015) found that the magnitude of the correlations among the five subscales (i.e., Cognitive, Behavioral, Family, Client-Centered, and Psychodynamic) were small to medium based on Rosenthal and Rosnow's (1984) guidelines (correlations are "small" if $r \geq .10$, "medium" if $.24 \leq r < .37$, and "large" if $r \geq .37$) except of the correlation between the Cognitive and Behavioral subscales which was large ($r = 0.59$) but not redundant (i.e., $r > 0.85$; Rosenthal & Rosnow, 1984), indicating that the subscales measure distinct techniques. Further, in the same sample, Smith et al. (2017) found that therapists delivering manual-based CBT in research and practice settings delivered the highest dosage of cognitive-behavioral techniques, followed by client-centered, family, and psychodynamic, whereas usual care therapists delivered the highest dosage of family therapy techniques followed by client-centered, psychodynamic, and cognitive-behavioral. Finally, therapists delivering manual-based CBT in the research setting were found to have delivered more cognitive-behavioral and fewer psychodynamic techniques in mid- and late-treatment than those in the practice setting (Smith et al., 2017). Overall, evidence suggests that TPOCS-RS can assess both EBPs (e.g., CBT) and non-EBPs (e.g., psychodynamic) and discriminate between treatment groups (CBT vs. usual care and research vs. practice settings), supporting the use of the TPOCS-RS as a measure of treatment differentiation in effectiveness research.

Although the TPOCS-RS was not originally designed as a measure of adherence, its design features may allow it to be used as such. First, it is scored using extensiveness ratings which have been described as more accurate for capturing adherence than alternative methods such as presence/absence or frequency (Hogue et al., 1996; McLeod et al., 2013). Second, the TPOCS-RS was expanded specifically to include practice elements that directly map onto the evidence-base for youth emotional and behavioral problems (e.g., behavioral activation for depression; parenting skills for behavior problems; McLeod et al., 2015). In fact, recent evidence has emerged suggesting that the TPOCS-RS is able to approximate protocol adherence to two manualized CBT treatments for youth anxiety (McLeod et al., 2021). In a sample of 796 treatment sessions of 55 youth with a primary anxiety disorder receiving a standard manualized CBT program, a modular manualized CBT program, or usual care, items in the TPOCS-RS Anxiety subscale (i.e., a subscale created to mirror the content found in the two CBT manuals) had interrater reliability scores ranging from ICC = .54 to ICC = .93 (*M* ICC = .80) in the standard manualized CBT group and ranging from ICC = .53 to ICC = .88 (*M* ICC = .74) in the modular manualized CBT group. Additionally, the TPOCS-RS Anxiety subscale displayed evidence of construct validity when used to estimate protocol adherence to two CBT protocols. Specifically, correlations between the TPOCS-RS Anxiety subscale and two protocol-specific adherence measures were medium for standard manualized CBT (*r* = .36) and large for modular manualized CBT (*r* = .46) based on Rosenthal and Rosnow's (1984) guidelines (correlations are "small" if $r \geq .10$, "medium" if $.24 \leq r < .37$, and "large" if $r \geq .37$), supporting convergent validity. Additionally, correlations between the TPOCS-RS Anxiety subscale and the remaining subscales (Psychodynamic, Family, Client-Centered) were small to medium in both the standard

manualized ($r = -.30$ to $r = .12$) and modular manualized ($r = -.25$ to $r = .01$) conditions,

supporting discriminant validity (McLeod et al., 2021). Overall, the TPOCS-RS scores displayed

evidence of reliability and construct validity when used to estimate protocol-adherence to two

separate manualized CBT treatments, indicating that the TPOCS-RS may be able to reliably

approximate protocol adherence.

Current Study

Considered together, evidence suggests that the TPOCS-RS may be suitable as a measure

of both adherence and differentiation and thus aid in the interpretation of implementation and

effectiveness trials. Current evidence is limited to samples of youth receiving CBT or usual care,

therefore this study aims to assess the ability of the TPOCS-RS to be used as a measure of

adherence to and differentiation from family therapy. The TPOCS-RS was used to code therapy

tapes from adolescents receiving treatment for externalizing problems assigned to one of three

treatment groups: usual care family therapy (UC-FT), usual care family therapy plus the

medication integration protocol (MIP), or nonfamily therapy usual care (UC-Other). Adherence

in the current study is defined as the delivery of core family therapy techniques. To assess

adherence, a Family Therapy subscale was generated that includes the TPOCS-RS items that are

considered to be core family therapy practice elements for adolescent externalizing problems

(e.g., multiparticipant interactions, parenting style). Differentiation is defined as the delivery of

techniques that are not considered core elements of family therapy, including both those that are

evidence-based for adolescent externalizing problems (e.g., CBT techniques such as functional

analysis of behavior and cognitive restructuring; Hogue et al., 2020) and those that are not (e.g.,

psychodynamic techniques; Chorpita et al., 2011). The existing TPOCS-RS subscales are well-

suited to capture non-evidence-based differentiation from family therapy (i.e., Psychodynamic,

Client-Centered). In addition, two subscales were created to assess evidence-based differentiation: (a) a non-family therapy (Non-FT) subscale that includes the family-focused TPOCS-RS items with evidence for treating adolescent externalizing problems that are not considered to be core elements of family therapy (e.g., operant strategies; McCart & Sheidow, 2016); and (b) a CBT subscale that includes the TPOCS-RS items that are considered to be core CBT practice elements for adolescent externalizing problems (e.g., functional analysis of behavior, cognitive distortion; Hogue et al., 2020). The creation of the TPOCS-RS Family Therapy, CBT, and Non-FT subscales is described in greater detail in the results section.

In order to produce scores that can aid in the interpretation of implementation and effectiveness trials, the TPOCS-RS should display certain psychometric characteristics. First, items included in the TPOCS-RS Family Therapy subscale should demonstrate adequate interrater reliability (i.e., consistency among ratings provided by multiple coders; ICC $\geq$ .40; Cicchetti, 1994; Hallgren, 2012), indicating that only a small portion of score variance is attributable to coders (rather than session content; Novick & Lewis, 1966).

Second, the TPOCS-RS Family Therapy subscale should display evidence of construct validity (i.e., the ability of a measure to accurately capture the construct it claims to assess; Foster & Cone, 1995). To evaluate the construct validity of the TPOCS-RS Family Therapy subscale, the current study used the multi-trait multi-method approach (Campbell & Fiske, 1959) in which the TPOCS-RS Family Therapy subscale was compared with measures of the same construct using multiple methods (i.e., observer- and therapist-rated adherence to family therapy techniques) providing evidence for convergent validity (Foster & Cone, 1995), and with measures of distinct constructs (i.e., treatment techniques outside of family therapy and client-therapist alliance) providing evidence for discriminant validity (Foster & Cone, 1995). More

specifically, to assess convergent validity, the TPOCS-RS Family Therapy subscale was compared with the corresponding Family Therapy subscales on the adherence measures used in the parent trials from which the coded tapes were drawn (i.e., the observer- and therapist-rated ITT-ABP). To assess discriminant validity, the TPOCS-RS Family Therapy subscale was compared with subscales representing other treatment modalities from both the TPOCS-RS (i.e., Non-FT, CBT, Psychodynamic, Client-Centered) and the observer- and therapist-rated ITT-ABP (i.e., CBT/MI) as well as a measure of client-therapist alliance. The multi-trait multi-method approach suggests that relations should be strongest between measures of the same construct and same method, followed by measures of the same construct and different methods, and weakest between measures of distinct constructs (Campbell & Fiske, 1959). In the current study, construct validity of the TPOCS-RS Family Therapy subscale is supported if the following pattern of correlations found in previous research are observed: (a) correlations are highest with measures of the same construct and same method (i.e., the corresponding observer-rated ITT-ABP Family Therapy subscale; Cecilione et al., 2021; Southam-Gerow et al., 2016); (b) second highest with measures of the same construct and different method (i.e., the corresponding therapist-rated ITT-ABP Family Therapy subscale; McLeod et al., 2021); (c) and lowest with measures of distinct constructs (i.e., the non-corresponding subscales of the TPOCS-RS, the observer- and therapist-rated ITT-ABP CBT/MI subscale, and a scale of client-therapist alliance; Carroll et al., 2000; Cecilione et al., 2021; Hogue, Dauber et al., 2008; McLeod et al., 2018; McLeod et al., 2021; Southam-Gerow et al., 2016). Additionally, construct validity is supported if the pattern of correlations produced by TPOCS-RS Family Therapy subscale scores are consistent with the pattern of correlations produced by the observer-rated ITT-ABP Family Therapy subscale scores.

Third, to be useful to researchers for interpreting results from implementation and effectiveness trials, the TPOCS-RS should be able to discriminate between treatment groups expected to differ in their delivery of therapeutic techniques (i.e., discriminative validity; Foster & Cone, 1995). To assess discriminative validity, study group differences in TPOCS-RS Family Therapy subscale scores were examined. Previous research has demonstrated that (a) UC-FT sessions included greater levels of family therapy techniques than both UC-Other sessions (Hogue, Dauber et al., 2014) and MIP sessions (Hogue et al., 2016) when coded using the ITTP-ABP; and (b) therapists at the clinic where the UC-FT and MIP groups were delivered demonstrated gold-standard levels of adherence to family therapy techniques based on inflation-adjusted self-reported ITT-ABP scores (Hogue et al., 2017). Therefore, discriminative validity is supported if TPOCS-RS Family Therapy subscale scores are highest for the UC-FT group, second highest for the MIP group, and lowest for the UC-Other group.

Finally, given that greater use of family therapy techniques has been shown to be related to clinical outcomes (Hogue et al., 2004; Hogue et al., 2006), to be useful to researchers attempting to interpret the effectiveness or lack thereof of family therapy, the TPOCS-RS Family Therapy subscale should predict the same pattern of clinical outcomes as established measures of adherence to family therapy (i.e., predictive validity; Cronbach & Meehl, 1955). Using the Family Therapy subscale of the ITT-ABP, Henderson et al. (2019) found that higher scores of adherence predicted decreases in adolescent-reported delinquency, externalizing behavior, and substance use, as well as parent-reported externalizing behavior. The current study examined the relation between the TPOCS-RS Family Therapy scores and the forementioned clinical outcomes and predictive validity is supported if higher TPOCS-RS Family Therapy scores significantly predict improvement in all four clinical outcomes.

Hypotheses

Hypothesis 1- Reliability

I hypothesize that interrater reliability will be at least "fair" (ICC $\geq$ .40; Cicchetti, 1994) for all TPOCS-RS items and subscales in the full sample.

Hypothesis 2- Convergent Validity

Hypothesis 2a. I hypothesize that in the full sample, the TPOCS-RS Family Therapy subscale scores will have a large correlation ($r \geq$ .37; Rosenthal & Rosnow, 1984) with the Family Therapy subscale of the observer-rated ITT-ABP.

Hypothesis 2b. I hypothesize that in the CASALEAP sample, correlations between the TPOCS-RS Family Therapy subscale and the Family Therapy subscale of the therapist-rated ITT-ABP will be medium (.24 $\leq r <$ .37; Rosenthal & Rosnow, 1984).

Hypothesis 3- Discriminant Validity

Hypothesis 3a. I hypothesize that in the full sample, correlations between the TPOCS-RS Family Therapy subscale and all other TPOCS-RS subscales (i.e., Non-FT, CBT, Psychodynamic, and Client-Centered) will be small-to-medium ($r <$ .37; Rosenthal & Rosnow, 1984) based on previous research demonstrating TPOCS-RS between-subscale correlations of these magnitudes (McLeod et al., 2015; McLeod et al., 2021).

Hypothesis 3b. I hypothesize that in the full sample, the correlation will be small-to-medium ($r <$ .37; Rosenthal & Rosnow, 1984) between the TPOCS-RS Family Therapy subscale and the CBT/MI subscale of the observer-rated ITT-ABP.

Hypothesis 3c. I hypothesize that in the CASALEAP sample, the correlation will be small ($r <$ .24; Rosenthal & Rosnow, 1984) between the TPOCS-RS Family Therapy subscale and the CBT/MI subscale of the therapist-rated ITT-ABP.

17

Hypothesis 3d. I hypothesize that the correlation will be small ($r < .24$; Rosenthal & Rosnow, 1984) between the TPOCS-RS Family Therapy subscale and a measure of client-therapist alliance (VTAS-R-SF) based on previous research assessing the correlation between the TPOCS-RS and measures of client-therapist alliance (McLeod et al., 2015; McLeod et al., 2021).

Hypothesis 4- Discriminative Validity

I hypothesize that the TPOCS-RS Family Therapy subscale scores will be significantly higher for the UC-FT group than for both the UC-Other group and the MIP group, and significantly higher for the MIP group than for the UC-Other group.

Hypothesis 5- Predictive Validity

I hypothesize that in the full sample, higher TPOCS-RS Family Therapy subscale scores will predict significant decreases in adolescent-reported delinquency, externalizing behavior, and substance use, as well as parent-reported externalizing behavior.

Method

Data Source

The current study uses data from a larger randomized naturalistic trial (CASALEAP; Hogue, Dauber, Henderson et al., 2015) and a small pilot study (CASALEAP-MIP, hereafter referred to as MIP; Hogue et al., 2016). The CASALEAP trial compared family therapy to nonfamily therapy to treat adolescent externalizing problems in routine practice settings. Youth with externalizing problems, including disruptive or delinquent behavior and substance use, were clinically referred to the research staff by high schools, family service agencies, and community programs serving youth in urban areas of a large northeastern city. Adolescents were randomly assigned to one of two groups: (a) Usual Care—Family Therapy (UC-FT); or (b) Usual Care—Other (UC-Other), which are described in greater detail below. Adolescents assigned to UC-FT

received treatment at a single community mental health clinic that routinely delivered non-manualized, structural-strategic family therapy for youth externalizing problems. Adolescents assigned to UC-Other received treatment at one of five treatment sites (including community mental health clinics, outpatient psychiatry clinics, and drug counseling centers) that did not routinely deliver family therapy. Youth assigned to the UC-FT group experienced greater reductions in youth-reported externalizing and internalizing symptoms, delinquency, and alcohol and drug use than youth in the UC-Other group.

The MIP trial evaluated a family-based protocol designed to integrate medication services into behavioral treatment planning for adolescents referred for primary externalizing problems (e.g., ODD, CD, SUD) with comorbid attention-deficit/hyperactivity disorder (ADHD). Youth were clinically referred to the research staff by high schools, family service agencies, and community programs serving youth in urban areas of a large northeastern city. Youth were also referred from the community mental health clinic that delivered non-manualized structural-strategic family therapy in the CASALEAP trial. All youth recruited for this pilot study were assigned to the Medication Integration Protocol (MIP) and were compared with a matched historical control (HC) group which consisted of adolescents who were assigned to the UC-FT group in the CASALEAP trial and met diagnostic criteria for ADHD. Youth in the MIP group were found to be more likely to complete a psychiatric evaluation and initiate medication for ADHD than youth in the HC group.

Participants

Youth Participants

CASALEAP. Youth were included in the CASALEAP trial if they (a) were between 12 and 18 years old; (b) had a primary caregiver who was willing to participate in treatment; (c) met

19

criteria for either the Mental Health (MH) or Substance Use (SU) track (described in more detail below); (d) were not enrolled in any other behavioral treatment; (e) expressed willingness to participate in treatment; and (f) had a health insurance plan that was accepted at the study treatment sites. Youth were excluded from the study if they (a) had an intellectual disability or autism spectrum disorder; (b) had a medical or psychiatric illness requiring hospitalization; (c) had current psychotic symptoms; or (d) had active suicidal ideation. A total of $N = 205$ youth were enrolled in the CASALEAP trial. Youth participant demographic information was presented by the parent study as separated by track and not by study group. However, it was noted that the only between-group differences in demographic information was that youth in the UC-Other group were more likely to have a household member involved in illegal activities.

Mental Health Track. Adolescent participants were eligible for the MH track if they met diagnostic criteria for either oppositional defiant disorder (ODD) or conduct disorder (CD) based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 2000); a total of $n = 130$ adolescents were eligible for the MH track. Participants in this track had a mean age of 15.3 years ($SD = 1.5$), 50% were male, 50% were female, 64% were Hispanic/Latinx/e, 20% were African American, 12% were multiracial, 5% had a racial/ethnic identity of "other race" and 50% had a primary caregiver with an income greater than $15,000. Adolescents met diagnostic criteria (as determined by research staff using the Mini International Neuropsychiatric Interview Version 5.0; Sheehan et al., 1998) for oppositional defiant disorder (88%), conduct disorder (44%), attention-deficit/hyperactivity disorder (73%), depression (40%), generalized anxiety disorder (13%), posttraumatic stress disorder (16%), and substance use disorder (5%), and 87% met criteria for multiple diagnoses.

Substance Use Track. Participants were eligible for the SU track if they (a) reported at least one day of alcohol use to intoxication or illegal drug use in the past 30 days or 30 days prior to living in a controlled environment; (b) endorsed at least one symptom of alcohol or substance dependence/abuse based on the DSM-IV-TR; and (c) met American Society on Addiction Medicine (2001) criteria for outpatient SU treatment. Adolescents who met criteria for both the MH and SU tracks were placed in the SU track. The SU track consisted of $n = 75$ participants who were 55% male, 45% female, and had an average age of 16.2 years ($SD = 1.3$). Self-reported race/ethnicity was 49% Hispanic/Latinx/e, 21% African American, 20% multiracial, and 9% "other race." Sixty-four percent of participants had a primary caregiver with an income greater than \$15,000. Diagnoses (as determined by research staff using the Mini International Neuropsychiatric Interview Version 5.0; Sheehan et al., 1998) included oppositional defiant disorder (85%), conduct disorder (68%), attention-deficit/hyperactivity disorder (75%), depression (45%), generalized anxiety disorder (24%), posttraumatic stress disorder (18%), and substance use disorder (69%), and 92% of participants met criteria for multiple diagnoses.

Current Study. A total of $N = 30$ youth from the CASALEAP trial were included in the current study (see Table 1). Youth were an average of 14.60 years old ($SD = 1.30$), were 50.0% male and 50.0% female, and were 66.7% Hispanic/Latinx/e, 13.3% African American, 13.3% Multiracial, and 6.7% other race. Sixty-three percent of participants had a primary caregiver with a yearly income greater than \$15,000.

The UC-FT group consisted of $n = 16$ youth (M age = 14.45, $SD = 1.37$; 50.0% male, 50.0% female; 75.0% Hispanic/Latinx/e, 12.5% Multiracial, 6.3% African American, 6.3% other race; 50.0% with a primary caregiver with an income greater than \$15,000) and the UC-Other group consisted of $n = 14$ youth (M age = 14.78, $SD = 1.23$; 50.0% male, 50.0% female; 57.1%

Hispanic/Latinx/e, 21.4% African American, 14.3% Multiracial, 7.1% other race; 78.6% with a primary caregiver with an income greater than $15,000).

MIP. Adolescents were eligible for inclusion in the MIP trial if they met the following criteria: (a) aged 13 to 17; (b) had a primary caregiver who was willing to participate in treatment; (c) met DSM-IV diagnostic criteria for ODD, CD, or substance use disorder (SUD); (d) met DSM-IV criteria for ADHD with or without onset prior to age 7; (e) were not prescribed medication for ADHD; and (f) were not enrolled in any behavioral treatment. Exclusion criteria included (a) bipolar disorder; (b) intellectual disability; (c) pervasive developmental disorder; (d) medical or psychiatric illness requiring hospitalization; (e) current psychotic symptoms; and (f) current suicidal ideation. A total of $N = 35$ participants were included in the MIP trial ($n = 14$ MIP and $n = 21$ HC). MIP participants were 50% male, 50% female, and had an average age of 15.3 years ($SD = 1.3$). Self-reported race/ethnicity was 50% Hispanic/Latinx/e, 43% African American, and 7% multiracial. Comorbid diagnoses included oppositional defiant disorder (93%), conduct disorder (64%), depression (29%), generalized anxiety disorder (14%), and substance use disorder (36%); multiple comorbid diagnoses were present in 93% of participants.

Current Study. A total of $N = 12$ youth from the MIP trial were included in the current study (see Table 1). Demographic information was not collected for two youth in the MIP trial. Youth were an average of 16.30 years old ($SD = 1.04$), were 41.7% male and 41.7% female, and were 41.7% Hispanic/Latinx/e, 33.3% African American, and 8.3% Multiracial. Caregiver income data were not reported for the MIP trial.

Table 1

Youth Descriptive Data and Group Comparisons

	M *(SD)* or %			
	UC-FT (*N* = 16)	UC-Other (*N* = 14)	MIP (*N* = 12)	*F* or χ2 *(p-value)*
Age	14.45 (1.37)	14.78 (1.23)	16.30 (1.04)	7.25 (.002)
Sex	-	-	-	0.00 (1.00)
Female	50.0	50.0	41.7	-
Male	50.0	50.0	41.7	-
Not reported	0.0	0.0	16.7	-
Race/Ethnicity	-	-	-	5.01 (.543)
Hispanic/Latinx/e	75.0	57.1	41.7	-
Black/African American	6.3	21.4	33.3	-
Multiracial	12.5	14.3	8.3	-
Other race	6.3	7.1	0.0	-
Not reported	0.0	0.0	16.7	-
Income >15k	50.0	78.4	-	2.63 (.142)

Note. UC = usual care, FT = family therapy, MIP = medication integration protocol

Therapist Participants

CASALEAP. A total of *N* = 34 therapists participated in the CASALEAP trial.

Therapists in the UC-FT group were all from a single community mental health clinic that

provided family therapy as the standard-of-care approach for youth externalizing problems.

Therapists in this group were marriage and family therapists, social workers with family therapy

training, or advanced trainees with family therapy experience. All therapists received regular

training and supervision from on-site supervisors experienced in family therapy to encourage

family-based case conceptualization and the use of family therapy treatment techniques. Seven

of the UC-FT therapists were female and seven were male. Seven identified as

Hispanic/Latinx/e, one identified and European American, and one identified as of other racial

background. Demographic information was not collected for five of the UC-FT therapists.

Therapist age ranged from 28 to 59 years and the average years of postgraduate therapy experience was 3.1 (*SD* = 4.3).

Therapists in the UC-Other group were from one of five outpatient clinics that did not routinely delivery family therapy. None of the UC-Other therapists or their supervisors were marriage and family therapists or completed postgraduate training in family therapy. Across sites, there were *n* = 20 participating therapists in the UC-Other group. Thirteen of the therapists were female and seven were male. Twelve identified as European American, three identified as Asian American, two identified as Hispanic/Latinx/e, and one identified as being of another racial background. Demographic information was not collected for two UC-Other therapists. Therapist age ranged from 24 to 45 years and the average years of postgraduate therapy experience was 3.2 (*SD* = 2.8).

Current Study. The current study included a total of 21 CASALEAP therapists (see Table 2) who had an average age of 33.65 (*SD* = 8.91), had an average of 3.20 years of post-degree therapy experience (*SD* = 3.49), were 19.0% male and 61.9% female, and were 28.6% Hispanic/Latinx/e, 19.0% Caucasian, 14.3% Asian American, 9.5% Multiracial, and 9.5% other race. Demographic information was not collected for four CASALEAP therapists.

The UC-FT group comprised *n* = 12 therapists (*M* age = 37.38, *SD* = 11.49; *M* years post-degree experience = 3.67, *SD* = 4.84; 8.3% male, 58.3% female; 41.7% Hispanic/Latinx/e, 8.3% Caucasian, 8.3% Multiracial, 8.3% other race). Demographic information was not collected for four UC-FT therapists. The UC-Other group comprised *n* = 9 therapists (*M* age = 30.33, *SD* = 4.12; *M* years post-degree experience = 2.89, *SD* = 2.52; 33.3% male, 66.7% female; 33.3% Caucasian, 33.3% Asian American, 11.1 % Hispanic/Latinx/e, 11.1% Multiracial, 11.1% other race).

MIP. There were a total of $N = 11$ therapists in the MIP trial ($n = 3$ MIP and $n = 8$ HC;

see Table 2), all of whom were from a single community mental health clinic that provided

family therapy as the standard-of-care for youth externalizing problems (the same clinic as in the

UC-FT group from the CASALEAP trial). Therapists delivering MIP ($n = 3$) were all female,

ranged in age from 29 to 32 years, and had an average of 3.0 years ($SD = 1.7$) of postgraduate

therapy experience. Two MIP therapists identified as Hispanic/Latinx/e and one identified as

European American. The current study will include all $n = 3$ therapists from the MIP group.

Table 2

Therapist Descriptive Data and Group Comparisons

	M *(SD)* or %			
	UC-FT ($N = 12$)	UC-Other ($N = 9$)	MIP ($N = 3$)	*F* or χ2 *(p-value)*
Age	37.4 (11.5)	30.3 (4.1)	30.3 (1.5)	1.90 (.180)
Sex	-	-	-	2.03 (.362)
Female	58.3	66.7	100	-
Male	8.3	33.3	0.0	-
Not reported	33.3	0.0	0.0	-
Race/Ethnicity	-	-	-	8.65 (.373)
Hispanic/Latinx/e	41.7	11.1	66.6	-
Caucasian	8.3	33.3	33.3	-
Asian	0.0	33.3	0.0	
Multiracial	8.3	11.1	0.0	-
Other race	8.3	11.1	0.0	-
Not reported	33.3	0.0	0.0	-
Years experience	3.7 (4.8)	2.9 (2.5)	3.0 (1.7)	0.98 (.907)

Note. UC = usual care, FT = family therapy, MIP = medication integration protocol, years experience = years post-degree therapy experience

Treatment Groups

CASALEAP

Across both treatment groups and all six treatment sites, therapists were not provided

with any external training or financial support and were not required to adjust their usual clinical

practices in any way. All sites routinely prescribed weekly treatment sessions, and time spent in weekly supervision was comparable across sites.

Usual Care—Family Therapy. The UC-FT group consisted of one community mental health clinic that routinely delivered family therapy for externalizing youth problems; the site did not use a specific manualized family therapy. To ensure that the routine care delivered at the UC-FT clinic adhered to family therapy principles, Hogue and Dauber (2013) used observational adherence benchmarking analyses comparing 15 randomly selected therapy tapes videotaped prior to the onset of the CASALEAP trial to adherence scores from an efficacy trial of MDFT (Liddle & Hogue, 2001). Results indicated that the therapists delivering family therapy at the community mental health clinic closely adhered to gold-standard levels of adherence produced by therapists in the controlled MDFT efficacy trial. Previous research on the CASALEAP study sample found that UC-FT therapists reported the strongest allegiance and skill in FT techniques (Hogue, Dauber et al., 2014). Further, UC-FT therapists reported utilizing FT techniques more than CBT, MI, or DC (Hogue, Dauber et al., 2014).

Usual Care—Other. The UC-Other group consisted of five clinics that represent an array of outpatient treatment settings for adolescent externalizing problems, including two community mental health clinics with organizational profiles similar to the UC-FT clinic, two outpatient child psychiatry clinics in teaching hospitals, and one addictions treatment clinic that had an adolescent program. None of the five clinics included staff therapists or supervisors who were trained or experienced in family therapy. Previous research on this study sample revealed that UC-Other therapists reported the strongest allegiance and skill in CBT and MI, and UC-Other therapists reported greater use of CBT, MI, and DC than UC-FT therapists, while UC-FT therapists reported greater use of FT than UC-Other therapists (Hogue, Dauber et al., 2014). The

current study will not include any sessions from the addictions treatment clinic because therapists at that clinic were unable to record treatment sessions.

MIP

Participants in the MIP group received treatment at a single community-based mental health clinic that featured family therapy as their routine standard of care for adolescent externalizing problems (the same clinic that treated UC-FT cases in the CASALEAP trial).

Medication Integration Protocol. The Medication Integration Protocol (MIP; Hogue & Bobek, 2013; Hogue, Bobek et al., 2014) is a modular family-based protocol that integrates medication services into behavioral treatment for adolescents with ADHD. The protocol contains five MIP Tasks, four of which can be delivered in the order that is clinically indicated (Task 1 always occurs first). Task 1 is *ADHD Assessment and Medication Consult*, in which the therapist consults with the medication prescriber to confirm a diagnosis of ADHD and the adolescent's eligibility for medication. In Task 2, *ADHD Psychoeducation and Client Acceptance*, the therapist reviews educational materials with the family that promote discussions about multiple aspects of ADHD, including common behavioral symptoms and impacts on socioemotional development and academic performance, prevalence rates, and the neurobiology of ADHD and how it relates to executive functioning and learning difficulties. The goals of Task 2 are for the therapist to promote acceptance of the ADHD diagnosis, establish realistic expectations for change, educate the family about the benefits and potential side effects of ADHD medication, and identify family-endorsed treatment goals. In Task 3, *ADHD Symptoms and Family Relations,* the therapist aims to help the family to understand how the adolescent's specific presentation of ADHD impacts their family, school, and peer functioning. The goals of Task 3 include the therapist helping the family alter negative attributions about ADHD-related behaviors, redefining

the adolescent referral problem as a family problem with family-based solutions, assessing the home environment to determine characteristics that promote or impede academic and treatment success, and assessing family readiness to make changes. In Task 4, *Medication and Family Decision Making,* the therapist educates the family about the potential benefits of medication in home, school, and peer environments, and discusses medication stigma, side effects, titration, substance misuse, and potential medication diversion. During Task 4, the family accepts, refuses, defers, or is determined to be ineligible for medication. Task 5, *Medication Management and Integration Planning*, involves therapist formulation of a case coordination framework for medication management that is tailored to each family. In the case coordination framework, the therapist and medication prescriber work together to support medication adherence and monitor benefits and side effects. The medication prescriber also arranges ongoing medication management visits and the therapist routinely checks in about medication issues during therapy sessions.

Measures

Treatment Process Measures

Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale (TPOCS-RS; McLeod et al., 2015; McLeod et al., 2021). The TPOCS-RS is a 47-item observational measure designed to assess the delivery of therapeutic techniques across five theory-based subscales: Family (8 items; e.g., "Multiparticipant Interaction"), Cognitive (4 items; e.g., "Cognitive Distortion"), Behavioral (9 items; e.g., "Behavioral Activation"), Psychodynamic (5 items; e.g., "Interpretation"), and Client-Centered (4 items; e.g., "Positive Regard"). The TPOCS-RS also contains 17 general items which represent common therapeutic techniques that are not associated with a specific treatment modality (e.g., "Homework"). For the

purposes of the current study, only the 30 items found across the five theory-based subscales were coded and included in analyses. Coders rate the degree to which therapists deliver each technique during an entire session on a 7-point extensiveness scale where 1 = *not at all*, 4 = *considerably*, and 7 = *extensively*. When coders rate extensiveness they consider both thoroughness (the depth, complexity, or persistence with which the therapist engages in a technique) and frequency (how often the therapist delivers the technique). The TPOCS-RS has demonstrated evidence of interrater reliability with mean ICCs ranging from .67 to .86 (McLeod et al., 2015; McLeod et al., 2021; Smith et al., 2017; Southam-Gerow et al., 2016), convergent validity (McLeod et al., 2021; Southam-Gerow et al., 2016), discriminant validity (McLeod et al., 2015; McLeod et al., 2021; Southam-Gerow et al., 2016), and discriminative validity (McLeod et al., 2021; Smith et al., 2017). The TPOCS-RS was used to code UC-FT, UC-Other, and MIP sessions in the current study.

Inventory of Therapy Techniques-Adolescent Behavior Problems (ITT-ABP; Hogue, Dauber et al., 2014; Hogue, Dauber, Lichvar et al., 2015). The ITT-ABP is a 25-item measure designed to capture the delivery of discrete treatment techniques associated with family therapy (FT), cognitive behavioral therapy (CBT), motivational interviewing (MI), and drug counseling (DC) approaches. Items associated with FT (e.g., arranged, coached, and/or helped process interactions among family members in session) and CBT (e.g., taught client new problem-solving, coping, and communication skills) were drawn from the Therapist Behavior Rating Scale (TBRS; Hogue et al., 1998) and items associated with MI (e.g., affirmed client's self-efficacy and ability to change problematic behavior, and praised client's change efforts) and DC (discussed cravings, triggers, and high-risk situations that lead to current or future drug use) were drawn from the Motivational Enhancement Therapy and Twelve Step Facilitation subscales

of the Yale Adherence and Competence Scale (YACS; Carroll et al., 2000). Exploratory and confirmatory factor analysis revealed a three-factor solution, resulting in the combination of the CBT and MI items into a single CBT/MI subscale, a FT subscale, and a DC subscale (Hogue, Dauber et al., 2014). The ITT-ABP assesses the extensiveness (i.e., thoroughness and frequency) of technique delivery based on a 5-point Likert-type scale: 1 = *not at all*, 2 = *a little bit*, 3 = *moderately*, 4 = *considerably*, and 5 = *extensively*. The ITT-ABP includes an observer-rated version and a therapist-rated version, both of which were included in the CASALEAP trial. Only the observer-rated version was used in the MIP trial. Both versions were included in the current study.

Interrater reliability between multiple independent observers was ICC = .87 for the FT subscale, ICC = .76 for the CBT/MI subscale, and was not reported for the DC subscale. Agreement between independent observers and therapists was ICC = .66 for the FT subscale, ICC = .06 for the CBT/MI subscale, and was not reported for the DC subscale (Hogue, Dauber, Lichvar et al., 2015). Further, the observer- and therapist-rated FT subscale has demonstrated evidence of convergent, discriminant, discriminative, and predictive validity (Henderson et al., 2019; Hogue, Dauber et al., 2014; Hogue, Dauber, Lichvar et al., 2015). Due to lack of variance in the DC subscale stemming from the inability to collect session recordings from the only clinic specializing in substance use and addiction treatment, the DC subscale was not included in analyses for the current study (Hogue, Dauber, Lichvar et al., 2015).

Vanderbilt Therapeutic Alliance Scale Revised Short Form (VTAS-R-SF; Shelef & Diamond, 2008). The VTAS-R-SF was used to assess client-therapist alliance in the CASALEAP and MIP trials. The VTAS-R-SF is a 5-item observer-rated measure designed to measure the collaborative and task-oriented working relationship between the client and the

therapist. The VTAS-R-SF is rated on a 6-point scale ranging from 0 = *not at all* to 5 = *a great deal*. The VTAS-R has demonstrated internal consistency ranging from α = .93 to α = .98 and interrater agreement ranging from ICC = .80 to ICC = .93 (Diamond et al., 1999). The VTAS-R-SF has also demonstrated convergence with the full-length VTAS-R (Shelef & Diamond, 2008).

Clinical Outcome Measures

In the CASALEAP trials, clinical outcome measures were administered at baseline and 3-, 6-, and 12-month follow-up. In the MIP trial, outcome measures were administered at baseline, 3-, and 6-month follow-up.

Child Behavior Checklist and Youth Self Report (CBCL and YSR; Achenbach & Rescorla, 2001). Youth externalizing and internalizing symptoms were measured using the CBCL and YSR which were rated by caregivers and adolescents, respectively. The CBCL and YSR are measures of youth emotional and behavioral symptoms with well documented evidence of reliability and validity (Achenbach et al., 2003; Achenbach & Rescorla, 2001) and include both internalizing and externalizing summary scales as well as the following syndrome scales: (a) Anxious/Depressed; (b) Depressed; (c) Somatic Complains; (d) Social Problems; (e) Thought Problems; (f) Attention Problems; (g) Rule-Breaking Behavior; and (h) Aggressive Behavior. The following DSM-oriented scales are also included: (a) Affective Problems; (b) Anxiety Problems; (c) Somatic Problems; (d) ADHD; (e) Oppositional Defiant Problems; and (f) Conduct Problems. Caregivers and youth rate youth symptoms on a three-point scale where 0 = *not true*, 1 = *somewhat or sometimes true*, and 2 = *very true or often true*. Analyses for the current study included only the externalizing summary scale scores.

National Youth Survey Self-Report Delinquency Scale (SRD; Elliott et al., 1985). Adolescent delinquent behaviors were assessed using the National Youth Survey Self-Report

Delinquency Scale, a 23-item self-report scale that assesses the frequency of delinquent behavior within five subscales: (a) Total Delinquency; (b) General Theft; (c) Crimes Against Persons; (d) Index Offenses; and (e) Drug Scales. Each item was rated on a 9-point Likert-type scale ranging from 1 = *never* to 9 = *2-3 times per day*. This measure has been widely used in research with adolescent clinical samples and is considered to be well-validated (Huizinga & Elliott, 1986; Piquero et al., 2002). Analyses for the current study used a total scale score.

Timeline Follow Back (TLFB; Sobell & Sobell, 1996). Adolescent substance use was measured with the TLFB method which assess adolescent-reported frequency and quantity of daily substance use using a calendar and memory aids to gather retrospective estimates. The TLFB method has demonstrated reliability and validity when used to assess alcohol and illicit drug use (Sobell & Sobell, 1996) and has been shown to be sensitive to change in substance use (e.g., Liddle et al., 2008; Robbins et al., 2011; Waldron et al., 2001). In the parent trials, adolescents reported the number of days they had engaged in any alcohol or illicit drug use for each month in the study period. For the current study, the TLFB was scored dichotomously as abstinent/not abstinent.

Study Procedures

Coding Procedures

The TPOCS-RS was coded by two female clinical psychology doctoral students, ages 28 and 29, who both identify as White. Coder training consisted of four phases. First, coders received didactic instruction by reading and discussing the coding manual. Second, coders coded 5 training sessions together and discussed discrepancies in real time. Third, coders coded 10 training sessions independently and participated in weekly meetings, during which the coding results were discussed. Across the first three phases, coders could discuss coding questions with

the TPOCS-RS creator (Dr. Bryce D. McLeod). Fourth, coders entered the certification phase, during which they were required to reach adequate item-level reliability, ICC (2,2) > .59 (Cicchetti, 1994), across 30 independently coded recordings. Training and certification tapes were selected to ensure variability across items and subscales.

Upon completion of the certification phase, coders independently coded randomly assigned sessions; both coders were naïve to treatment group and coded sessions in a randomly assigned order. Coder drift (Margolin et al., 1998) was prevented in two ways. First, coders met regularly during the independent-coding process to discuss any questions. Second, interrater reliability was examined on an ongoing basis to ensure between-coder reliability. If item-level reliability fell below the acceptable level (ICC < .60) for any item, additional training occurred, which included reexamination of the coding manual by both coders, group discussions about coding discrepancies, and co-coding of challenging items. Coders scored entire therapy sessions and every session was double-coded. To reduce measurement error, item scores were averaged across the two coders. To minimize missing TPOCS-RS data, code sheets were routinely examined for missing ratings and if an item was erroneously left unscored, the coder was asked to recode the video; therefore, there were no missing scores for the TPOCS-RS.

Sampling of Treatment Sessions

In the CASALEAP trial, 444 sessions were targeted for audio or video recording, and 157 session recordings were collected (35%). Of the 287 sessions targeted for recording that were not collected, 41% were from families who did not consent to allow recordings, 26% were from cases treated in a UC-Other clinic that did not allow recording, and 33% were not recorded due to technological or logistical challenges. Recordings were 66% UC-FT sessions (28 cases seen

by 14 therapists) and 34% UC-Other sessions (17 cases seen by 11 therapists), and included both early treatment (sessions 1-5; 44%) and late treatment (sessions 6 or higher; 56%).

In the MIP trial, a total of 119 video recorded sessions were selected for coding (80 MIP and 39 HC) from 21 cases (12 MIP and 9 HC). For each case, two sessions were randomly selected from available recordings from the early phase of treatment (sessions 2-5; 65%) and one session was selected from the later phase (session 6 or later; 35%).

For the current study, a total of $N = 103$ sessions were coded, including 54 CASALEAP (35 UC-FT and 19 UC-Other) sessions and 49 MIP sessions randomly selected from the archived pool of 157 CASALEAP sessions and 119 MIP sessions. Included tapes come from $N = 42$ cases ($n = 16$ UC-FT, $n = 14$ UC-Other, $n = 12$ MIP) treated by $N = 24$ therapists ($n = 12$ UC-FT, $n = 9$ UC-Other, $n = 3$ MIP). The number of cases per therapist ranged 1 to 9 ($M = 2.23$) in the full sample, 1 to 4 ($M = 1.45$) in UC-FT, 1 to 3 ($M = 1.56$) in UC-Other, and 1 to 9 ($M = 3.67$) in MIP. The number of sessions per case ranged from 1 to 10 in the full sample ($M = 2.65$), 1 to 7 ($M = 2.13$) in UC-FT, 1 to 2 ($M = 1.36$) in UC-Other, and 2 to 10 ($M = 4.45$) in MIP.

Results

Preliminary Analyses

Sample Bias

Youth Participants. The current study included 30 of the 205 youth participants from the CASALEAP parent study and 12 of the 15 youth participants in the MIP parent study. The 175 CASALEAP youth who were not included in the current study did not differ from the included sample in sex (χ^2 [1, $N = 205$] = 0.04, $p = .839$), or race and ethnicity (χ^2 [3, $N = 205$] = 1.32, $p = .724$); however, non-included youth were significantly older ($M = 15.84$, $SD = 1.40$), than included youth ($M = 14.60$, $SD = 1.30$, t[203] = 4.50, $p < .001$). This indicates that the

included CASALEAP sample largely represents the full CASALEAP sample, but may underrepresent older adolescents. The three MIP youth who were not included in the current study did not differ from the included sample in sex (χ^2 [1, N = 13] = 0.26, p = .612), race and ethnicity (χ^2 [2, N = 13] = 0.45, p = .800), or age (t[11] = -0.13, p = .898), suggesting that the included MIP sample is representative of the full MIP sample.

To identify differences in youth characteristics across treatment groups in the current study, youth-level demographics were compared between the UC-FT, UC-Other, and MIP groups (see Table 1). There were no differences in youth sex (χ^2 [2, N = 40] = 0.00, p = 1.00) or race and ethnicity (χ^2 [6, N = 40] = 5.01, p = .543), but there were significant differences in age (F[2, 37] = 7.25, p = .002). Due to multiple comparisons, unequal group sizes and homogeneity of variances, post hoc comparisons between the three treatment groups were conducted using a Bonferonni test (McHugh, 2011). Post-hoc analyses revealed that youth in the MIP group (M = 16.30, SD = 1.04) were significantly older than youth in both the UC-FT group (M = 14.45, SD = 1.37, p = .002, 95% C.I. = [0.60, 3.12]) and the UC-Other group (M = 14.78, SD = 1.23, p = .016, 95% C.I. = [0.23, 2.82]). There were no significant differences in age between the UC-FT and UC-Other groups (p = 1.00, 95% C.I. = [-1.47, 0.82]). This suggests that the MIP sample is representative of an older population than the CASALEAP sample.

Therapist Participants. The current study included 21 of the 34 therapists from the CASALEAP parent study. All three therapists from the MIP parent trial were included in the present study. The included CASALEAP therapists did not differ from the excluded therapists in sex (χ^2 [1, N = 30] = 0.78, p = .376), race and ethnicity (χ^2 [4, N = 30] = 5.23, p = .265), age (t[28] = -0.85, p = .404), or years of post-degree therapy experience (t[26] = -0.42, p = .678). This suggests that included therapists are representative of therapists from the parent studies.

To examine differences in therapist characteristics across treatment groups in the current study, therapist-level demographics were compared between the UC-FT, UC-Other, and MIP groups (see Table 2). Groups did not differ in therapist sex (χ^2 [2, $N = 20$] = 2.03, $p = .362$), race and ethnicity (χ^2 [8, $N = 20$] = 8.65, $p = .373$), age (F[2, 17] = 1.90, $p = .180$), or years of post-degree therapy experience (F[2, 15] = 0.98, $p = .907$), indicating that therapists from the three treatment groups represent similar populations.

Missing Data

There were no missing TPOCS-RS data for coded sessions. Patterns of missing data were assessed for the observer- and therapist-rated ITT-ABP, VTAS-R-SF, ATAS, CBCL/YSR, SRD, and TLFB following the procedures recommended by Schafer and Graham (2002). Little's Missing Completely at Random (MCAR) test was used to identify if missing data exhibited an identifiable pattern. Specifically, Little's MCAR test reveals if data are missing completely at random (MCAR; i.e., not a function of any observed or unobserved variables), missing at random (MAR; i.e., a function of other observed variables in the dataset), or missing not at random (MNAR; i.e., a function of the variable being examined; Schafer & Graham, 2002).

Separate tests were run for variables collected at the session level (i.e., ITT-ABP observer and therapist report, VTAS-R-SF) and at the client level (i.e., TLFB, SRD, YSR, CBCL). Data were considered missing completely at random for both the session-level variables (Little's MCAR test Chi Square = 56.81, $df = 41$, $p = .051$) and the client-level variables (Little's MCAR test Chi Square = 347.51, $df = 352$, $p = .558$), suggesting that missing data are a random sample of all data and are independent of observed and non-observed variables.

Data Distribution

Data distribution was examined for the full sample and separately for each of the three groups (UC-FT, UC-Other, and MIP). Means, standard deviations, ranges, and normality were explored for TPOCS-RS items and subscales (i.e., Family Therapy, Non-FT, CBT, Psychodynamic, Client-Centered; subscale generation is described below). Normality was determined by evaluating skewness and kurtosis and determining if they fall outside the range of a normal distribution (i.e., -2 to 2; George & Mallery, 2016). For non-normal distributions, outliers were identified (i.e., *z*-score > 3.00; Tabachnik & Fidell, 2007) and considered for removal.

In the full sample, item-level ranges varied notably based on the treatment orientation with which the items are associated (see Table 3). The full possible range of scores was 6 points (extensiveness was rated from 1 to 7). Cognitive items ranged from 0 to 1 (*M* range = 0.50, *SD* = 0.50), behavioral items ranged from 0 to 3.5 (*M* range = 1.69, *SD* = 1.10), psychodynamic items ranged from 2 to 4 (*M* range = 3.13, *SD* = 0.85), family items ranged from 2 to 6 (*M* range = 4.19, *SD* = 1.65), and client-centered items all had a range of 4.5. The subscale with the highest range was Client-Centered (3.67), followed by Family Therapy (2.80), Non-FT (2.50), Psychodynamic (1.75), and CBT (0.90). Findings suggest that extensiveness scores displayed a limited range for cognitive and behavioral techniques, approximately half the full range for psychodynamic techniques, and close to the full range for family and client-centered techniques.

All items were positively skewed except *Client Perspective,* which was negatively skewed. Item-level skewness ranged from -0.60 to 10.15, and item-level kurtosis ranged from -1.36 to 103. Skewness and kurtosis fell outside the range of a normal distribution (i.e., -2 to 2; George & Mallery, 2016) for 17 out of 25 items in the full sample. Of note, items with a normal

distribution (*Explores Past, Targets Others, Family Roles, Validates Client, Positive Regard, Client-Perspective*) were observed as occurring between 49 and 103 times ($M = 85$, $SD = 20.95$), whereas items with a non-normal distribution were observed as occurring between 1 and 49 times ($M = 15.29$, $SD = 15.19$). All subscales were positively skewed, and two subscales (Non-FT and CBT) had non-normal distributions. Data (items and subscales) were examined for outliers (z-score > 3.00). There were an average of 1.78 outliers per TPOCS-RS item and an average of 1.60 outliers per TPOCS-RS subscale. For both items and subscales, outliers were determined to be scores at the high end of the range rather than data error, so they were not removed. For descriptive statistics by treatment group, see Appendix A.

Table 3

TPOCS-RS Item and Subscale Descriptive Data and Interrater Reliability in the Full Sample

Item	Full Sample (N = 103)					
	N	Range	*M(SD)*	ICC	S	K
Cognitive Ed*	1	0.5	1.01 (0.05)	NV	10.15	103.00
Cognitive Distortion	4	1.0	1.03 (0.15)	.797	5.60	31.76
Coping Skills*	0	0.0	1.00 (0.00)	NV	-	-
Functional Analysis	4	2.5	1.04 (0.15)	.901	7.88	67.68
Relaxation	1	1.5	1.02 (0.15)	.889	10.15	103.00
Respondent*	0	0.0	1.00 (0.00)	NV	-	-
Operant – Child*	4	1.0	1.02 (0.13)	.557	5.87	37.10
Skill Building	25	3.5	1.31 (0.67)	.913	2.57	7.19
Behavioral Activation*	3	1.0	1.02 (0.12)	NV	6.79	49.16
Monitoring	5	1.5	1.04 (0.19)	.750	5.87	37.51
Modeling*	24	2.5	1.24 (0.50)	.715	2.23	4.86
Transference	3	2.0	1.03 (0.22)	.761	7.62	62.07
Explores Past	49	4.0	1.72 (1.02)	.789	1.41	1.13
Client Resistance	12	3.0	1.12 (0.42)	.767	4.69	25.89
Interpretation	49	3.5	1.47 (0.66)	.716	1.84	4.19
Targets Others	72	5.0	3.08 (1.74)	.887	0.16	-1.36
Recruits Others	34	2.0	1.29 (0.48)	.756	1.71	2.41
Parenting Style	19	3.0	1.23 (0.61)	.809	3.33	11.57
Operant – Parent	10	2.0	1.08 (0.29)	.772	4.78	25.72
Parenting Skills	19	4.5	1.19 (0.57)	.808	5.03	33.28
Multiparticipant Interactions	43	5.0	1.70 (1.14)	.856	1.96	3.39
Family Roles	89	6.0	2.82 (1.35)	.759	0.54	-0.06
Validates Client	102	4.5	2.94 (0.90)	.650	0.41	-0.27
Positive Regard	95	4.5	2.40 (0.92)	.655	0.93	1.05
Client Perspective	103	4.5	4.21 (1.07)	.727	-0.60	-0.09
FT Subscale	-	2.8	2.02 (0.70)	.898	0.23	-0.93
Non-FT Subscale	-	2.5	1.14 (0.34)	.848	4.08	22.27
CBT Subscale	-	0.9	1.09 (0.17)	.912	2.41	6.31
Psych Subscale	-	1.8	1.34 (0.41)	.813	1.51	1.97
CC Subscale	-	3.7	3.19 (0.68)	.764	0.53	0.40

Note. N = the number of times an item was observed as occurring during a session by at least one coder; ICC = intraclass correlation coefficient; S = skewness; K = kurtosis; NV = ICCs not calculated due to lack of variance; *Not included in TPOCS-RS subscales used in present study

Interrater Reliability

The interrater reliability of all included TPOCS-RS items and subscales (Family Therapy, Non-FT, CBT, Psychodynamic, and Client-Centered) was examined using intraclass correlation coefficients (ICCs). The model ICC (2,2) based on a two-way random effects model was used because it provides a reliability estimate of the mean score of multiple coders and allows for the generalization of findings to other samples (Koo & Li, 2016; Shrout & Fleiss, 1979). Following the recommendations set forth by Cicchetti (1994), ICCs below .40 are considered *poor*, between .40 and .59 are considered *fair*, between .60 and .74 are considered *good*, and .75 and above are considered *excellent*. ICCs were calculated for the full sample (see Table 3) and for each treatment group separately (see Appendix A). It was hypothesized that TPOCS-RS Family Therapy item- and subscale-level interrater reliability would be at least "fair" (ICC $\geq$ .40; Cicchetti, 1994) for the full sample.

In the full sample, item-level ICCs ranged from .56 to .91 (*M* ICC = .77, *SD* = .09) and all but one item (*Operant – Child*) displayed at least *good* interrater reliability. *Operant* – Child (ICC = .56) displayed *fair* reliability; of note, this item was not included in any of the TPOCS-RS subscales. For the TPOCS-RS Family Therapy subscale, ICC = .90 in the full sample, ICC = .81 in the UC-FT group, ICC = .88 in the UC-Other group, and ICC = .89 in the MIP group. ICCs for the other TPOCS-RS subscales (i.e., Non-FT, CBT, Psychodynamic, Client-Centered) in the full sample ranged from .76 to .91 (*M* ICC = .83, *SD* = .06). All TPOCS-RS subscales displayed *excellent* reliability. ICCs could not be calculated for four items (*Cognitive Education, Coping Skills, Respondent, Behavioral Activation*) due to a lack of coder variance. Findings suggest that the TPOCS-RS items can be coded reliably by independent observers.

Subscale Generation

Five subscales were generated for the TPOCS-RS (Family Therapy, Non-FT, CBT, Psychodynamic, Client-Centered). Subscale scores were calculated by first averaging items across coders and then averaging all items in each subscale, consistent with the scoring method used for the ITT-ABP subscales (Hogue, Dauber, Lichvar et al., 2015). To examine the ability of the TPOCS-RS to estimate adherence to family therapy, a specific Family Therapy (FT) subscale was created (see Table 4). Although the TPOCS-RS contains a Family Focus subscale, not all items are exclusive to the family therapy treatment modality. For example, the item "operant strategies—parent" refers to techniques that aim to change a child's behavior by manipulating environmental contingencies, which is a behavioral technique (Graziano & Diament, 1992). An item was included in the Family Therapy subscale if it (a) is considered a core element of family therapy based on the literature (Hogue et al., 2017; Hogue et al., 2019; Hogue et al., 2021) and expert consultation; (b) is not considered a core element of another treatment modality represented by the TPOCS-RS subscales (i.e., CBT, Psychodynamic, or Client-Centered); and (c) had a full sample ICC of at least .40 (Cicchetti, 1994).

Table 4

TPOCS-RS Family Therapy Subscale

Item	Description
Targets other participants	Clients other than the target child are asked to modify their affect, behavior, or cognitions or when a Cognitive, Behavioral, Psychodynamic, or Client-Centered item is directed at someone besides the target child.
Recruits others	Attempts to recruit/retain parents and other supports (e.g., family members) for future sessions.
Parenting style	Helps and/or encourages parents to modify their parenting practices with a goal of improving the parent-child relationship.
Multiparticipant interactions	Sets-up, encourages, or processes interactions in the therapy session between participants.
Family members' roles	Therapist teaches or emphasizes how emotional/behavioral problems may be caused or maintained by the environmental context/family dynamics, etc.

Note. Item descriptions are from the unpublished TPOCS-RS scoring manual (McLeod, 2014)

The current study assessed the ability of the TPOCS-RS to measure both evidence-based and non-evidence-based differentiation from family therapy. As such, two subscales were created to assess evidence-based differentiation: (a) a Non-FT subscale that includes the family-focused TPOCS-RS items with evidence for treating adolescent externalizing problems that are not considered to be core elements of family therapy (see Table 5); and (b) a CBT subscale that includes the TPOCS-RS items that are considered to be core CBT practice elements for adolescent externalizing problems (Hogue et al., 2020; see Table 6). An item was included in the Non-FT subscale if (a) it is a family-based technique with evidence for treating adolescent externalizing problems based on the literature (e.g., McCart & Sheidow, 2016) and expert consultation but is not considered a core element of family therapy; (b) it is not considered a core element of another treatment modality represented by the TPOCS-RS subscales (i.e., Family

Therapy, CBT, Psychodynamic, or Client-Centered); and (c) had a full sample ICC of at least .40

(Cicchetti, 1994).

Table 5

TPOCS-RS Non-FT Subscale

Item	Description
Operant strategies—parent	Teaches principles of operant interventions (e.g., positive reinforcement, response cost, behavior reward system), sets up operant system and/or employs operant intervention.
Parenting skills	Parent-focused interventions are designed to improve the ability to set limits, ensure adequate supervision, and monitor child behavior (appropriate or inappropriate), activities, and involvements.

Note. Item descriptions are from the unpublished TPOCS-RS scoring manual (McLeod, 2014)

An item was included in the CBT subscale if (a) it was considered a core element of CBT

for adolescent externalizing problems based on the literature (Hogue et al., 2020) and expert

consultation; (b) it was not considered a core element of another treatment modality represented

by the TPOCS-RS subscales (i.e., Family Therapy, Non-FT, Psychodynamic, or Client-

Centered); and (c) had a full sample ICC of at least .40 (Cicchetti, 1994). Three items initially

considered for inclusion in the CBT subscale were ultimately excluded. One item was not

observed as occurring in any of the sessions (*Coping Skills*), and two items (*Cognitive Education*

and *Behavioral Activation)* had no coder variance. Therefore, they did not meet the criteria of

having a full sample ICC of at least .40.

Table 6

TPOCS-RS CBT Subscale

Item	Description
Functional analysis of behavior	Performs functional analysis of a target behavior/Teaches A-B-C (Antecedent-Behavior-Consequences) model.
Monitoring	The therapist teachers and/or encourages a child to monitor his/her behavior, affect, or cognitions.
Cognitive distortion	Teaches and/or encourages the client to identify and/or restructure cognitive distortions.
Relaxation strategies	Teaches/encourages client to use relaxation, meditation, or pleasant mental imagery.
Skill building	Interventions that focus upon building behavioral skills (e.g., goal attainment, social skills, assertiveness training).

Note. Item descriptions are from the unpublished TPOCS-RS scoring manual (McLeod, 2014)

To assess non-evidence-based differentiation from family therapy, two existing TPOCS-RS subscales were used: Psychodynamic (see Table 7) and Client-Centered (see Table 8). All items in these subscales were retained, as they had full sample ICCs of at least .40 (Cicchetti, 1994).

Table 7

TPOCS-RS Psychodynamic Subscale

Item	Description
Addresses transference	Discusses or interprets the client's interaction with the therapist.
Explores past	Discusses client's past experiences prior to treatment.
Addresses client resistance	Explicitly identifying, processing and/or discussing client's resistance to therapy and/or resistance to change.
Interpretation	Comments on client behavior and/or relates that behavior to an aspect of the client's characteristics, general functioning, and/or past experiences.

Note. Item descriptions are from the unpublished TPOCS-RS scoring manual (McLeod, 2014)

Table 8

TPOCS-RS Client-Centered Subscale

Item	Description
Validates client	Validates client's feelings and/or treatment goals.
Positive regard	Responds to client in warm and compassionate manner.
Client perspective	Attempts to understand client's point of view/Probes for client's unique perspective.

Note. Item descriptions are from the unpublished TPOCS-RS scoring manual (McLeod, 2014)

Primary Analyses

Correlation magnitudes are interpreted based on Rosenthal and Rosnow's (1984) guidelines (correlations are "small" if $r \geq .10$, "medium" if $.24 \leq r < .37$, and "large" if $r \geq .37$). The majority of correlations described below were computed in the full sample ($N = 103$); however, since therapists did not rate the ITT-ABP in the MIP trial, correlations between the TPOCS-RS Family Therapy subscale and the therapist-rated ITT-ABP subscales (FT and CBT/MI) were computed only in the CASALEAP sample ($N = 54$ sessions). Additionally, follow-up contrasts using Fisher *r*-to-*z* transformation were calculated for the absolute value of correlations primarily in the full sample. Correlations between the TPOCS-RS Family Therapy subscale and the therapist-rated ITT-ABP FT or CBT/MI subscales were compared to correlations between the TPOCS-RS Family Therapy subscale and all other subscales and scales conducted in the CASALEAP sample only.

Convergent Validity

The convergent validity of the TPOCS-RS Family Therapy subscale was evaluated by examining the magnitude of Pearson product-moment correlations among scores on the observer-rated TPOCS-RS Family Therapy subscale and the observer- and therapist-rated ITT-ABP FT subscales. The convergent validity of a measure is supported if scores produced by the

measure are highly related to scores produced by other measures of the same construct, and these relations are expected to be higher between scores from measures with similar methods than between measures with different methods (Campbell & Fiske, 1959). Thus, it was hypothesized that the TPOCS-RS Family Therapy subscale scores would have a large correlation ($r \geq .37$) with another observer-rated subscale of family therapy techniques (i.e., the observer-rated ITT-ABP FT subscale) and a medium correlation ($.24 \leq r < .37$) with a therapist-rated subscale of family therapy techniques (i.e., the therapist-rated ITT-ABP FT subscale). Correlations (see Tables 9 and 10) were large between the TPOCS-RS Family Therapy subscale and both the observer-rated ITT-ABP FT subscale in the full sample ($r = .660$), and the therapist-rated ITT-ABP FT subscale in the CASALEAP sample ($r = .622$). The correlations between the TPOCS-RS Family Therapy subscale and ITT-ABP FT subscale were consistent with hypotheses for the observer-rated version, larger than hypothesized for the therapist-rated version, and overall supportive of convergent validity.

Follow-up contrasts were computed to determine if the TPOCS-RS Family Therapy subscale evidenced stronger relations with a measure of the same construct using the *same* method (i.e., observer-rated ITT-ABP FT subscale) than to a measure of the same construct using a *different* method (i.e., therapist-rated ITT-ABP FT subscale). In the CASALEAP sample, the correlation between the TPOCS-RS Family Therapy subscale and the observer-rated ITT-ABP FT subscale ($r = .716$) was not significantly different from the correlation between the TPOCS-RS Family Therapy subscale and the therapist-rated ITT-ABP FT subscale ($r = .622$, $z = 0.85$, $p = .395$). Though correlation magnitudes are supportive of convergent validity, follow-up contrast findings do not align with expected differences across methods (i.e., observer- vs. therapist-rated).

Discriminant Validity

The discriminant validity of scores on the TPOCS-RS Family Therapy subscale was evaluated by examining the magnitude and pattern of Pearson product-moment correlations among scores on the TPOCS-RS Family Therapy subscale and scores from (a) subscales of observer-rated non-family therapy techniques (i.e., TPOCS-RS Non-FT, CBT, Psychodynamic, and Client-Centered subscales; observer-rated ITT-ABP CBT/MI subscale C); (b) a subscale of therapist-rated non-family therapy techniques (i.e., therapist-rated ITT-ABP CBT/MI subscale); and (c) a scale of client-therapist alliance (i.e., VTAS-R-SF). Of note, the VTAS-R-SF was only coded for a sample of the MIP sessions in the parent trial, which differs from the sample for which the TPOCS-RS was coded. Thus, VTAS-R-SF scores for the session closest to the date of the sessions for which the TPOCS-RS was coded were used. Correlations are presented in Tables 9 and 10.

It was hypothesized that correlations between the TPOCS-RS Family Therapy subscale and all observer-rated subscales of non-family therapy techniques (i.e., the TPOCS-RS subscales [Non-FT, CBT, Psychodynamic, and Client-Centered] and the observer-rated ITT-ABP CBT/MI subscale) would be small-to-medium ($r < .37$). In the full sample, correlations with the TPOCS-RS Family Therapy subscale were medium for the TPOCS-RS Non-FT subscale ($r = .331$) and the TPOCS-RS Psychodynamic subscale ($r = .254$), and small for the TPOCS-RS CBT subscale ($r = -.147$), the TPOCS-RS Client-Centered subscale ($r = .018$), and the observer-rated ITT-ABP CBT/MI subscale ($r = -.101$). This suggests that the TPOCS-RS Family Therapy subscale has low-to-moderate relations with observer-rated measures of related but distinct constructs, consistent with hypotheses and supportive of discriminant validity.

Given that relations are expected to be higher between scores from measures of the same method than between measures of different methods (Campbell & Fiske, 1959), it was hypothesized that the correlation between the TPOCS-RS Family Therapy subscale and a measure of therapist-rated non-family therapy techniques (i.e., the therapist-rated ITT-ABP CBT/MI subscale) would be small ($r < .24$). In the CASALEAP sample, the correlation between the TPOCS-RS Family Therapy subscale and the therapist-rated ITT-ABP CBT/MI subscale was small ($r = .107$), supportive of hypotheses and discriminant validity. To determine if the TPOCS-RS Family Therapy subscale was more highly related to observer-rated measures of non-family therapy techniques than to a therapist-rated measure of non-family therapy techniques, follow-up contrasts compared the correlation between the TPOCS-RS Family Therapy subscale and the therapist-rated ITT-ABP CBT/MI subscale to the correlations between the TPOCS-RS Family Therapy subscale and the non-corresponding TPOCS-RS subscales, as well as the observer-rated ITT-ABP CBT/MI subscale. The correlation between the TPOCS-RS Family Therapy subscale and the therapist-rated ITT-ABP CBT/MI subscale ($r = .107$) was not significantly different from the correlation between the TPOCS-RS Family Therapy subscale and (1) the TPOCS-RS Non-FT subscale ($r = .345$; $z = 1.27$, $p = .204$); (2) the TPOCS-RS CBT subscale ($r = .192$; $z = 0.44$, $p = .660$); (3) the TPOCS-RS Psychodynamic subscale ($r = .251$; $z = 0.75$, $p = .453$); (4) the TPOCS-RS Client-Centered subscale ($r = .000$; $z = -0.54$, $p = .589$); or (5) the observer-rated ITT-ABP CBT/MI subscale ($r = .070$; $z = -0.18$, $p = .857$). These findings indicate that the TPOCS-RS Family Therapy subscale was not more highly related to observer-rated measures of non-family therapy techniques than to therapist-rated measures of non-family therapy techniques, inconsistent with expected differences.

Finally, it was hypothesized that the correlation between the TPOCS-RS Family Therapy subscale and a measure of client-therapist alliance (i.e., VTAS-R-SF) would be small ($r < .24$). Across all treatment groups (but only including sessions for which a VTAS-R-SF score was available; $N = 58$), the correlation between the TPOCS-RS Family Therapy subscale and the VTAS-R-SF was small (-.194), consistent with hypotheses and supportive of discriminant validity.

Table 9

Correlations Between the TPOCS-RS Family Therapy Subscale and the non-corresponding TPOCS-RS subscales, the observer-rated ITT-ABP subscales, and the VTAS-R-SF in the Full Sample

	OR ITT-ABP FT	TPOCS-RS Non-FT	TPOCS-RS CBT	TPOCS-RS Psych	TPOCS-RS CC	OR ITT-ABP CBT/MI	VTAS-R-SF
TPOCS-RS FT	.660**	.331**	-.147	.254	.018	-.101	-.194
OR ITT-ABP FT	-	.257*	.002	.309*	.022	.087	-.028
TPOCS-RS Non-FT	-	-	-.118	-.045	-.285*	-.103	.113
TPOCS-RS CBT	-	-	-	.115	.332**	.305*	-.084
TPOCS-RS Psych	-	-	-	-	.180	.124	-.092
TPOCS-RS CC	-	-	-	-	-	.199*	.160
OR ITT-ABP CBT/MI	-	-	-	-	-	-	.304*

Note. OR = observer-rated, FT = family therapy, CBT = cognitive behavioral therapy, MI = motivational interviewing, Psych = psychodynamic, CC = client-centered
* $p < .05$, ** $p < .001$

Table 10

Correlations Between the TPOCS-RS Family Therapy Subscale and the non-corresponding TPOCS-RS subscales, the therapist-rated ITT-ABP subscales, and the VTAS-R-SF in the CASALEAP Sample

	TR ITT-ABP FT	OR ITT-ABP FT	TPOCS-RS Non-FT	TPOCS-RS CBT	TPOCS-RS Psych	TPOCS-RS CC	OR ITT-ABP CBT/MI	TR ITT-ABP CBT/MI	VTAS-R-SF
TPOCS-RS FT	.622**	.716**	.345*	-.192	.251	.000	.070	.107	-.058
TR ITT-ABP FT	-	.620**	.340*	-.272*	.127	-.215	-.050	.595**	-.097
OR ITT-ABP FT	-	-	.215	-.008	.289*	.126	.229	.251	.100
TPOCS-RS Non-FT	-	-	-	-.169	-.109	-.385*	-.171	.095	.254
TPOCS-RS CBT	-	-	-	-	.139	.326*	.379*	-.007	-.069
TPOCS-RS Psych	-	-	-	-	-	.245	.170	.061	-.072
TPOCS-RS CC	-	-	-	-	-	-	.285	-.138	.241
OR ITT-ABP CBT/MI	-	-	-	-	-	-	-	.140	.228
TR ITT-ABP CBT/MI	-	-	-	-	-	-	-	-	-.088

Note. TR = therapist-rated, OR = observer rated, FT = family therapy, CBT = cognitive behavioral therapy, MI = motivational interviewing, Psych = psychodynamic, CC = client-centered
* $p < .05$, ** $p < .001$

Construct Validity: Combining Convergent and Discriminant

The construct validity of the TPOCS-RS Family Therapy subscale is supported if it is

more highly related to measures of the same construct (e.g., adherence to family therapy

techniques), than to measures of distinct constructs (e.g., non-family therapy techniques, client-therapist alliance; Campbell & Fiske, 1959). Follow-up contrasts were computed to determine if the correlations produced by the TPOCS-RS Family Therapy subscale followed this pattern. The correlation between the TPOCS-RS Family Therapy subscale and another measure of adherence (i.e., ITT-ABP FT) was expected to be significantly higher than the correlations between the TPOCS-RS Family Therapy subscale and measures of non-family therapy techniques (i.e., TPOCS-RS Non-FT, CBT, Psychodynamic, Client-Centered and observer-rated ITT-ABP CBT/MI) and a measure of client-therapist alliance (i.e., VTAS-R-SF).

In the full sample, the correlation between the TPOCS-RS Family Therapy subscale and another observer-rated subscale of adherence to family therapy techniques (i.e., observer-rated ITT-ABP FT subscale; $r = .660$) was significantly stronger than the correlations between the TPOCS-RS Family Therapy subscale and observer-rated subscales of non-family therapy techniques including (1) the TPOCS-RS Non-FT subscale ($r = .331$; $z = 3.13$, $p = .002$); (2) the TPOCS-RS Psychodynamic subscale ($r = .254$; $z = 3.72$, $p < .001$); (3) the TPOCS-RS CBT subscale ($r = .147$; $z = 4.50$, $p < .001$); (4) the TPOCS-RS Client-Centered subscale ($r = .018$; $z = 5.41$, $p < .001$); and (5) the observer-rated ITT-ABP CBT/MI subscale ($r = .101$; $z = 4.77$, $p < .001$). The correlation between the TPOCS-RS Family Therapy subscale and the ITT-ABP FT subscale ($r = .660$) was also significantly stronger than the correlation between the TPOCS-RS Family Therapy subscale and a scale of client-therapist alliance (i.e., the VTAS-R-SF; $r = .194$; $z = 3.52$, $p < .001$). In the full sample, follow-up contrasts revealed that the TPOCS-RS Family Therapy subscale had a significantly stronger correlation with another measure of observer-rated adherence to family therapy than with observer-rated measures of non-family therapy techniques and client-therapist alliance, supportive of construct validity.

In the CASALEAP sample, the correlation between the TPOCS-RS Family Therapy subscale and a therapist-rated subscale of adherence to family therapy techniques (i.e., therapist-rated ITT-ABP FT subscale; $r = .622$) was significantly stronger than the correlations between the TPOCS-RS Family Therapy subscale and (1) subscales of observer-rated non-family therapy techniques (i.e., TPOCS-RS Non-FT subscale [$r = .345$; $z = 1.86$, $p = .031$], TPOCS-RS Psychodynamic subscale [$r = .251$; $z = 2.38$, $p = .009$], TPOCS-RS CBT subscale [$r = .192$; $z = 2.70$, $p = .007$], TPOCS-RS Client-Centered subscale [$r = .000$; $z = 3.68$, $p < .001$], observer-rated ITT-ABP CBT/MI subscale [$r = .070$, $z = 3.25$, $p = .001$]); (2) a therapist-rated subscale of non-family therapy techniques (i.e., ITT-ABP CBT/MI subscale; $r = .107$; $z = 3.14$, $p = .002$); and a scale of client-therapist alliance (i.e., VTAS-R-SF; $r = .058$; $z = 3.08$, $p = .002$). In the CASALEAP sample, the TPOCS-RS Family Therapy subscale had a significantly stronger correlation with a therapist-rated measure of adherence than with therapist- and observer-rated measures of non-family therapy techniques as well as a measure of client-therapist alliance, supportive of construct validity.

Overall, follow-up contrasts in the full sample and CASALEAP sample revealed that the TPOCS-RS Family Therapy subscale is most highly related to other measures of adherence to family therapy techniques (i.e., same construct; observer- and therapist-rated ITT-ABP FT subscales) than to measures of non-family therapy techniques (i.e., related distinct constructs; TPOCS-RS Non-FT, CBT, Psychodynamic, and Client-Centered subscales and observer- and therapist-rated ITT-ABP CBT/MI subscales), and to a measure of client-therapist alliance (i.e., independent distinct construct; VTAS-R-SF). This pattern of correlations is consistent with the multi-trait multi-method matrix (Cambpell & Fiske, 1959) and supportive of construct validity.

The construct validity of the TPOCS-RS Family Therapy subscale is further supported if its pattern of correlations with other subscales and scales is similar to correlation patterns produced by another observational subscale of adherence to family therapy techniques (i.e., observer-rated ITT-ABP FT subscale). As seen in Tables 9 and 10, in the present study, the observer-rated ITT-ABP FT subscale had a large correlation with other measures of adherence to family therapy (i.e., TPOCS-RS Family Therapy subscale, $r = .660$; therapist-rated ITT-ABP FT subscale, $r = .620$), small-to-medium correlations with subscales of non-family therapy techniques (i.e., TPOCS-RS Non-FT, $r = .257$; TPOCS-RS CBT, $r = .002$; TPOCS-RS Psychodynamic, $r = .309$; TPOCS-RS Client-Centered, $r = .022$; observer-rated ITT-ABP CBT/MI, $r = .087$; and therapist-rated ITT-ABP CBT/MI, $r = .251$), and a small correlation with a scale of client-therapist alliance (i.e., VTAS-R-SF, $r = -.028$). Results indicate that the pattern of correlations produced by the TPOCS-RS Family Therapy subscale mirrors the pattern of correlations produced by the observer-rated ITT-ABP FT subscale, supporting construct validity.

Discriminative Validity

To explore discriminative validity, group differences in TPOCS-RS Family Therapy subscale scores were examined by computing adjusted least square means (LSMs) to account for the nested data design (sessions nested within clients nested within therapists; Barber, Foltz, Crits-Cristoph, & Chittams, 2004), and conducting pairwise comparisons between treatment groups with Bonferonni adjustment. For discriminative validity analyses only, the TPOCS-RS Family Therapy subscale score was derived from the highest Family Therapy item score for each session, consistent with previous research exploring the ability of the TPOCS-RS to identify group differences (e.g., Smith et al., 2017). Given that not all family therapy practices would be expected to be delivered in a single session, averaging all items included in the family therapy

subscale may underestimate the dosage and subsequently underestimate group differences. It was hypothesized that TPOCS-RS Family Therapy subscale scores would be (a) significantly higher for the UC-FT group than for the UC-Other group; (b) significantly higher for the UC-FT group than for the MIP group; and (c) significantly higher for the MIP group than the UC-Other group.

The adjusted mean for the TPOCS-RS Family Therapy subscale score (see Table 11) was 4.81 (SE = 0.21) for the UC-FT group, 3.47 (SE = 0.18) for the MIP group, and 2.18 (SE = 0.29) for the UC-Other group. Pairwise comparisons revealed significantly higher TPOCS-RS Family Therapy scores in the UC-FT group than in both the MIP group (t[96] = 4.80, p < .001) and the UC-Other group (t[96] = 7.29, p < .001). TPOCS-RS Family Therapy scores were also significantly higher for the MIP group than the UC-Other group (t[96] = 3.76, p < .001). Taken together, scores on the TPOCS-RS Family Therapy subscale were highest for the UC-FT group, second highest for the MIP group, and lowest for the UC-Other group. The observed group differences in TPOCS-RS Family Therapy subscale scores are consistent with hypothesized group differences and supportive of discriminant validity.

Table 11

Least Square Means and Differences of Least Square Means

| | M | | | Pairwise Comparisons | | | | | |
| | | | | UC-FT x MIP | | UC-FT x UC-Other | | UC-Other x MIP | |
	UC-FT	MIP	UC-Other	t	p	t	p	t	p
TPOCS-RS FT	4.81	3.47	2.18	4.80	<.001*	7.29	< .001*	3.76	<.001*

Note. *Bonferonni adjusted p-value < .05

Predictive Validity

To examine predictive validity, latent growth curve modeling (Duncan et al., 1999) was conducted using Mplus Version 8.8 (Muthén & Muthén, 1998-2017) to examine the relation

between TPOCS-RS Family Therapy subscale scores and change in clinical outcomes over time. For latent growth curve modeling a sample size of at least $N = 100$ is generally recommended (Shi et al., 2021). However, it has been suggested that for a model hypothesizing a single latent variable underlying four measured indicators, samples close to $N = 50$ may be sufficient (Hamilton et al., 2003; Tanaka, 1987). The current study had a sample of $N = 42$ cases which had measured indicators at four time points (baseline, 3-months, 6-months, and 12-months). Given the limited sample size as well as Henderson et al.'s (2019) findings that greater use of family therapy techniques predicted improvements in clinical outcomes across both UC-FT and UC-Other groups, models were conducted in the full sample.

Clinical outcomes included scores on the Externalizing subscale of the Child Behavior Checklist (CBCL) and Youth Self Report (YSR), total scores on the Self-Report Delinquency scale (SRD), and substance use abstinence/non-abstinence based on the Timeline Follow-Back (TLFB). Separate models were conducted for each clinical outcome, which were measured at four time points (baseline, 3 months, 6 months, 12 months). Given that the MIP trial did not include a 12-month follow-up, the 12-month time point was classified as missing for all outcome measures in the MIP group. Missing data were handled with robust maximum likelihood estimation given that data were missing completely at random (Little & Rubin, 2002). Following procedures outlined in Henderson et al. (2019), TPOCS-RS Family Therapy scores were averaged across all sessions for each client, and a sandwich estimator was used to account for the nesting of clients within therapists (Diggle et al., 2002). Additionally, given treatment group differences in age and that the UC-FT group was found to have greater improvement in clinical outcomes than the UC-Other group in the parent trial (Hogue, Dauber, Henderson et al., 2015), treatment group was included as a covariate. For effect size estimates, β coefficients from LGC

models estimating fully standardized effects were used. It was hypothesized that higher scores on the TPOCS-RS Family Therapy subscale would predict improvements in each of the clinical outcomes (i.e., CBCL Externalizing scale scores, YSR Externalizing scale scores, SRD scale scores, and TLFB scores of substance use abstinence/non-abstinence). As seen in Table 12, latent growth curve models revealed that scores on the TPOCS-RS Family Therapy subscale did not significantly predict change in clinical outcomes over time for any included outcome. These findings are inconsistent with hypotheses and not supportive of predictive validity.

Table 12

Coefficients and Standard Errors for Growth Parameters and Association with TPOCS-RS

Family Therapy subscale scores by treatment outcome

| Outcome | Growth factor mean | | Growth factor variance | | FT technique use | | |
	Intercept coeff (*SE*)	Slope coeff (*SE*)	Intercept coeff (*SE*)	Slope coeff (*SE*)	Intercept coeff (*SE*)	Slope coeff (*SE*)	β
YSR Ext	3.56 (5.28)	-2.08 (1.61)	50.21[*] (13.40)	0[b]	4.09 (2.18)	-0.14 (0.70)	-0.22
CBCL Ext	-14.38[*] (4.23)	0.84 (3.12)	61.36[*] (14.78)	0[b]	8.18 (1.82)	-0.24 (1.18)	-0.17
SRD	-32.72[*] (12.42)	-2.52 (5.01)	359.35[*] (181.71)	0[b]	12.08[*] (4.15)	1.35 (1.49)	0.44
TLFB	0[a]	-0.92 (1.56)	16.60 (12.52)	0[b]	2.26 (1.35)	-0.33 (0.47)	-0.33

Note. Growth factor mean = mean growth trajectory collapsed across treatment conditions and adjusted for adherence-outcome relations and treatment condition; growth factor variance = individual variation around mean growth trajectory; Coeff = regression coefficient; *SE* = standard error

[a]Value fixed to 0 for model identification, necessary for fitting latent growth curve models with categorical outcomes

[b]Value fixed to 0 to facilitate model convergence

*p < .05

For the purposes of comparing the magnitude of the adherence-outcome relations in the current study to meta analytic findings, bivariate correlations were computed between the TPOCS-RS Family Therapy subscale score and outcome scores (i.e., CBCL Externalizing scale, YSR Externalizing scale, SRD scale, TLFB abstinence/non-abstinence) at the 3-month follow-up time point. This time point was chosen because it is closest to post-treatment (the average number of sessions attended was 8.5 in the CASALEAP sample and 6.7 in the MIP sample). TPOCS-RS Family Therapy subscale scores were not significantly correlated with any outcome scores. Effect sizes were positive and ranged from small to medium (YSR Externalizing $r = .324$; CBCL Externalizing $r = .211$; SRD $r = .231$; TLFB $r = .087$). Results indicate that there was a small, nonsignificant relation between TPOCS-RS Family Therapy scores and clinical outcome measure scores such that higher TPOCS-RS Family Therapy scores were associated with higher scores on the outcome measures (indicative of higher levels of symptoms).

Discussion

The present study aimed to determine if the TPOCS-RS subscale scores provide evidence of reliability and validity when used to assess adherence to and differentiation from family therapy in adolescents with externalizing problems. This study expands upon prior research findings that the TPOCS-RS can estimate adherence to and differentiation from CBT in samples of youth with a primary anxiety disorder (McLeod et al., 2015; McLeod et al., 2021). To estimate adherence to family therapy in the present study, a Family Therapy subscale comprising TPOCS-RS items that map onto core family therapy practice elements was generated (Hogue et al., 2017; Hogue et al., 2019). Differentiation was assessed using existing (i.e., Psychodynamic and Client-Centered) and newly generated (i.e., CBT and Non-FT) TPOCS-RS subscales. The interrater reliability of the TPOCS-RS items and subscales was examined. Construct validity was

assessed by exploring the magnitude and pattern of the relations between the TPOCS-RS Family Therapy subscale and other TPOCS-RS subscales (e.g., CBT), other observer- and therapist-rated measures of therapeutic techniques (i.e., ITT-ABP), and a measure of client-therapist alliance (i.e., VTAS-R-SF). To determine discriminative validity, scores on the TPOCS-RS Family Therapy subscale were examined for group differences. Finally, predictive validity was assessed by exploring whether the TPOCS-RS Family Therapy subscale predicted changes in clinical outcomes consistent with previous research. Overall, findings supported the reliability, construct validity, and discriminative validity of the TPOCS-RS in our sample, but not predictive validity.

Reliability

All TPOCS-RS items included in the subscales achieved at least *good* interrater reliability in the full sample (ICC ≥ .60; Cicchetti, 1994), with the majority of items (79%) having reliability in the *excellent* range (ICC ≥ .75; Cicchetti, 1994). The average interrater reliability of the TPOCS-RS items (*M* ICC = .77) was consistent with mean TPOCS-RS item-level interrater reliability in past research (*M* ICC range = .61 to .84; e.g., McLeod & Weisz, 2010; McLeod et al., 2015; McLeod et al., 2021). Subscale-level interrater reliability was *excellent* for all included TPOCS-RS subscales in the full sample, consistent with prior research examining the interrater reliability of the TPOCS-RS subscales (TPOCS-RS subscale ICC range = .72 to .94, *M* ICC = .86; McLeod et al., 2015; McLeod et al., 2021). Reliability findings add to existing evidence suggesting that the TPOCS-RS items and subscales can be reliably coded by independent observers (McLeod et al., 2015, McLeod et al., 2021) and that reliability evidence extends to sessions of family therapy for adolescent externalizing problems.

Notably, for three items considered for inclusion in the TPOCS-RS CBT subscale (*Cognitive Education, Coping Skills, Behavioral Activation*) ICCs were unable to be calculated in the full sample due to a lack of coder variance stemming from low base rates; thus the items were ultimately excluded from analyses. These findings are consistent with past research observations of low base rates for some TPOCS-RS items, especially when used to rate usual clinical care (McLeod et al., 2015; McLeod et al., 2021). Further, visual inspection of item-level ICCs separated by treatment group revealed that for all three treatment groups, at least eight items (primarily those associated with the cognitive and behavioral modalities) had low base rates such that ICCs could not be calculated, but that the specific items for which base rates were low varied across treatment groups. Given that the TPOCS-RS is intended to capture a wide array of therapeutic techniques for multiple problem areas found across several theoretical orientations (McLeod et al., 2015), low base rates for some items in a given sample or treatment group are expected. This suggests that although some items were infrequently observed in the current sample of adolescents with externalizing problems, items should be retained and further explored in new samples (e.g., adolescents receiving CBT to treat externalizing problems).

Construct Validity: Convergent and Discriminant

The magnitude and pattern of correlations in the present sample support the construct validity of TPOCS-RS Family Therapy subscale scores. Specifically, construct validity is supported when correlations are highest between measures of the same construct and same method, second highest between measures of the same construct and different methods, and lowest between measures of distinct constructs (i.e., the multi-trait multi-method matrix; Campbell & Fiske, 1959). The TPOCS-RS Family Therapy subscale scores were highly correlated with scores from other subscales of the same construct (i.e., adherence to family

therapy techniques; observer- and therapist-rated ITT-ABP FT subscales), and these correlations were significantly higher than the correlations between the TPOCS-RS Family Therapy subscale scores and subscale/scale scores of distinct constructs (i.e., non-family therapy techniques and client-therapist alliance; TPOCS-RS Non-FT, CBT, Psychodynamic, Client-Centered subscales; observer- and therapist-rated ITT-ABP CBT/MI subscales; VTAS-R-SF) which ranged from small to medium. Taken together, TPOCS-RS Family Therapy subscale scores were more highly related to other subscales assessing adherence to family therapy (same construct) than to subscales and scales assessing differentiation and client-therapist alliance (distinct constructs), supporting construct validity.

Visual inspection revealed that the TPOCS-RS Family Therapy subscale scores were more highly related to observer-rated subscale scores of family therapy adherence (same method) than to therapist-rated subscale scores of family therapy adherence (different method), consistent with the multi-trait multi-method approach to assessing construct validity, but follow-up contrasts failed to reveal a significant difference. Though research has generally found low to moderate concordance between observer and therapist ratings of treatment fidelity (McLeod et al., 2022), Hogue et al. (2015, 2022) have found strong reliability between therapist and independent observer ratings of the ITT-ABP FT subscale (ICC = .64 to ICC = .75) when used to rate therapy sessions focused on adolescent externalizing problems. Thus, though these findings do not align with the multi-trait multi-method approach, they are consistent with past research utilizing the ITT-ABP in similar samples, providing additional support for construct validity.

Moreover, the pattern of correlations produced by TPOCS-RS Family Therapy subscale scores mirrored the pattern of correlations produced by scores from another observer-rated subscale of family therapy techniques (i.e., the observer-rated ITT-ABP FT subscale) further

supporting the construct validity of TPOCS-RS Family Therapy subscale scores. Although the TPOCS-RS was not originally designed to assess adherence to and differentiation from family therapy, the similarity in correlation patterns among scores produced by the TPOCS-RS subscales and scores produced by the observer-rated ITT-ABP subscales (which *were* originally designed to assess adherence to and differentiation from family therapy), suggests that TPOCS-RS subscale scores may be useful for this purpose.

Overall, the magnitudes and patterns of the correlations between the TPOCS-RS Family Therapy subscale and subscales/scales of adherence to core family therapy techniques, non-family therapy techniques, and client-therapist alliance are largely consistent with hypotheses and supportive of construct validity. These patterns are aligned with those produced by an observational measure designed to assess adherence to and differentiation from family therapy and are consistent with patterns seen in past research exploring the construct validity of treatment integrity measures (e.g., Carroll et al., 2000; Hogue, Dauber et al., 2008; McLeod et al., 2015; McLeod et al., 2021; Southam-Gerow et al., 2016).

Discriminative Validity

The ability of the TPOCS-RS to differentiate between treatment groups expected to differ in their delivery of therapeutic techniques (i.e., discriminative validity) was assessed by examining treatment group differences in TPOCS-RS Family Therapy subscale scores. It was expected that the UC-FT and MIP groups would have higher levels of adherence to core family therapy elements than the UC-Other group given that UC-FT and MIP therapists were from community mental health clinics where family therapy was routinely delivered, and UC-Other therapists were not (Hogue, Dauber, Henderson et al., 2015; Hogue et al., 2016). Additionally, it was expected that UC-FT therapists would have higher levels of adherence to core family

therapy elements than MIP therapists because a portion of the MIP sessions was intended to focus on ADHD medication, thus reducing the amount of time available to deliver core family therapy techniques (Hogue et al., 2016). Group differences in TPOCS-RS Family Therapy subscale scores were found to be highest for the UC-FT group, second highest for the MIP group, and lowest for the UC-Other group, consistent with hypotheses. These findings are aligned with previous research demonstrating the same pattern of scores from the ITT-ABP FT subscale (Hogue, Dauber et al., 2014; Hogue et al., 2016). Findings support the discriminative validity of the TPOCS-RS Family Therapy subscale, indicating that it may be an effective tool for comparing levels of adherence across treatment groups, an essential element of interpreting results from effectiveness and implementation trials (McLeod et al., 2021).

Predictive Validity

The predictive validity of the TPOCS-RS Family Therapy subscale was explored by examining the ability of TPOCS-RS Family Therapy subscale scores to predict change in clinical outcomes over time consistent with previous research findings that greater use of family therapy techniques predicted greater decreases in delinquency, externalizing behavior problems, and substance use (Henderson et al., 2019). Latent growth curve models revealed that TPOCS-RS Family Therapy subscale scores did not predict changes in adolescent- or parent-reported externalizing behavior problems (i.e., YSR and CBCL Externalizing scales, respectively), adolescent-reported delinquency (i.e., SRD scale), or adolescent-reported substance use (i.e., TLFB) assessed at four time points over a 12-month follow-up period. Findings do not support the predictive validity of the TPOCS-RS Family Therapy subscale.

Current findings are inconsistent with existing evidence that more extensive use of family therapy techniques predicts improvements in clinical outcomes in samples of adolescents with

conduct and substance use problems (e.g., Baldwin et al., 2012; Henderson et al., 2019; Hogue et al., 2008; Robbins et al., 2011). This inconsistency may be due to differences in sample composition and size. Specifically, 28.6% of youth included in the present study were from the MIP trial, which selected youth with comorbid externalizing problems and ADHD, whereas past research identifying a significant family therapy adherence – outcome relation did not (e.g., Baldwin et al., 2012; Henderson et al., 2019; Hogue et al., 2008; Robbins et al., 2011). Though family therapy has shown modest effectiveness in improving clinical outcomes among adolescents with ADHD (Robin, 2014), systematic reviews of psychosocial treatments for youth with ADHD suggest that a combination of stimulant medication and behavior therapy is most effective (Evans et al., 2014; Pelham & Fabiano, 2008). Additionally, 43% of youth in the MIP trial did receive stimulant medications. Thus, the inclusion of youth with comorbid ADHD in the present study may have diminished the family therapy adherence – outcome relation either due to family therapy techniques being less effective for this population or to stimulant medication use accounting for a larger proportion of the population change in outcomes. Alternatively, present study findings that scores of adherence to family therapy techniques did not significantly predict changes in clinical outcomes over time may be explained by our sample size ($N = 42$ cases). Specifically, though there is some evidence that a sample size close to $N = 50$ may be sufficient (Tanaka, 1987), it has also been suggested that as sample size decreases, there is an increase in estimation errors resulting in underestimation of true effects (Burchinal, 1989; Hamilton et al., 2003). In the present study, the sample of included cases decreased at each time point as participants were lost to follow-up, and no MIP cases ($N = 12$) had data at the 12-month follow-up time point, suggesting that there may not have been sufficient sample size to detect effects of family therapy technique use on change in clinical outcomes over the extended follow-up period.

Effect sizes for the relation between TPOCS-RS Family Therapy subscale scores and clinical outcomes at the timepoint closest to post-treatment (i.e., 3-month follow-up) were small-to-medium ($r = .087$ to $r = .324$) and nonsignificant. Although these findings are inconsistent with predicted associations between family therapy adherence and clinical outcomes, they are aligned with meta-analyses exploring the relation between treatment integrity and clinical outcomes which have found that the relation is inconsistent and overall smaller than expected. Specifically, meta-analyses exploring the relation between treatment integrity and clinical outcomes in youth psychotherapy have found small effect sizes (e.g., $r = .03$ to $r = .10$; Collyer et al., 2020; $r = -.09$ to $r = -.11$; Martinez, 2020). The inconsistent relation between treatment integrity and clinical outcomes in youth psychotherapy suggests the need to explore moderating factors (e.g., therapist competence, client characteristics, medication use). Future research should continue to explore the family therapy adherence – outcome relation in large samples of youth with behavior problems both with and without comorbid ADHD, controlling for stimulant medication use and other relevant client, therapist, and process-level factors.

Limitations

The present study has several limitations worth noting. First, as previously stated, the included sample size may have been insufficient to detect family therapy adherence – outcome effects. Second, MIP sessions (47.6% of included sessions) did not include therapist-rated treatment integrity data or consistent scores of client-therapist alliance, therefore validity analyses involving those constructs (i.e., correlations with therapist-rated ITT-ABP subscales or VTAS-R-SF) were primarily conducted in the CASALEAP sample, limiting the generalizability of construct validity findings. Generalizability was also limited by the inability to collect session recordings from the only clinic in the parent trials that treated substance use disorders (Hogue,

Dauber, Henderson et al., 2015), thus, the applicability of reliability and validity findings to populations of adolescents with substance misuse is unknown. Finally, the items included in the TPOCS-RS Family Therapy subscale were selected based on theoretical and empirical research identifying the core family therapy elements and expert consultation, but were not confirmed using confirmatory factor analysis. Therefore, it is possible that some relevant items were excluded. For example, the item *Interpretation* (i.e., therapist comments on and suggests the meaning of client behavior and relates that behavior to an aspect of the client's life; McLeod & Weisz, 2010) is part of the TPOCS-RS Psychodynamic subscale and was not included in the TPOCS-RS Family Therapy subscale. However, a core element of family therapy is reframing problems from adolescent-focused to family/relational-focused (i.e., relational reframe; Hogue et al., 2019), which may involve therapists interpreting the meaning of an adolescent's behavior and relating it to family functioning. Thus, the TPOCS-RS Family Therapy subscale should be confirmed with factor analysis (Tavakol & Wetzel, 2020).

Future Directions

A single generic measure that can assess adherence and differentiation, like the TPOCS-RS, may be useful for quality improvement efforts. Specifically, quality improvement efforts in the mental health field aim to maximize clinical improvements by identifying how structural and process elements influence client outcomes (McLeod et al., 2013). A generic measure assessing a variety of prescribed and non-prescribed techniques for treating multiple youth mental and behavioral health problems can capture process elements by identifying what treatment techniques were delivered to determine what techniques are effective, for whom, and under what circumstances (Norcross & Wampold, 2011). While several elements of the TPOCS-RS make it a useful tool for quality improvement efforts, its use requires resources not conducive to

community settings. Specifically, the TPOCS-RS is intended to be rated by independent observers who rate an entire session (McLeod et al., 2015). Community mental health settings are unlikely to have the financial resources to hire independent raters or the temporal resources for existing staff to observe entire sessions. Preliminary evidence supports the feasibility of therapists self-rating adherence to family therapy techniques and differentiation (i.e., motivational interventions) and suggests that therapists did so with moderate accuracy compared to independent observers (Hogue et al., 2022). Thus, future research should explore the concurrence of independent observers and therapists using the TPOCS-RS to rate their therapeutic technique delivery at the end of their sessions.

Conclusion

Overall, findings largely support the ability of the TPOCS-RS to produce reliable and valid subscale scores that estimate adherence to and differentiation from core family therapy techniques. The present study adds to previous research findings that the TPOCS-RS items can be combined to assess treatment adherence (McLeod et al., 2021) and that TPOCS-RS subscales can assess differentiation (McLeod et al., 2015) by being the first to explore its ability to do so in a sample of adolescents receiving family therapy or usual care for externalizing problems. A single non-problem, non-treatment specific measure that can assess both adherence and differentiation can serve as a more efficient tool for identifying what, and to what extent, treatment techniques were delivered in an effectiveness or implementation trials, aiding in the interpretation of observed or non-observed group differences (McLeod et al., 2021).

References

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology*, *32*(3), 328-340.

Achenbach, T. M., & Rescorla, L. A. (2001). *ASEBA school age forms and profiles.* Burlington, VT: ASEBA.

Alexander, J. F., & Parsons, B. V. (1973). Short-term behavioral intervention with delinquent families: Impact on family process and recidivism. *Journal of Abnormal Psychology, 81*(3), 219-225.

Alexander, J. E, & Parsons, B. V. (1982). Functional family therapy. Monterey, CA: Brooks/Cole.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* Washington, DC: American Psychiatric Association.

American Society on Addiction Medicine (2001). *ASAM patient placement criteria for the treatment of substance related disorders* (2nd rev. ed.). Chevy Chase, MD: Author.

Baldwin, S. A., Christian, S., Berkeljon, A., & Shadish, W. R. (2012). The effects of family therapies for adolescent delinquency and substance abuse: A meta-analysis. *Journal of Marital and Family Therapy*, *38*(1), 281-304.

Barber, J. P., Foltz, C., Crits-Christoph, P., & Chittams, J. (2004). Therapists' adherence and competence and treatment discrimination in the NIDA Collaborative Cocaine Treatment Study. *Journal of Clinical Psychology, 60*(1), 29-41.

Becker, S. J., & Curry, J. F. (2008). Outpatient interventions for adolescent substance abuse: A quality of evidence review. *Journal of Consulting and Clinical Psychology*, *76*(4), 531-543.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., ... Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443-451.

Borduin, C. M., Mann, B. J., Cone, L. T., Henggeler, S. W., Fucci, B. R., Blaske, D. M., & Williams, R. A. (1995). Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology, 63*(4), 569-578.

Borntrager, C., Chorpita, B. F., Higa-McMillan, C. K., Daleiden, E. L., & Starace, N. (2013). Usual care for trauma-exposed youth: Are clinician-reported therapy techniques evidence-based? *Children and Youth Services Review*, *35*(1), 133-141.

Brookman-Frazee, L., Haine, R. A., Baker-Ericzén, M., Zoffness, R., & Garland, A. F. (2010). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(3), 254-269.

Burchinal, M. (1989). *Comparison of models for estimating individual growth curves*. Paper presented at the biennial meeting of the Society for Research in Child Development. Kansas City, MO, April 27- 30, 1989.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Carroll, K. M., Martino, S., & Rounsaville, B. J. (2010). No train, no gain? *Clinical Psychology: Science and Practice, 17*, 36–40.

Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., ... Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*(3), 225-238.

Carroll, K. M., & Nuro, K. F. (2002). One size cannot fit all: A stage model for psychotherapy manual development. *Clinical Psychology: Science and Practice*, *9*(4), 396-406.

Cecilione, J. L., McLeod, B. D., Southam-Gerow, M. A., Weisz, J. R., & Chorpita, B. F. (2021). Examining the Relation between Technical and Global Competence in Two Treatments for Youth Anxiety. *Behavior Therapy,* 1-13.

Center for Behavioral Health Statistics and Quality. (2018). *2017 National Survey on Drug Use and Health: Detailed Tables*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566-579.

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., ... Starace, N. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, *18*(2), 154-172.

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, *7*(1), 5-20.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and

    standardized assessment instruments in psychology. *Psychological Assessment, 6*(4),

    284-290.

Collyer, H., Eisler, I., & Woolgar, M. (2020). Systematic literature review and meta-analysis of

    the relationship between adherence, competence and outcome in psychotherapy for

    children and adolescents. *European Child & Adolescent Psychiatry*, *29*, 417-431.

Cox, J. R., Martinez, R. G., & Southam-Gerow, M. A. (2019). Treatment integrity in

    psychotherapy research and implications for the delivery of quality mental health

    services. *Journal of Consulting and Clinical Psychology*, *87*(3), 221-233.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological*

    Bulletin, *52*(4), 281-302.

Dauria, E. F., McWilliams, M. A., & Tolou-Shams, M. (2018). Substance use prevention and

    treatment interventions for court-involved, non-incarcerated youth. In P. M. Monti, S. M.

    Colby, & T. O. Tevyaw (Eds.), *Brief interventions for adolescent alcohol and substance*

    abuse (pp. 213–241). The Guilford Press.

Decker, S., Carroll, K. M., Nich, C., Canning-Ball, M., & Martino, S. (2013). Correspondence of

    motivational interviewing adherence and competence ratings in real and roleplayed client

    sessions. *Psychological Assessment, 25*, 306–312.

Diamond, G. M., Liddle, H. A., Hogue, A., & Dakof, G. A. (1999). Alliance-building

    interventions with adolescents in family therapy: A process study. *Psychotherapy:*

    Theory, Research, Practice, Training, *36*(4), 355-368.

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of longitudinal data* (2nd ed.).

    United Kingdom: Oxford University Press.

Drake, R. E., Goldman, H. H., Leff, H. S., Lehman, A. F., Dixon, L., Mueser, K. T., & Torrey, W. C. (2001). Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services, 52*(2), 179-182.

Eccles, M. P., Armstrong, D., Baker, R., Cleary, K., Davies, H., Davies, S., ... Sibbald, B. (2009). An implementation research agenda. *Implementation Science, 4*(18).

Ehrenreich-May, J., Southam-Gerow, M. A., Hourigan, S. E., Wright, L. R., Pincus, D. B., & Weisz, J. R. (2011). Characteristics of anxious and depressed youth seen in two different clinical contexts. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(5), 398-411.

Elliott, D. S., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use* (Vol. 27, pp. 91-92). Beverly Hills, CA: Sage Publications.

Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review*, *11*(3), 75-113.

Evans, S. W., Owens, J. S., & Bunford, N. (2014). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, *43*(4), 527-551.

Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent psychology*, *37*(1), 215-237.

Faw, L., Hogue, A., Johnson, S., Diamond, G. M., & Liddle, H. A. (2005). The Adolescent Therapeutic Alliance Scale (ATAS): Initial psychometrics and prediction of outcome in family-based substance abuse prevention counseling. *Psychotherapy Research*, *15*(1-2), 141-154.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*(3), 248-260.

Garland, A. F., Accurso, E. C., Haine-Schlagel, R., Brookman-Frazee, L., Roesch, S., & Zhang, J. J. (2014). Searching for elements of evidence-based practices in children's usual care and examining their impact. *Journal of Clinical Child & Adolescent Psychology*, *43*(2), 201-215.

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(1-2), 15-26.

George, D., & Mallery, P. (2016). *Frequencies*. In *IBM SPSS Statistics 23 Step by Step* (pp. 115-125). Routledge.

Gillespie, M. L., Huey Jr, S. J., & Cunningham, P. B. (2017). Predictive validity of an observer-rated adherence protocol for multisystemic therapy with juvenile drug offenders. *Journal of Substance Abuse Treatment*, *76*, 1-10.

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education*, *74*(3), 249-266.

Graziano, A. M., & Diament, D. M. (1992). Parent behavioral training: An examination of the paradigm. *Behavior Modification*, *16*(1), 3-38.

Grella, C. E., Hser, Y. I., Joshi, V., & Rounds-Bryant, J. (2001). Drug treatment outcomes for adolescents with comorbid mental and substance use disorders. *The Journal of Nervous and Mental Disease*, *189*(6), 384-392.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23-34.

Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003). *The Effect of Sample Size on Latent Growth Models.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Hawke, L. D., Koyama, E., & Henderson, J. (2018). Cannabis use, other substance use, and co-occurring mental health concerns among youth presenting for substance use treatment services: Sex and age differences. *Journal of Substance Abuse Treatment*, *91*, 12-19.

Henderson, C. E., Hogue, A., & Dauber, S. (2019). Family therapy techniques and one-year clinical outcomes among adolescents in usual care for behavior problems. *Journal of Consulting and Clinical Psychology*, *87*(3), 308-312.

Henggeler, S. W., Rodick, J. D., Borduin, C. M., Hanson, C. L., Watson, S. M., & Urey, J. R. (1986). Multisystemic treatment of juvenile offenders: Effects on adolescent behavior and family interaction. *Developmental Psychology*, *22*(1), 132-141.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents*. Guilford Press.

Hogue, A., & Bobek, M. (2013). *Medication Integration Protocol (MIP): Provider manual*. New York City, NY: The National Center on Addiction and Substance Abuse, Columbia University.

Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A. (2017). Distilling the core elements of family therapy for adolescent substance use: Conceptual and empirical solutions. *Journal of Child & Adolescent Substance Abuse*, *26*(6), 437-453.

Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A.

(2019). Core elements of family therapy for adolescent behavior problems: empirical

distillation of three manualized treatments. *Journal of Clinical Child & Adolescent

Psychology*, *48*(1), 29-41.

Hogue, A., Bobek, M., MacLean, A., Miranda, R., Wolff, J. C., & Jensen-Doss, A. (2020). Core

elements of CBT for adolescent conduct and substance use problems: Comorbidity,

clinical techniques, and case examples. *Cognitive and Behavioral Practice*, *27*(4), 426-

441.

Hogue, A., Bobek, M., Porter, N., Dauber, S., Southam-Gerow, M. A., McLeod, B. D., &

Henderson, C. E. (2021). Core elements of family therapy for adolescent behavioral

health problems: Validity generalization in community settings. *Journal of Clinical Child

& Adolescent Psychology,* 1-13.

Hogue, A., Bobek, M., Porter, N., MacLean, A., Bruynesteyn, L., Jensen-Doss, A., &

Henderson, C. E. (2022). Therapist self-report of fidelity to core elements of family

therapy for adolescent behavior problems: Psychometrics of a pragmatic quality indicator

tool. *Administration and Policy in Mental Health and Mental Health Services

Research*, *49*(2), 298-311.

Hogue, A., Bobek, M., Tau, G. Z., & Levin, F. R. (2014). Clinical strategies for integrating

medication interventions into behavioral treatment for adolescent ADHD: The

Medication Integration Protocol. *Child & Family Behavior Therapy*, *36*(4), 280-304.

Hogue, A., & Dauber, S. (2013). Assessing fidelity to evidence-based practices in usual care:

The example of family therapy for adolescent behavior problems. *Evaluation and

Program Planning*, *37*, 21-30.

Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C., Inclan, J., ... Liddle, H. A.

(2008). Assessing fidelity in individual and family therapy for adolescent substance

abuse. *Journal of Substance Abuse Treatment*, *35*(2), 137-147.

Hogue, A., Dauber, S., & Henderson, C. E. (2014). Therapist self-report of evidence-based

practices in usual care for adolescent behavior problems: Factor and construct

validity. *Administration and Policy in Mental Health and Mental Health Services

Research*, *41*(1), 126-139.

Hogue, A., Dauber, S., & Henderson, C. E. (2017). Benchmarking family therapy for adolescent

behavior problems in usual care: Fidelity, outcomes, and therapist performance

differences. *Administration and Policy in Mental Health and Mental Health Services

Research*, *44*(5), 626-641.

Hogue, A., Dauber, S., Henderson, C. E., Bobek, M., Johnson, C., Lichvar, E., & Morgenstern, J.

(2015). Randomized trial of family therapy versus nonfamily treatment for adolescent

behavior problems in usual care. *Journal of Clinical Child & Adolescent

Psychology*, *44*(6), 954-969.

Hogue, A., Dauber, S., Lichvar, E., Bobek, M., & Henderson, C. E. (2015). Validity of therapist

self-report ratings of fidelity to evidence-based practices for adolescent behavior

problems: Correspondence between therapists and observers. *Administration and Policy

in Mental Health and Mental Health Services Research*, *42*(2), 229-243.

Hogue, A., Dauber, S., Samuolis, J., & Liddle, H. A. (2006). Treatment techniques and outcomes

in multidimensional family therapy for adolescent behavior problems. *Journal of Family

Psychology*, *20*(4), 535-543.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008).

Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, *76*(4), 544-555.

Hogue, A., Henderson, C. E., Ozechowski, T. J., & Robbins, M. S. (2014). Evidence base on outpatient behavioral treatments for adolescent substance use: Updates and recommendations 2007–2013. *Journal of Clinical Child & Adolescent Psychology*, *43*(5), 695-720.

Hogue, A., Lichvar, E., & Bobek, M. (2016). Pilot evaluation of the medication integration protocol for adolescents with ADHD in behavioral care: Treatment fidelity and medication uptake. *Journal of Emotional and Behavioral Disorders*, *24*(4), 223-234.

Hogue, A., & Liddle, H. A. (2009). Family-based treatment for adolescent substance abuse: Controlled trials and new horizons in services research. *Journal of Family Therapy, 31*(2), 126-154.

Hogue, A., Liddle, H. A., Dauber, S., & Samuolis, J. (2004). Linking Session Focus to Treatment Outcome in Evidence-Based Treatments for Adolescent Substance Abuse. *Psychotherapy:* Abuse. *Psychotherapy: Theory, Research, Practice, Training*, *41*(2), 83-96.

Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training*, *33*(2), 332-345.

Huey Jr, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of

change in multisystemic therapy: Reducing delinquent behavior through therapist

adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology*, *68*(3), 451-467.

Huizinga, D., & Elliott, D. S. (1986). Reassessing the reliability and validity of self-report delinquency measures. *Journal of Quantitative Criminology*, *2*(4), 293-327.

Keyes, K. M., Brady, J. E., & Li, G. (2015). Effects of minimum legal drinking age on alcohol and marijuana use: Evidence from toxicological testing data for fatally injured drivers aged 16 to 25 years. *Injury Epidemiology*, *2*(1), 1-10.

Krishna, R., Maithreyi, R., & Surapaneni, K. M. (2010). Research bias: A review for medical students. *Journal of Clinical and Diagnostic Research*, *4*(2), 2320-2324.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Liddle, H. A., Dakof, G. A., Turner, R. M., Henderson, C. E., & Greenbaum, P. E. (2008). Treating adolescent drug abuse: A randomized trial comparing multidimensional family therapy and cognitive behavior therapy. *Addiction*, *103*(10), 1660-1670.

Liddle, H. A. & Hogue, A. (2001) Multidimensional Family Therapy for adolescent substance abuse. In: Wagner, E. F. & Waldron, H. B., eds. *Innovations in Adolescent Substance Abuse Interventions*, pp. 229–261. London: Pergamon/Elsevier Science.

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data (2nd ed.)*. New York, NY: Wiley.

Margolin, G., Oliver, P., Gordis, E., O'Hearn, H., Medina, A., Ghosh, C., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child & Family Psychology Review, 1*, 195–213.

Martinez, R. G. (2020). *A meta-analysis investigating the correlation between treatment integrity and youth client outcomes* [Doctoral dissertation, Virginia Commonwealth University]. VCU Scholars Compass.

McCart, M. R., & Sheidow, A. J. (2016). Evidence-based psychosocial treatments for adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology*, *45*(5), 529-563.

McHugh, M. L. (2011). Multiple comparison analysis testing in ANOVA. *Biochemia Medica*, *21*(3), 203-209.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*(2) 73-84.

Mcleod, B. D. (2014). *Scoring manual for the Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale*. Unpublished manual.

McLeod, B. D., Martinez, R. G., Southam-Gerow, M. A., Weisz, J. R., & Chorpita, B. F. (2021). Can a Single Measure Estimate Protocol Adherence for Two Psychosocial Treatments for Youth Anxiety Delivered in Community Mental Health Settings?. *Behavior Therapy*.

McLeod, B. D., Porter, N., Hogue, A., Becker-Haimes, E. M., & Jensen-Doss, A. (2022). What is the Status of Multi-Informant Treatment Fidelity Research? *Journal of Clinical Child & Adolescent Psychology*, *52*(1) 74-94.

McLeod, B. D., Smith, M. M., Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2015). Measuring treatment differentiation for implementation research: The Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale. *Psychological Assessment, 27*(1), 314-325.

McLeod, B. D., Southam-Gerow, M. A., Rodríguez, A., Quinoy, A. M., Arnold, C. C., Kendall, P. C., & Weisz, J. R. (2018). Development and initial psychometrics for a therapist competence instrument for CBT for youth anxiety. *Journal of Clinical Child & Adolescent Psychology*, *47*(1), 47-60.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice*, *20*(1), 14-32.

McLeod, B. D., Southam-Gerow, M. A., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review*, *38*(4), 541-546.

McLeod, B. D., & Weisz, J. R. (2010). The therapy process observational coding system for child psychotherapy strategies scale. *Journal of Clinical Child & Adolescent Psychology, 39*(3), 436-443.

McLeod, J. D., Uemura, R., & Rohrman, S. (2012). Adolescent mental health, behavior problems, and academic achievement. *Journal of Health and Social Behavior*, *53*(4), 482-497.

Merikangas, K. R., & McClair, V. L. (2012). Epidemiology of substance use disorders. *Human Genetics*, *131*(6), 779-789.

Mizock, L., & Harkins, D. (2011). Diagnostic bias and conduct disorder: Improving culturally sensitive diagnosis. *Child & Youth Services, 32*(3), 243-253.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Norcross, J. C., & Wampold, B. E. (2011). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology*, *67*(2), 127-132.

Novick, M. R., & Lewis, C. (1966). Coefficient alpha and the reliability of composite

    measurements. *Psychometrika, 32*, 1-13.

Pelham, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidencebased assessment of attention

    deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and*

    Adolescent Psychology, 34, 449–476.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues

    and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365-

    383.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy

    research: Analysis of the studies and examination of the associated factors. *Journal of*

    Consulting and Clinical Psychology, *75*(6), 829-841.

Piquero, A. R., Macintosh, R., & Hickman, M. (2002). The validity of a self-reported

    delinquency scale: Comparisons across gender, age, race, and place of

    residence. *Sociological Methods & Research*, *30*(4), 492-529.

Regan, J., Park, A. L., & Chorpita, B. F. (2019). Choices in treatment integrity: Considering the

    protocol and consultant recommendations in child and adolescent therapy. *Journal of*

    Clinical Child & Adolescent Psychology, *48*(sup1), S79-S89.

Riedinger, V., Pinquart, M., & Teubert, D. (2017). Effects of systemic therapy on mental health

    of children and adolescents: A meta-analysis. *Journal of Clinical Child & Adolescent*

    Psychology, *46*(6), 880-894.

Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C., & Szapocznik, J.

    (2011). Therapist adherence in brief strategic family therapy for adolescent drug

    abusers. *Journal of Consulting and Clinical Psychology*, *79*(1), 43-53.

Robin, A. L. (2014). Family therapy for adolescents with ADHD. *Child and Adolescent Psychiatric Clinics*, *23*(4), 747-756.

Rohde, P., Lewinsohn, P. M., Seeley, J. R., Klein, D. N., Andrews, J. A., & Small, J. W. (2007). Psychosocial functioning of adults who experienced substance use disorders as adolescents. *Psychology of Addictive Behaviors*, *21*(2), 155.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis.* McGraw-Hill.

Rowe, C. L. (2012). Family therapy for drug abuse: Review and updates 2003–2010. *Journal of Marital and Family Therapy*, *38*(1), 59-81.

Santisteban, D. A., Coatsworth, J. D., Perez-Vidal, A., Kurtines, W. M., Schwartz, S. J., LaPerriere, A., & Szapocznik, J. (2003). Efficacy of brief strategic family therapy in modifying Hispanic adolescent behavior problems and substance use. *Journal of Family Psychology*, *17*(1), 121-133.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 32-43.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., ... Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(20), 22-33.

Shelef, K., & Diamond, G. M. (2008). Short form of the revised Vanderbilt Therapeutic Alliance

       Scale: Development, reliability, and validity. *Psychotherapy Research*, *18*(4), 433-443.

Shi, D., DiStefano, C., Zheng, X., Liu, R., & Jiang, Z. (2021). Fitting latent growth models with

       small sample sizes and non-normal missing data. *International Journal of Behavioral*

       Development, *45*(2), 179-192.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability.

       Psychological Bulletin, *86*(2), 420-428.

Smith, M. M., McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Kendall, P. C., &

       Weisz, J. R. (2017). Does the delivery of CBT for youth anxiety differ across research

       and practice settings? *Behavior Therapy*, *48*(4), 501-516.

Sobell, L. C., & Sobell, M. B. (1996). *Timeline Followback user's guide: A calendar method for*

       assessing alcohol & drug use. Toronto, Canada: Addiction Research Foundation.

Southam-Gerow, M. A., Chorpita, B. F., Miller, L. M., & Gleacher, A. A. (2008). Are children

       with anxiety disorders privately referred to a university clinic like those referred from the

       public mental health system? *Administration and Policy in Mental Health and Mental*

       Health Services Research, *35*(3), 168-180.

Southam-Gerow, M. A., McLeod, B. D., Arnold, C. C., Rodríguez, A., Cox, J. R., Reise, S. P., ...

       Kendall, P. C. (2016). Initial development of a treatment adherence measure for

       cognitive–behavioral therapy for child anxiety. *Psychological Assessment*, *28*(1), 70-80.

Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2003). Youth with anxiety disorders in

       research and service clinics: Examining client differences and similarities. *Journal of*

       Clinical Child and Adolescent Psychology, *32*(3), 375-385.

Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-

Smith, J. K. (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 1043-1052.

Staff, J., Patrick, M. E., Loken, E., & Maggs, J. L. (2008). Teenage alcohol use and educational attainment. *Journal of Studies on Alcohol and Drugs*, *69*(6), 848-858.

Stanhope, V., Manuel, J. I., Jessell, L., & Halliday, T. M. (2018). Implementing SBIRT for adolescents within community mental health organizations: A mixed methods study. *Journal of Substance Abuse Treatment*, *90*, 38-46.

Szapocznik, J., Hervis, O. E., & Schwartz, S. (2003). *Therapy manuals for drug addiction: Brief strategic family therapy for adolescent drug abuse.* Bethesda, MD: National Institute on Drug Abuse.

Tabachnik, B. G., & Fidell, S. L. (2007). *Using multivariate statistics.* Boston: Pearson Education Inc.

Tanaka, J. S. (1987). " How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, 134-146.

Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2013). The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis. *Journal of Substance Abuse Treatment*, *44*(2), 145-158.

Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, *11*, 245-247.

Waldron, H. B., Slesnick, N., Brody, J. L., Turner, C. W., & Peterson, T. R. (2001). Treatment

outcomes for adolescent substance abuse at 4- and 7-month assessments. *Journal of Consulting and Clinical Psychology, 69*(5), 802–813.

Waldron, H. B., & Turner, C. W. (2008). Evidence-based psychosocial treatments for adolescent substance abuse. *Journal of Clinical Child & Adolescent Psychology*, *37*(1), 238-261.

Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry*, *70*(7), 750-761.

Weisz, J. R., Southam-Gerow, M. A., Gordis, E. B., Connor-Smith, J. K., Chu, B. C., Langer, D. A., ... Weiss, B. (2009). Cognitive–behavioral therapy versus usual clinical care for youth depression: An initial test of transportability to community clinics and clinicians. *Journal of Consulting and Clinical Psychology*, *77*(3), 383-396.

Wong, S. S., Zhou, B., Goebert, D., & Hishinuma, E. S. (2013). The risk of adolescent suicide across patterns of drug use: A nationally representative study of high school students in the United States from 1999 to 2009. *Social Psychiatry and Psychiatric Epidemiology*, *48*(10), 1611-1620.

Wu, L. T., Gersing, K., Burchett, B., Woody, G. E., & Blazer, D. G. (2011). Substance use disorders and comorbid Axis I and II psychiatric disorders among young psychiatric patients: Findings from a large electronic health records database. *Journal of Psychiatric Research*, *45*(11), 1453-1462.