2023

Initial Psychometric Evaluation of an Observer-Rated Measure of Shared Decision Making Practices in Youth Psychotherapy

Jennifer LC Herbst
Virginia Commonwealth University

Initial Psychometric Evaluation of an Observer-Rated Measure of Shared Decision Making Practices in Youth Psychotherapy

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

by

Jennifer Lynn Cecilione Herbst, M.S.
Bachelor of Arts, University of Richmond, 2015
Master of Science, Virginia Commonwealth University, 2019

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

This dissertation project was be supported by grant funding from the National Institute of Mental Health (1F31 MH 127924 - 01A1)

Virginia Commonwealth University
Richmond, Virginia
October 11, 2023

Acknowledgement

The completion of this dissertation would not have been possible without the continued love and support of my wonderful husband, Kevin Herbst. Thank you for being my shoulder to cry (a lot) on and for always cheering me up when I get discouraged. Thank you also to the Ceciliones—Anthony, Linda, and Kristen. You guys taught me how to hustle for my dreams at all costs. Your unwavering love and belief in me have helped me reach heights I never thought I could.

I would also like to thank my mentor and dissertation chair, Dr. Bryce McLeod. I sincerely appreciate all the hard work you have committed to my growth as a young scholar, especially your dedication to helping me obtain my F31—I could not have done it without you. Thank you to my committee members, Drs. Beth Heller, Michael Southam-Gerow, Rosalie Corona, David Langer, and Michael Broda for kindly agreeing to help guide me throughout this process. Dr. Heller, thank you for helping me find my path and for constantly going out of your way to support my clinical pursuits. I would also like to thank Emma Dear, Tyler Lehman, Lindsay Edwards, Grace Li, and Alexys Weihl for being integral members of the coding team.

To my amazing cohort, lab mates, friends, and UR family, thank you for bringing so much light to the many dark moments along this long and arduous trek of graduate studies. I would not have been able to make it through without you all.

Finally, I would like to dedicate this dissertation to the memory of my grandparents—Arnold Martins and Anna Cecilione. You both sacrificed so much and fought battles on and off the battlefield for me to get to where I am today. I know you'll both be with your Dr. Jen on graduation day, I love you mostest.

Table of Contents

Abstract

INITIAL PSYCHOMETRIC EVALUATION OF AN OBSERVER-RATED MEASURE OF SHARED DECISION MAKING PRACTICES IN YOUTH PSYCHOTHERAPY

By: Jennifer Lynn Cecilione Herbst, 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

SDM is a process by which clinicians and clients collaboratively make treatment decisions; this process may improve the alliance and clinical outcomes. SDM in youth psychotherapy is understudied, and a critical gap in the field is the paucity of established measures designed to assess SDM practices in youth psychotherapy. This project aimed to fill this gap by examining the score reliability and validity of a novel observer-rated measure of SDM designed for youth psychotherapy—the Shared Decision Making Practices in Youth Mental Health scale (SDM-YMH)—and using the measure to investigate linkages between SDM practices, the alliance, and clinical outcomes. Therapy tapes (N = 150) and accompanying treatment integrity and alliance data were utilized to develop the SDM-YMH. The 38 youth (52.6% male, 60.5% White) in these tapes ranged in age from 8 to 13 years old (M = 9.84, SD = 1.65). Findings provide support for the content validity of SDM-YMH items and support for the inter-rater reliability of SDM-YMH subscale scores (Youth Total ICC[2,2] = .64; Caregiver Total ICC[2,2] = .78) and approximately 38% of items (Youth ICCs: -0.2 to 0.61; Caregiver ICCs: -0.2 to 0.73). However, the construct and predictive validity of SDM-YMH item and subscale scores were not supported. Further research is needed to clarify the psychometric properties of the SDM-YMH, particularly with a more recent sample in which clinicians were explicitly instructed to deliver SDM practices.

Literature Review

Youth Mental Health

Mental health disorders cause significant impairment for youth. In the United States alone, up to 33% of youth struggle with a mental health disorder (CDC, 2020; NAMI, 2020). Mental health disorders in childhood can lead to significant impairment and health concerns in adulthood (Fryers & Brugha, 2013; Hofstra, Van Der Ende, & Verhulst, 2002). Hence, these mental health disorders are important to treat in childhood before they cause further harm and distress later in life.

Thankfully, hundreds of evidence-based programs (EBPs) have been developed to address mental health concerns in youth (e.g., Coping Cat to address youth anxiety; Kendall & Hedtke, 2006). Many of these EBPs have demonstrated positive outcomes in clinical trials in decreasing mental health disorder symptomatology in youth (Weisz et al., 2017). However, the effectiveness of these EBPs in community mental health centers, where most youth receive psychotherapy to address mental health concerns (Cummings et al., 2016), has been disappointing (Weisz et al., 2013). Thus, it is vital to optimize the effectiveness of EBPs for youth being treated in community mental health centers. One way to accomplish this may be to better adapt EBPs to fit individual clients and their caregiver(s).

Adapting EBPs for Youth Psychotherapy

Adapting EBPs to fit client preferences in community mental health centers may help improve care quality (i.e., agreement about treatment activities and goals, improved clinical outcomes; Heisler et al., 2003; Loh et al., 2006). A poor fit between EBPs and client preferences may be one reason why these programs are not always effective in community settings (Chorpita et al., 2011; Schiffman et al., 2011; Weisz et al., 2013). For instance, a poor fit between an EBP

and client preferences can lead to premature dropout and poor clinical outcomes (Aarons et al., 2009; Nicolas & Schwartz, 2012). Adapting EBPs to fit client preferences, ideals (Chamber & Norton, 2016; Connor-Smith & Weisz, 2003; Marques et al., 2019), and cultural values (Aggarwal et al., 2016; Alegría et al., 2018; Whitely, 2009) may thus result in better care quality (i.e., therapeutic relationship and outcomes).

Yet, clinicians may struggle to adapt EBPs to fit the preferences of clients, especially youth and their caregivers (Aarons et al., 2009; Chambers & Norton, 2016). This may be in part due to an absence of guidelines and practices that clinicians can use to adapt EBPs (Marques et al., 2019; Stirman Wiltsey et al., 2013, 2015). To address this gap, guidelines are needed to help clinicians better adapt EBPs to the needs of their individual clients (Baumann et al., 2017). Shared decision making (SDM) has the potential to fill this gap.

Shared Decision Making

There is no one, agreed-upon definition of SDM that has been used consistently across studies (Makoul & Clayman, 2006; Shay & Lafta, 2015; Street et al., 2009). While some definitions have focused on both the process *and* product of SDM (e.g., SDM being a process by which both the client and clinician are explicit about their values and treatment preferences *and* arrive at an agreed decision; Elwyn et al., 2001), several others have described SDM as the interactive process of providers and patients sharing information to collaboratively make healthcare decisions (Alegría et al., 2018; Langer & Jensen-Doss, 2018). The current project will utilize the latter as a working definition of SDM, as it does not conflate both process (conversing about treatment decisions) and product (the agreed-upon decision being made). Rather, operationalizing SDM as a *process* of collaboration regarding treatment decisions allows for the

efforts to collaborate to be distinct from the product of the collaboration (e.g., whether the decision was agreed upon by all parties).

This process of SDM may help clinicians better adapt EBPs to fit the preferences, ideals, and cultural values of youth clients and their caregivers (Langer & Jensen-Doss, 2018). That is, if clinicians have a solid understanding of their clients' goals, preferences, and values, they will likely be better able to adapt EBPs to better fit their clients' specific goals, preferences, and values. Thus, via SDM, clinicians may better understand their clients wishes and provide the best fitting treatment. Yet, most SDM research to date has been conducted in medical settings (Shay & Lafta, 2015) and focused on client-provider collaboration when discussing treatment options for a medical problem (e.g., cancer; Mandelblatt et al., 2006; Schrager et al., 2020; or HIV; Fuller et al., 2017; Okoli et al., 2020). This research indicates that SDM leads to increased client satisfaction (Shay & Lafta, 2015) and improved health outcomes (Durand et al., 2014; Sepucha et al., 2018). In adult mental health research, SDM has been shown to improve care quality (De Las Cuevas, & Peñate, 2014; Slade, 2017) and help clinicians adapt EBPs to fit the needs of clients from minoritized communities (Aggarwal et al., 2016; Alegría et al., 2018; Whitley, 2009). Thus, research has established the potential of SDM to improve care quality in medical settings and adult mental health treatment.

Potential Role of SDM in Youth Psychotherapy

However, SDM is understudied in youth mental health treatment. It has been posited that SDM can potentially help clinicians adapt EBPs to better fit individual clients (Langer & Jensen-Doss, 2018). Figure 1 (see Appendix) illustrates how *SDM practices* (i.e., techniques clinicians use to promote SDM, such as asking clients to share opinions regarding treatment activities; Elwyn et al., 2001; Langer & Jensen-Doss, 2018; Shay & Lafta, 2015) can play an instrumental

role in improving the effectiveness of EBPs delivered in community mental health centers (Langer & Jensen-Doss, 2018; Loh et al., 2006; Street et al., 2008).

As SDM is an interactive process that involves the sharing of information and collaboration between clinicians and clients/caregivers, there are several perspectives from which to examine SDM (e.g., studying how youth's involvement in SDM impacts their outcomes). Yet, it is particularly important to first examine the practices clinicians can implement to encourage SDM, as opposed to what consumers of mental health care can do to initiate SDM. This may help clinicians better promote equitable, client-centered care that honors client autonomy without placing an increased burden on clients and/or caregivers to start the SDM process. Thus, the current study will focus on SDM practices that can be used to engage families in conversations about treatment decisions. SDM practices are hypothesized to help clinicians engage youth and caregivers in discussions that allow them to adapt treatments, including EBPs (Langer & Jensen-Doss, 2018), which promotes SDM and improves care quality. For example, if clinicians implement SDM practices (e.g., present pros and cons of an EBP and elicit client feedback), they may be better equipped to develop a treatment plan that best suits that client.

The mechanism of change in this model (Figure 1) is the alliance (i.e., the affective and collaborative components of the client-clinician relationship; Elvins & Green, 2008; Shirk & Saiz, 1992). That is, SDM practices directed at the youth and caregiver are expected to promote SDM, which, in turn, leads to a stronger alliance (Ackerman & Hilsenroth, 2003; Fjermestad et al., 2016; Fjermestad et al., 2020). For example, clinicians implementing SDM practices may learn more about their clients' goals for treatment as well as their values and preferences as they relate to certain treatment activities. With this information, clinicians can collaboratively select

treatment activities with clients, thereby developing a tailored treatment plan and increasing agreement upon treatment activities, which is an important component of the alliance (Bordin, 1979; McLeod, 2011). By explicitly eliciting and honoring their clients' preferences and values, clinicians may also increase the therapeutic bond between themselves and their clients—another important component of the alliance (Elvins & Green, 2008; Shirk & Saiz, 1992). A strong youth-clinician and caregiver-clinician alliance has been consistently demonstrated to play a pivotal role in promoting positive youth clinical outcomes (Karver et al., 2018; McLeod, 2011). As this model illustrates, SDM practices have the potential to promote SDM and improve care quality in community mental health centers (Grim et al., 2016; Langer & Jensen-Doss, 2018; Simmons et al., 2016). While SDM has been largely understudied in youth psychotherapy, there is some evidence from extant research to support the effectiveness of certain SDM practices across various settings.

Demonstrated Effectiveness of SDM in Psychotherapy

SDM originated and has a rich history in the adult medical field (e.g., Elwyn et al., 2000; Stiggelbout et al., 2012). Yet, more modest evidence is emerging that supports the effectiveness of SDM in the mental health field. A recent literature review was conducted to identify common clinician-delivered, evidence-based SDM practices in psychotherapy, and a total of 14 studies were identified (Cecilione Herbst, 2022). This review focused on randomized controlled trials (RCTs), as this study design is the gold standard for demonstrating causation (Kazdin, 2016). Studies identified in this review examined the effect of SDM interventions on theoretically-relevant outcomes—SDM, the alliance, clinical outcomes. (See Appendix for additional information on these studies).

Of the identified studies, 12 were conducted in mental health settings, and two in substance use settings. Only two of these studies involved youth samples (Langer et al., 2022; Westermann et al., 2013); whereas the remaining studies included adult patients. SDM practices were deemed to be *evidence-based* if they demonstrated a statistically significant increase in at least one outcome variable and had a methodologically sound design. Methodological rigor was determined via the following: (a) low risk of bias (on the basis of the Cochrane Risk of Bias Tool Version 2; Cochrane, 2021), (b) inclusion of appropriate outcome measures (e.g., measures that had demonstrated construct validity), (c) potential for generalizability of findings (e.g., included a heterogeneous sample), and (d) inclusion of an assessment of protocol adherence (i.e., extent to which a clinician delivers the treatment as articulated in the protocol or treatment manual; Southam-Gerow et al., 2016).

Practices that Increased SDM

Nine of the 14 identified studies assessed SDM as an outcome variable. Four of these studies contained methodological flaws that interfered with the interpretability of findings (e.g., inappropriate assessment of outcome variables). Alegría et al. (2018), Aoki et al. (2019), Langer et al., (2021), and Westermann et al. (2013) demonstrated significant increases in SDM as a result of their interventions. Notably, Aoki et al. (2019), Langer et al. (2022), and Westermann et al. (2013) operationally defined SDM as client/caregiver-reported perceived involvement in decision making and client/caregiver-reported decisional conflict; whereas Alegría et al. (2018) operationalized SDM as an observer-report of the SDM process.

Across these studies, common SDM practices included: (a) clearly introducing the SDM process, (b) explaining the clients' diagnosis and providing corresponding psychoeducation, (c) exploring client's background and values, (d) eliciting goals for therapy, I providing information

about potential treatment options, (f) reviewing pros and cons of each option, (g) developing the treatment plan collaboratively with the client based on their preferences/values, and (h) establishing a plan for monitoring progress. Importantly, Langer et al. (2022) is the only study to date that examined the active inclusion and participation of youth in the SDM process. In their protocol, clinicians were instructed to emphasize the youth's perspective in treatment planning. Moreover, they also included explicit instructions to discuss who would be involved in treatment (e.g., youth alone, youth and caregivers).

Practices that Increased the Alliance and Clinical Outcomes

Seven of the 14 studies measured the alliance as an outcome, yet only two of these studies demonstrated a statistically significant association between the SDM intervention and the alliance (Hamann et al., 2020; Metz et al., 2019). However, findings from these studies should be interpreted with caution due to methodological concerns (Cecilione Herbst, 2022). Moreover, four studies included clinical outcomes as an outcome variable, and only one study (with considerable methodological flaws, e.g., non-randomized sample, unexplained unbalanced rates of missingness from treatment and placebo groups; Joosten et al., 2009) demonstrated a significant increase in clinical outcomes due to their SDM intervention. Notably, Langer et al. (2022) was underpowered. As such, it is possible that with an adequately powered sample, they may have found statistically significant differences in alliance and outcomes between the SDM and control groups.

Adapting SDM Practices for Youth Psychotherapy

As few clinical trials have examined SDM in youth psychotherapy, further research is needed to understand and identify effective SDM practices in youth mental health treatment. Notably, SDM in youth psychotherapy would be distinct from SDM in adult psychotherapy

populations and from SDM in medical settings in three primary ways. One, SDM in youth psychotherapy would need to be developmentally appropriate for youth clients (e.g., use appropriate language and examples when communicating treatment options to children as compared to adolescents; Barry & Pickard, 2008; Weisz & Hawley, 2002). Two, SDM in youth psychotherapy should involve clients' caregivers, as caregivers are important stakeholders in the youth psychotherapy process, are involved in many treatment activities, and often fund sessions for youth clients (McLeod et al., 2013; Langer et al., 2022). Lastly, distinct from SDM in medical settings that often involve one treatment decision (surgery A vs. B, e.g., Mandelblatt et al., 2006), SDM in youth psychotherapy would involve many decision points throughout treatment (Chorpita et al., 2011). That is, SDM in youth psychotherapy could be used to *adapt* the implementation of an EBP to better fit clients' preferences and values. As aforementioned, since EBPs have demonstrated disappointing clinical outcomes in community mental health centers (where most youth receive psychotherapy), it is important to study possible avenues that can be used to increase the effectiveness of EBPs (e.g., adapting EBPs to best fit clients and caregivers values and preferences via SDM).

While the aforementioned evidence-based SDM practices have been demonstrated to increase theoretically relevant outcomes (e.g., SDM), most of this evidence comes from the adult (not youth) mental health field (e.g., Alegría et al., 2018). Given the differences between SDM in adult medical/mental health and SDM in youth psychotherapy, it is important that these practices be adapted to and examined in the context of youth psychotherapy. Thus, Langer and Jensen-Doss (2018) proposed a theoretical model detailing how providers might utilize SDM in youth psychotherapy, which was in part formulated based on common SDM elements identified in a large-scale review of SDM (Clayman & Makoul, 2009). Specifically, they offered the following

steps: “(a) identify stakeholder roles, (b) specify decisions, (c) present available options, (d) determine pros and cons (e) design treatment plan, (f) monitor progress, and (g) return to SDM process” (p. 825). These steps largely map on to extant evidence-based practices (e.g., providing information about potential treatment options; Aoki et al., 2019) and have demonstrated effectiveness in a preliminary RCT with youth clients (Langer et al., 2022).

Critical Gap in the Literature: Measurement of SDM Practices

While further research is undoubtedly needed to better clarify SDM in youth psychotherapy, a major hinderance to this work is the lack of established measures of SDM practices specific to youth psychotherapy. To conduct high-quality research that examines practices clinicians can use to encourage SDM with their youth psychotherapy clients, it would be advantageous to utilize an observer-rated measure of SDM practices, as observer-rated measures are considered the “gold standard” when assessing in-session provider behaviors (Hogue et al., 2015). In their review, Gärtner et al. (2018) demonstrated that the vast majority of measures designed to assess SDM were developed for adult medical settings. Of the 40 measures identified, about half (n = 18), were observer-rated, of which 14 were designed for generic adult healthcare settings, two for adult oncology, one for pediatric palliative care, and one for adult psychiatry. Notably, none of the identified measures assessed SDM in youth psychotherapy. Gärtner et al. (2018) also concluded that there were significant flaws in the development of about half of the observer-rated measures, including psychometric concerns (e.g., poor inter-rater reliability).

Of the observer-rated measures that demonstrated favorable psychometric properties, the most commonly used (e.g., Observing Patient Involvement Scale or OPTION, Elwyn, 2003), demonstrated “good” inter-rater reliability (intraclass correlation coefficients or ICCs $\geq$ .60;

Cicchetti, 1994). Gärtner et al. (2018) also commented on the degree to which previous studies had established the construct validity of their scores (i.e., degree to which scores on a measure assess the intended construct; Kazdin, 2016). A common method for assessing construct validity among these studies was to correlate SDM scores with scores on measures of similar and dissimilar constructs (i.e., multi-trait multi-method; Campbell & Fiske, 1959). Construct validity was considered "confirmed" in the identified studies if (a) the correlation between SDM scores and scores on theoretically similar measures (e.g., confidence in decision making, Knapp et al., 2009) was "small" to "large" ($r \geq .10$; Rosenthal & Rosnow, 1984), and (b) SDM scores were less strongly correlated with dissimilar constructs (e.g., anxiety, Edwards et al., 2003). However, about half of the identified measures' scores did not demonstrate construct validity. Thus, a critical gap in the SDM field, and SDM in youth psychotherapy in particular, is the lack of a psychometrically sound, observer-rated measure of SDM practices in youth psychotherapy.

Current Study

EBPs can help reduce the burden of mental health disorders for youth, but their effectiveness may be bolstered if they better fit the needs of individual clients (Chambers & Norton, 2016; Marques et al., 2019). SDM may help clinicians adapt EBPs, but research is needed to determine if SDM practices help improve care quality within the context of EBP delivery in community mental health centers (Cheng et al., 2017; Eliacin et al., 2015; Grim et al., 2016; Hale et al., 2020; Langer & Jensen-Doss, 2018). To conduct this research, a psychometrically sound, observer-rated measure of SDM practices designed for youth psychotherapy is needed (Blumenthal-Barby et al., 2019; Langer & Jensen-Doss, 2018).

The paucity of SDM research in youth mental health treatment may be due, at least in part, to the absence of established measures designed to assess SDM practices for youth

psychotherapy. Most SDM measures (e.g., OPTION, Elwyn et al., 2005) are designed for medical or adult mental health populations (Gärtner et al., 2018). Given the differences between these populations and youth psychotherapy, a measure of SDM practices specific to youth psychotherapy is needed. This study aimed to fill this gap with a novel measure of SDM practices in youth psychotherapy—the Shared Decision Making Practices in Youth Mental Health scale (SDM-YMH). The SDM-YMH items were designed to map onto previously identified SDM practices (Cecilione Herbst, 2022; Clayman & Makoul, 2009; Elwyn et al., 2005) and theory regarding SDM in youth psychotherapy (Langer & Jensen-Doss, 2018; see *Shared Decision Making Practices in Youth Mental Health Scale Development* below for more detail).

It is important to establish the reliability and validity of SDM-YMH scores, especially given that poor psychometric evidence has been a consistent hinderance in the measurement of SDM (Gärtner et al., 2018). Moreover, it is important to continue examining the links between SDM practices, the alliance, and clinical outcomes (especially in youth receiving EBPs in community mental health centers, where most youth receive psychotherapy services). While there is theoretical grounding for this association (i.e., the link between SDM practices, the alliance, and clinical outcomes; Fjermestad et al., 2016; Langer & Jensen-Doss, 2018), minimal empirical evidence exists to support these linkages in the context youth psychotherapy (Langer et al., 2022). This study aimed to accomplish these goals by examining the reliability and validity of SDM-YMH scores in a sample of youth receiving evidence-based psychotherapy in community settings and by evaluating the link between SDM practices, the alliance, and clinical outcomes.

Hypotheses

Hypothesis 1

It was hypothesized that SDM-YMH item and subscale scores would demonstrate at least "good" inter-rater reliability (ICCs ≥ .60; Cicchetti, 1994; Gärtner et al., 2018).

Hypothesis 2

It was hypothesized that SDM-YMH item and subscale scores would demonstrate greater shared variance with theoretically similar constructs, such as goal and session agenda setting (convergent) and less shared variance with theoretically dissimilar constructs, such as adherence to cognitive-behavioral therapy or family therapy practices (discriminant; Campbell & Fiske, 1959; Foster & Cone, 1995; Kazdin, 2016). Specifically, SDM-YMH scores were expected to demonstrate (a) "large" correlations (r's ≥ .36; Rosenthal & Rosnow, 1984) with scores on measures of similar constructs (e.g., goal setting) without being redundant (r's < .7; Kline, 1979) and (b) "small" (.10 ≤ r ≤ .23; Rosenthal & Rosnow) to non-significant (p > .05) correlations with scores on measures of dissimilar constructs (e.g., protocol adherence, competence; Gärtner et al., 2018).

Hypothesis 3

It was hypothesized that greater use of SDM practices (i.e., higher SDM-YMH scores) would be predictors of positive change in the alliance and youth clinical outcomes throughout treatment. That is, higher observed SDM practices were hypothesized to predict a stronger alliance (youth-clinician, caregiver-clinician; Langer & Jensen-Doss, 2018) and greater improvement in youth clinical (internalizing) outcomes (Ackerman & Hilsenroth, 2003; Fjermestad et al., 2020; Langer & Jensen-Doss, 2018)

Hypothesis 3a. Higher SDM-YMH subscale scores were hypothesized to predict the slope of change in youth-clinician alliance (e.g., Fjermestad et al., 2020; Langer & Jensen-Doss, 2018).

Hypothesis 3b. Higher SDM-YMH subscale scores were hypothesized to predict the slope of change in caregiver-clinician alliance (e.g., Fjermestad et al., 2020; Langer & Jensen-Doss, 2018).

Hypothesis 3c. Higher SDM-YMH subscale scores were hypothesized to predict the slope of change in youth-reported internalizing symptoms (e.g., Fjermestad et al., 2020; Langer & Jensen-Doss, 2018; Loh et al., 2006; Street et al., 2008)

Hypothesis 3d. Higher SDM-YMH subscale scores were hypothesized to predict the slope of change in caregiver-reported youth internalizing symptoms (e.g., Fjermestad et al., 2020; Langer & Jensen-Doss, 2018; Loh et al., 2006; Street et al., 2008).

Method

Shared Decision Making Practices in Youth Mental Health Scale Development

An iterative process was used to develop the SDM-YMH. The framework used to guide the development of the SDM-YMH is Clark and Watson's (1995) principles for quality scale development, which are consistent with the Standards for Educational and Psychological Testing guidelines (American Education Research Association, 2014) and gold-standard methods (Carroll, 2000; Hogue et al., 2008; McLeod et al., 2018). In this framework, validity is determined by examining a measure's content, construct, and external validity. Content validity, involving a precise and detailed conceptualization of the items (i.e., SDM practices), is established by (a) determining the nature and scope of the measure of interest and (b) creating item content. Next, construct validity is established by testing the measure with a representative

sample and evaluating item distributions, reliability, and score convergent and discriminant validity using an MTMM approach (Campbell & Fiske, 1959). Once construct validity evidence is obtained, external validity (e.g., predictive validity) can be established.

Item and Subscale Development. SDM-YMH items were designed to capture SDM practices delivered by clinicians to youth and caregivers in youth psychotherapy. Several sources were utilized in the generation of items. First, core SDM practices that were identified as part of a large-scale review (Clayman & Makoul, 2009) were used as the basis of the SDM-YMH items (e.g., presenting treatment options to patients, presenting evidence for treatment options, discussing patients' values, agreeing on a treatment plan). Then, these items were cross-referenced with items on the OPTION (Elwyn et al., 2005) and OPTION-5 (Elwyn et al., 2013), which are the two most used observer-rated measures of SDM in mental health settings (Gärtner et al., 2018). Items included on the OPTION or OPTION-5 that were missing from Clayman and Makoul (2009)'s list of core SDM practices were then added to the SDM-YMH's item list. Next, SDM practices resulting from a review of RCTs of SDM interventions in mental health settings (Cecilione Herbst, 2022) were integrated into the item list. Lastly, proposed adaptations of SDM to youth mental health were integrated into the SDM-YMH item list and corresponding item definitions based on the theoretical model developed by Langer and Jensen-Doss (2018).

To capture SDM practices in youth psychotherapy, the SDM-YMH has two subscales: (1) Youth: SDM practices directed at youth and (2) Caregiver: SDM practices directed at caregivers. The items were expected to be the same across subscales (e.g., exploring values as they relate to treatment decisions; Aoki et al., 2019), but item exemplars demonstrated the differences in delivery of SDM practices across target individuals (e.g., clinician exploring caregiver's parenting philosophies versus clinician exploring youth's preferences regarding

where to begin on a fear ladder). Coders were asked to recognize to whom the SDM practice was directed and code accordingly. The inclusion of subscales, such as Youth and Caregiver, is similar to existing therapy process measures (McLeod & Weisz, 2005; Shelef & Diamond, 2008) and was expected to help determine the role SDM practices directed toward youth and caregivers play in youth psychotherapy (Haine-Schlagel & Walsh, 2015; Langer & Jensen-Doss, 2018; Liverpool et al., 2020). See Table 1 for the full item list.

Content Validity: Youth and Caregiver Items

A multi-step process was used to establish the content validity of the SDM-YMH items. First, definitions for each item were written according to the following criteria: (a) conciseness, (b) one practice per item, and (c) focus on observable behavior (Ware, Tugenberg, & Dickey, 2003). Second, an adapted Delphi technique was used to arrive at expert consensus about the items (Normand et al., 1998). The Delphi technique is an iterative group judgment approach to establish expert consensus. The items were distributed to 14 experts in the fields of youth mental health and SDM (i.e., individuals who have published on shared decision-making in the past; e.g., Margarita Alegría, Alegría et al., 2018; Amanda Jensen-Doss, Langer & Jensen-Doss, 2018) who were asked to rate the content validity of each item ("Is this practice 'essential', 'useful, but not essential', or 'not necessary' to SDM in youth mental health treatment?") and provided the opportunity to edit or add items.

Of the 14 experts contacted to assess content validity, eight replied (Amanda Jensen-Doss, Richard Street, Adrian Edwards, Lindsay Holly, Shaun Liverpool, Laura Shay, Blanche Wright, Jennifer Elston Lafata). Content validity ratios (CVR) were calculated for each item using the following formula: CVR = (ne - N/2)/(N/2) where ne = number of experts indicating "essential", N = number of experts. CVR values range from +1 to -1, with positive values

Table 1

Shared Decision Making Practices in Youth Mental Health (SDM-YMH) Items

Item	Definition	CVR
1. Elicit Goals for Therapy	Clinician encourages client/caregiver to share their goals for therapy.	.75
2. Define/Explain Problem	Clinician (1) identifies diagnosis/presenting problem(s) AND/OR (2) teaches client/caregiver about diagnosis/presenting problem(s).	0.00
3. Decision Identification	Clinician identifies that a treatment decision needs to be made.	1.00
4. Define Roles (desire for involvement)	Clinician encourages client/caregiver to share their thoughts/feelings regarding (1) how they would like to be involved in treatment decisions, (2) who else should be involved in treatment decisions, AND/OR (3) discusses the client/caregiver's ability to participate in the decision-making process.	1.00
5. Present Options	Clinician identifies or shares options for treatment (including the option of no action).	.75
6. Equipoise	Clinician shares that there is more than one path to desired treatment outcomes.	.75
7. Present Evidence	Clinician shares evidence-based information with client/caregiver as it relates to a treatment option.	.75
8. Patient Values/Preferences	Clinician (1) encourages client/caregiver to share their preferences/values during a discussion of treatment planning AND/OR (2) relates client/caregiver preferences/values explicitly to treatment decision.	.50
9. Discuss Pros/Cons (benefits/risks/costs)	Clinician encourages client/caregiver to share their opinions regarding the benefits/costs of treatment option(s).	.50
10. Mutual Agreement	Clinician discusses and/or processes the differing opinions of client and caregiver OR of clinician and family in an effort to reach a middle ground during treatment planning discussions.	0.00
11. Make or Explicitly Defer Decision	Clinician engages the client/caregivers in a discussion to decide on (1) a specific treatment option OR (2) that the treatment decision is being delayed to a future time.	.75
12. Check/Clarify Understanding	Clinician evaluates client/caregiver understanding of treatment decision(s) being made.	.50
13. Arrange Follow Up (to discuss treatment plan)	Clinician discusses when/why treatment decision/plan would need to be revisited with client/caregiver.	.50
14. Clinician Recommendations/ Knowledge*	Clinician shares their opinion/recommendation regarding how to proceed in treatment.	-.25

Note. CVR = Content Validity Ratio; ***** Removed from measure

indicating more than half of the experts rated the item as essential (i.e., the item has at least some content validity). Items with CVRs below the threshold for the number of experts (0.75 for 8 experts, $p < .05$; Ayre & Scally, 2014) were considered for exclusion. Of the original 14 items, seven had CVRs of 0.75 or higher, six had CVRs between 0.50 and 0, and one had a CVR less than zero. The item with a negative CVR (Clinician Recommendations/Knowledge) was removed from the SDM-YMH before coding commenced. Items with CVRs below the threshold of 0.75 were retained as long as (a) at least half of the experts agreed that it was a necessary SDM practice and (b) no raters listed it as "not necessary." A liberal approach to retaining items was taken to ensure any potentially useful SDM practices were not omitted. Additionally, the lower-than-anticipated number of expert raters increased the difficulty of obtaining CVRs of 0.75 or higher (see Table 1 above).

Scoring Strategy. As the SDM-YMH was designed to assess clinician-delivered SDM practices, an extensiveness scoring strategy, a standard and commonly used scoring strategy (Carroll, 2000; Hogue et al., 2008; Southam-Gerow et al., 2016), was used to estimate the extent to which clinicians delivered SDM practices during a session using a 7-point Likert-type scale: 1 (*not at all*), 5 (*considerably*), and 7 (*extensively*). This strategy was used to generate scores on SDM-YMH items.

Coding Manual Development. Once items were finalized, a draft of the SDM-YMH coding manual was produced. Procedures for developing exemplary therapy process coding manuals were followed, including item definitions, coding guidelines, and item distinctions (Hogue et al., 1996). SDM practices were not expected to vary across children, adolescents, and caregivers, but *how* the practices were delivered differed (e.g., use of cartoons with children; focus on the cost associated with treatment length with caregivers). Thus, item definitions and

exemplars illustrated how SDM practice delivery may vary across youth and caregivers. The coding manual was developed in collaboration with Drs. McLeod and Langer. Additionally, pilot coding was conducted to inform item definitions, exemplars, and distinctions.

Initial Evaluation of the Reliability and Validity of the SDM-YMH Items and Subscales

Data Source

This study used data from 150 psychotherapy sessions from 38 youth with primary anxiety problems treated by 26 clinicians from the Child Services and Treatment Enhancement Project Multisite Trial (Child STEPs Multisite Trial; Weisz et al., 2012), and the resulting data were used to examine the link between SDM practices, the alliance, and clinical outcomes. A multi-trait multi-method (MTMM) approach was used to evaluate the convergent and discriminant validity of scores on the SDM-YMH items and subscales (Youth, Caregiver) scores (Campbell & Fiske, 1959), using archived treatment integrity and alliance data from the Child STEPs Multisite Trial, which was generated as part of the Treatment Integrity Measurement Study (TIMS; R01 MH1086539; PIs McLeod, Southam-Gerow).

Treatment Integrity Measurement Study (TIMS)

TIMS was designed to develop observer-rated treatment integrity (i.e., degree to which an intervention is delivered as intended; Allen et al., 2012; Perepletchikova et al., 2007) measures. To achieve this goal, all sessions from three studies for youth anxiety were coded: (a) Coping Cat Trial (Kendall et al., 2008), (b) Youth Anxiety Study (Southam-Gerow et al., 2010), and (c) Child STEPs Multisite Trial (Weisz et al., 2012). The current study used archived TIMS data from the Child STEPs Multisite Trial, an effectiveness trial conducted in community mental health centers. Youth in the Child STEPs Multisite Trial had to meet DSM-IV-TR (American Psychiatric Association, 2000) criteria or demonstrate clinically elevated anxiety, depression, or

conduct problems. Youth in the Child STEPs Multisite Trial were randomly allocated to one of three conditions: modular manualized treatment (Modular), standard manualized treatment (Standard), or usual care. Treatment integrity measures developed in TIMS (protocol adherence, competence, differentiation) were used to evaluate the construct validity (convergent, discriminant) of SDM-YMH item and subscale scores. Notably, TIMS only included youth with primary anxiety problems.

Child STEPs Multisite Trial data are relevant to the current study because the Child STEPs Multisite Trial (a) examined the effectiveness of EBPs and thus provides an opportunity to examine SDM practices in the context of EBPs delivered in community mental health centers and (b) collected weekly clinical outcome data, which provides an opportunity to determine if SDM-YMH subscale scores predict youth clinical outcomes. While the Child STEPs Multisite Trial included youth with presenting problems besides anxiety (as detailed below), this study only included youth with a primary presenting problem of anxiety, as treatment integrity data is only available for the anxiety subsample. Leveraging the existing treatment integrity data was essential to evaluate the construct validity of SDM-YMH scores in this sample.

Child STEPs Multisite Trial Participants

Youth. Child STEPs Multisite Trial sessions (N = 150) were coded with the SDM-YMH in the current study. Data are from 38 youth with primary anxiety problems who met the following criteria: (a) had at least two audible sessions and (b) received an EBP from only one study clinician. The 38 youth participants (52.6% male, 47.4% female) in these therapy sessions ranged from 8 to 13 years old (M = 9.84, SD = 1.65). Youth identified the following racial/ethnic identities: 60.5% White, 5.3% Black or African American, 2.6% Asian American, 2.6% Hispanic or Latinx, 26.3% Multiracial, and 2.6% Other. See Table 2 below for further detail.

Table 2

Youth Demographic Information

	***M* (*SD*) or %**		
	Modular	**Standard**	***t* or χ^2 value (p-value)**
N	16	22	
Age	9.94 (1.88)	9.77 (1.51)	.30 (.77)
Sex			.15 (.70)
Female	43.8	50.0	
Race/Ethnicity			8.18 (.15)
White	43.8	72.7	
Black	12.5	0.0	
Asian American	0.0	4.5	
Latinx	0.0	4.5	
Multiracial	37.5	18.2	
Other	6.3	0.0	
CBCL (pre) scores			
Total	63.63 (10.39)	65.27 (7.49)	-.57 (.57)
Internalizing	69.56 (9.33)	70.00 (6.72)	-.17 (.87)
Externalizing	55.06 (11.64)	59.00 (11.28)	-1.05 (.30)
Anxiety	69.88 (7.76)	69.64 (7.20)	.10 (.92)

Note. CBCL = Child Behavior Checklist

Caregivers. The 44 caregivers ranged in age from 25 to 74 (*M* = 39.38 years, *SD* = 9.44) and identified as the following: 65.9% female, 29.5% male, 4.5% did not report. Moreover, 38.7% of families reported a household annual income below $40,000. See Table 4 for detail.

Clinicians. The 26 clinicians (82.1% female, 17.9% male) were employed by the community mental health centers and randomly assigned to condition (Modular or Standard). Most clinicians (85%) held a master's degree and ranged from 27 to 59 years old (*M* = 40.34, *SD* = 9.67). Clinicians identified as the following: 42.9% White, 14.3% Black or African American, 21.4% Asian American, 3.6% Multiracial, and 3.6% Other. See Table 3 below for detail.

Child STEPs Multisite Trial Treatment Conditions

Modular. The Modular Approach to Therapy for Children with Anxiety, Depression, and Conduct Problems (MATCH; Chorpita & Weisz, 2005) protocol was administered in the Modular condition. Modules in MATCH correspond to the content delivered in the following EBPs: (a) Coping Cat (Kendall & Hedtke, 2006) a cognitive-behavioral treatment (CBT) program for youth anxiety, (b) Primary And Secondary Control Enhancement Training (PASCET; Weisz, Weersing, Valeri, & McCarty, 1999), a CBT program for youth depression, and (c) Defiant Children (Barkley, 2013) a parent training program for youth conduct problems. Flowcharts for each problem area detailed a default sequence of modules, but clinicians could deviate when clinically relevant. The current study only focused on youth with a primary presenting concern of anxiety.

Standard. In the standard condition, three manualized treatments with a prescribed order of sessions were administered: (a) Coping Cat, (b) PASCET, and (c) Defiant Child. Coping Cat was the primary treatment used in the included tapes.

Table 3

Clinician Demographic Information

	***M* (*SD*) or %**		
	Modular	**Standard**	***t* or χ^2 value (p-value)**
N	10	16	
Age	35.20 (6.81)	43.56 (9.96)	-2.33 (.03)
Sex			.01 (.94)
Female	80.00	81.25	
Years of Experience	3.67 (1.68)	8.67 (9.80)	-1.51 (.15)
Race/Ethnicity			4.67 (.32)
White	50.00	56.25	
Black	.00	12.50	
Asian American	40.00	12.50	
Not reported	.00	12.50	
Other	10.00	6.25	
Area of Specialty			3.19 (.67)
Social worker	40.00	37.50	
Behavior specialist	30.00	12.50	
Psychologist	20.00	12.50	
Mental health counselor	10.00	25.00	
Not reported	.00	6.25	
Other	.00	6.25	

Table 4

Caregiver Demographic Information

	***M* (*SD*) or %**		
	Modular	**Standard**	***t* or χ^2 value (p-value)**
N	19	25	
Age	39.88 (7.37)	39.00 (10.93)	.29 (p = .78)
Sex			
Female	14 (73.7%)	15(60%)	p = .56
Income <$40,000	56%	73%	

Assessment Plan and Measures

Treatment integrity (protocol adherence, competence, differentiation) and alliance data were generated as part of TIMS. All available treatment sessions have been coded from each anxiety case in the Child STEPs Multisite Trial with the treatment integrity and alliance measures in Table 5. Youth clinical outcome data were collected at each session.

SDM Practices

The Shared Decision Making Practices in Youth Mental Health scale (SDM-YMH) was used to assess clinician-directed practices that promote SDM with youth and caregivers (i.e., SDM practices). The SDM-YMH is a 13-item observed-rated measure used to assess SDM practices in youth psychotherapy and is comprised of two subscales (i.e., Youth and Caregiver). The scope of the SDM-YMH is restricted to one contributing factor of SDM—clinician behaviors (see Figure 1). As aforementioned, examining the steps clinicians can take to promote SDM—i.e., SDM practices—(as opposed to focusing on actions taken by clients and/or caregivers) is an important first step in determining how providers can better promote client-centered care that honors client autonomy. Other contributing factors of SDM (e.g., client/caregiver involvement in conversations regarding treatment planning) are outside the scope of the SDM-YMH. Importantly, SDM itself (i.e., interactive process of clinicians and

clients collaborating and making healthcare decisions) is also outside the scope of SDM-YMH. That is, SDM-YMH items only focus on clinician behaviors contributing to the collaborative conversations about making treatment decisions.

Table 5

Key Constructs and Associated Measures for the Current Study

Construct	Measure	Type	Informant	Sampling	Role in Current Study
SDM Practices	SDM-YMH (Youth & Caregiver Subscales)	O	O	4 sessions	--
Treatment Goals	TPOCS-RS (Goal Setting & Session Agenda Items)	O	O	Weekly	Convergent
Protocol Adherence	CBAY-A Model Subscale	O	O	Weekly	Discriminant
Competence	CBAY-C Model Subscale	O	O	Weekly	Discriminant
Differentiation	TPOCS-RS (Psychodynamic Item & Family Subscale)	O	O	Weekly	Discriminant
Alliance	TPOCS-A (Youth & Caregiver Subscales)	O	O	Weekly	Discriminant & Predictive
Clinical Outcomes	WPC—Internalizing	S	C, Y	Weekly	Predictive

Note. Measure type: S=survey; O=observation; Informant: O=observer; C=caregiver; Y = youth; SDM-YMH = Shared Decision Making Practices in Youth Mental Health Scale; TPOCS-RS = Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale, CBAY-A = Cognitive Behavioral Therapy Adherence Scale for Youth Anxiety, CBAY-C = Cognitive Behavioral Therapy for Anxiety in Youth Competence Scale, TPOCS-A = Therapy Process Observational Coding System for Child Psychotherapy—Alliance, WPC = Weekly Phone Checklist

Of note, in this study, “treatment decisions” were operationally defined as *decision points throughout treatment that may reasonably impact the course or outcome of treatment in a meaningful way*. For example, the following were considered treatment decisions in this study: (a) whether or not to engage in exposures, (b) what order to engage in certain treatment activities (e.g., whether to do relaxation before or after thought challenging), (c) what presenting problem to address first in treatment (e.g., anxiety or depression), or (d) whether to continue or discontinue treatment. SDM-YMH items were not coded when clinicians were using a collaborative style for the purposes of rapport building. That is, clinician behaviors that could be considered collaborative and similar to some SDM practices (e.g., presenting options and eliciting patient preferences with regards to what game to play in session) but were not delivered in the context of engaging clients/caregivers in a conversation about treatment decisions were not coded. The following are examples of clinician questions that would not be codable with the SDM-YMH: (a) “What do you want to earn for following directions this week?”, (b) “We have 10 minutes left, what game would you like to play?”, or (c) “What are some fun things you and your family do together?".

After watching or listening to entire therapy sessions, coders rated each item on a 7-point Likert-type extensiveness (i.e., frequency *and* thoroughness of the delivery of the practice element) scale from 1 (*not at all*) to 7 (*extensively*). SDM-YMH Youth subscale scores demonstrated “good” inter-rater reliability (ICC[2,2] = .62; Cicchetti, 1994), and SDM-YMH Caregiver subscale scores demonstrated “excellent” inter-rater reliability (ICC[2,2] = .81; Cicchetti, 1994). See *Shared Decision Making Practices in Youth Mental Health Scale* below for a detailed description of scale development and coding procedures.

Treatment Goals. Relevant items from the Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale (TPOCS-RS; McLeod et al., 2015) were used to establish convergent validity of SDM-YMH scores. The TPOCS-RS is a 47-item observer-rated measure of how extensively clinicians deliver discrete practice elements (e.g., relaxation, exposure) found across multiple treatment approaches for youth psychotherapy (i.e., CBT, family, client-centered, and psychodynamic). After watching or listening to entire therapy sessions, coders rated each item on a 7-point Likert-type extensiveness scale from 1 (*not at all*) to 7 (*extensively*). TPOCS-RS item and subscale scores have demonstrated reliability and construct validity in previous studies (McLeod et al., 2015; McLeod & Weisz, 2010; Southam-Gerow et al., 2010). For example, the "Goal Setting" item demonstrated "excellent" inter-rater reliability (ICC = .82) and "Session Agenda" demonstrated "good" inter-rater reliability (ICC = .62) in a previous study (McLeod et al., 2015). "Goal Setting" (clinicians initiating discussions of long-term treatment goals) and "Session Agenda" (clinicians initiating discussion of session goals) item scores were used to assess the convergent validity of SDM-YMH items and subscale scores, as these items both involve clinicians engaging clients and caregivers in conversations related to treatment decisions. In this study, TPOCS-RS Goal Setting scores demonstrated "good" inter-rater reliability (ICC[2,2] = .70; Cicchetti, 1994); whereas TPOCS-RS Session Agenda scores demonstrated "poor" inter-rater reliability (ICC[2,2] = .38; Cicchetti, 1994). Since the 95% confidence interval did not contain zero, this item was retained in analyses (Rousson et al., 2003; Tan & Tan, 2010). These criteria for inclusion/exclusion of items are used throughout this project (see below).

Treatment Integrity

Protocol Adherence. Protocol adherence is the degree to which a clinician delivers the treatment as prescribed in the protocol (Southam-Gerow et al., 2016). In this study, protocol adherence was measured via the observer-rated Cognitive Behavioral Therapy Adherence Scale for Youth Anxiety (CBAY-A; Southam-Gerow et al., 2016). The CBAY-A is a 22-item measure designed to capture the delivery of practice elements commonly found in CBT protocols for youth anxiety. The Model subscale of the CBAY-A was used in the current study as a measure of protocol adherence, as the 11 items of the Model subscale map onto key practice elements of CBT for anxiety (i.e., Psychoeducation—Anxiety, Emotion Education, Fear Ladder, Relaxation, Cognitive—Anxiety, Problem Solving, Self-Reward, Coping Plan, Exposure Preparation, Exposure, Exposure Debrief; Kendall, 1994; McLeod et al., 2018; Southam-Gerow et al., 2016). These same 11 items also comprise the Model subscale of the CBAY-C (see below). After watching or listening to entire therapy sessions, coders rated each item on a 7-point Likert-type extensiveness scale from 1 (*not at all*) to 7 (*extensively*). CBAY-A Model subscale scores have demonstrated reliability and validity in previous studies (McLeod et al., 2018; McLeod et al., 2019; Southam-Gerow et al., 2016). For example, the CBAY-A item scores demonstrated "fair" to "excellent" inter-rater reliability in previous studies (ICC[2,2]s ranged from .48 to .80; M = .77, SD = .15; Southam-Gerow et al., 2016) In this study, scores on the CBAY-A Model subscale demonstrated "excellent" (Cicchetti, 1994) inter-rater reliability (subscale ICC[2,2] = .82).

Competence. Broadly, competence refers to the quality with which treatment is delivered (Perepletchikova & Kazdin, 2005). While competence has been further delineated into technical (skillfulness and responsiveness in delivering techniques found in a specific treatment; Barber et al., 2007; McLeod et al., 2018) and global (clinical skills and judgment that cut across treatment

modalities; Barber et al., 2007; Brown et al., 2018) domains, measures designed to assess these subsets of competence may not be able to completely distinguish between the two (Cecilione et al., 2021). Thus, only one measure of competence was utilized in this study to assess competence—the CBT for Anxiety in Youth Competence Scale (CBAY-C; McLeod et al., 2018). The CBAY-C is a 23-item observer-rated measure to assess technical competence in delivering core practice elements commonly found in CBT protocols for youth anxiety. Similar to the CBAY-A, the Model subscale of the CBAY-C was used in the current study as a measure of competence, as the 11 items of the Model subscale map onto key practice elements of CBT for anxiety as well as mirror those of the CBAY-A Model subscale. After watching or listening to entire treatment tapes, coders rated each item on a 7-point Likert-type competence scale (skillfulness *and* responsiveness of the delivery of each practice element) from 1 (*very poor*) to 7 (*excellent*). CBAY-C item and subscale scores have demonstrated reliability and validity in previous studies; for example, CBAY-C item scores have demonstrated inter-rater reliability at the item level (ICC[2,2]s ranged from .37 to .80; M = .67, SD = .11), as well as representativeness, convergent and discriminant validity (see McLeod et al., 2018; McLeod et al., 2019). In this study, the CBAY-C Model subscale demonstrated "excellent" (Cicchetti, 1994) inter-rater reliability (subscale ICC[2,2] = .81).

Differentiation. Differentiation refers to whether a treatment under investigation includes proscribed elements (Perepletchikova & Kazdin, 2005). For example, if a clinician who was instructed to implement individual CBT for youth anxiety incorporates a considerable number of practice elements commonly found in systemic family therapy, this would constitute differentiation. This study used the TPOCS-RS Psychodynamic and Family subscales to represent differentiation (i.e., delivery of non-CBT practices). Scores on the TPOCS-RS

subscales have demonstrated reliability and validity (McLeod et al., 2015; McLeod & Weisz, 2010; Southam-Gerow et al., 2010). In McLeod et al. (2015), the Psychodynamic subscale scores demonstrated "good" (ICC = .74) inter-rater reliability, and the Family subscale scores demonstrated "excellent" (ICC = .94) inter-rater reliability. In this study, TPOCS-RS Family subscale scores demonstrated "excellent" (Cicchetti, 1994) inter-rater reliability (subscale ICC[2,2] = .95). However, in this study, TPOCS-RS Psychodynamic subscale scores demonstrated "poor" (Cicchetti, 1994) inter-rater reliability (mean item ICC[2,2] = .16). Therefore, three items (Addresses Client Resistance, Explores Past, Addresses Transference) were excluded from the subscale due to poor inter-rater reliability (ICCs < .40; Cicchetti 1994) with 95% confidence intervals including zero (Rousson et al., 2003; Tan & Tan, 2010). Only one item (Interpretation) was retained in the TPOCS-RS Psychodynamic Item (ICC[2,2] = .34; 95% CI: .09—.52). Throughout the remainder of the document, this item (Interpretation) will be referred to as the TPOCS-RS Psychodynamic Item.

Together, the CBAY-A Model scale, CBAY-C Model scale, and TPOCS-RS Family subscale and Psychodynamic Item scores were used to assess discriminant validity of SDM-YMH item and subscale scores.

Alliance. The alliance refers to the affective and collaborative components of the client-provider relationship (Elvins & Green, 2008; Shirk & Saiz, 1992) and has been linked to positive clinical outcomes in youth psychotherapy (McLeod, 2011). The Therapy Process Observational Coding System-Alliance Scale (TPOCS-A; McLeod & Weisz, 2005) is a 9-item observer-rated measure of youth- and caregiver-clinician alliance. That is, the TPOCS-A is comprised of two subscales—Youth and Caregiver—each including the same 9 items. Coders watched or listened to entire therapy tapes and then rated items on a 6-point Likert-type scale ranging from 0 (*not at*

all) to 5 (*a great deal*). TPOCS-A scores have demonstrated reliability, internal consistency, convergent, and predictive validity in previous studies (Fjermestad et al., 2012; Liber et al., 2010; McLeod et al., 2017). For example, in McLeod and Weisz (2005), the TPOCS-A Youth subscale scores demonstrated "fair" to "good" inter-rater reliability (ICC[2,2]s ranged from .40 to .75; M = .59, SD = .10) and convergent validity with a self-report alliance instrument (TASC; Shirk & Saiz, 1992).

In this study, TPOCS-A Youth scale scores demonstrated "excellent" (Cicchetti, 1994) inter-rater reliability (subscale ICC[2,2] = .87) as well as "excellent" internal consistency (α = .90; Cronbach, 1951). Yet, in this study, TPOCS-A Caregiver subscale scores demonstrated "poor" (Cicchetti, 1994) inter-rater reliability (mean item ICC[2,2] = .32). Therefore, three items (Hostile Manner, Positive Affect, Client Discomfort, Client/Therapist Discomfort, Client Noncompliance) were considered for exclusion from the subscale score due to poor inter-rater reliability (ICCs < .40; Cicchetti 1994) with 95% confidence intervals including zero. However, these items were retained, since the TPOCS-A Caregiver subscale has demonstrated validity and reliability in previous works (e.g., mean item-level ICC = .61, SD = .10 in McLeod & Weisz, 2005). Thus, all items were retained to improve interpretability of findings. The TPOCS-A Caregiver subscale scores demonstrated "good" (Cicchetti 1994) inter-rater reliability (ICC[2,2] = .52) and "excellent" internal consistency (α = .90; Cronbach, 1951) in the current sample.

In the current study, the TPOCS-A subscale (Youth, Caregiver) scores were used to assess the discriminant and predictive validity of the SDM-YMH subscale scores. See Table 6 below for descriptive data.

Table 6. *Descriptive Data for Construct Validity Measures*

Item	N	Range	Min.	Max.	Mean(SD)	Skewness	Kurtosis	ICC(2,2)
TPOCS-RS Goal Setting	150	6.00	1.00	7.00	2.32(1.34)	1.16	1.17	.70
TPOCS-RS Session Agenda	150	4.00	1.00	5.00	2.21(0.87)	0.71	0.37	.38
CBAY-A	150	6.00	1.00	7.00	4.27(1.56)	-0.44	-0.70	.82
CBAY-C	122	5.00	2.00	7.00	4.00(1.06)	.030	-0.40	.81
TPOCS-RS Family	150	2.50	1.00	3.50	1.31(0.47)	2.14	4.96	.95
TPOCS-RS Psychodynamic	96	1.00	1.00	2.00	1.04(0.15)	4.47	21.46	.34
TPOCS-A Youth	129	3.50	0.94	4.44	3.35(0.69)	-1.23	1.54	.87
TPOCS-A Caregiver	65	2.63	1.13	3.75	2.26(0.52)	0.36	0.70	.52

Note. TPOCS-RS = Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale, CBAY-A = Cognitive Behavioral Therapy Adherence Scale for Youth Anxiety, CBAY-C = Cognitive Behavioral Therapy for Anxiety in Youth Competence Scale, TPOCS-A = Therapy Process Observational Coding System for Child Psychotherapy—Alliance

Clinical Outcomes. The 12-item Weekly Phone Checklist (WPC) Youth and Caregiver measures yield Total, Internalizing, and Externalizing scales, which were adapted from the Child Behavior Checklist and Youth Self Report (Achenbach & Rescorla, 2011) to track clinical outcomes. Higher scores indicate greater clinical impairment and distress. Favorable reliability and validity data have been reported (Achenbach & Rescorla, 2011). The WPC Internalizing scale (Youth-, Caregiver-report) was used to assess predictive validity of the SDM-YMH subscale scores.

Observer-Rated Coding Procedures for the SDM-YMH

Coders, Coder Training, and Coding Procedures

Coders for this study included the primary author and five additional coders (60% bachelor's level, 40% undergraduate level; 80% female, 20% male, 80% White, 20% Asian American, 40% with zero prior coding experience). Training for SDM-YMH began with the independent review of the coding manual and a review of exemplar sessions that represented each item. Archived treatment integrity ratings from TIMS were used to identify sessions where treatment goals were discussed with youth and caregivers to ensure variability in SDM practices across SDM-YMH subscales. Next, coders independently coded 10 tapes to establish mutual understanding of item definitions. Coders then independently coded 20 sessions to establish initial inter-rater reliability. Coders were considered ready for independent coding ("certified") when ICCs for the most commonly used items (i.e., items that were observed in at least half of the certification tapes—Elicit Goals for Therapy, Define/Explain Problem, Present Options, Patient Values/Preferences, Check/Clarify Understanding, Make/Defer Decision, Make/Defer Decision,) were at least "good" (ICC ≥ .60; Cicchetti, 1994). Then, coders were assigned to code sessions at random. Due to the availability of coders, the primary author was Coder 1 (i.e., coded

all sessions in the sample; *N* = 150), and the remaining co-coders were Coder 2 (i.e., were randomly assigned the remainder of the tapes to double code). At bi-weekly coding meetings, coders discussed difficult sessions, and ICCs were continually examined to prevent coder drift (Margolin et al., 1998).

Session Sampling

Four sessions were selected from each case (*N*=152; see *Power Analysis*). To sample different treatment phases, treatment was divided into beginning, middle, and end stages, created by dividing the number of sessions by three (McLeod & Weisz, 2010). Two sessions were randomly selected from the beginning stage of treatment, and one session was selected from the middle and end stages. Two sessions were selected from the beginning stage of treatment, as more conversations regarding treatment planning were expected to occur at the outset of treatment, increasing the opportunity to observe clinicians implementing SDM practices. Archived TIMS data was used to ensure that youth and caregivers were present in at least two sessions. This procedure is similar to established methods (e.g., McLeod & Weisz, 2005) and was expected to provide opportunities to observe clinicians using SDM practices to discuss treatment decisions with youth and caregivers. For cases where there were not enough sessions in which caregivers were present (*N* = 6 cases), a randomly selected tape from that same phase of treatment was selected as a substitute. There were also two cases in which an insufficient number of tapes were available from each phase. Therefore, the final sample size was *N* = 150.

Observer-Rated Coding Procedures for Other Included Measures

As aforementioned, scores on other key observer-rated measures were generated as part of the TIMS (R01 MH1086539; PIs McLeod, Southam-Gerow) project prior to the commencement of the current project. The procedures for generating scores on these observer-

rated measures (i.e., TPOCS-RS, TPOCS-A, CBAY-A, and CBAY-C)were similar to those used to generate SDM-YMH scores and are detailed below.

The CBAY-C coding team was comprised of two female clinical psychology doctoral students (50% Latinx, 50% White). The CBAY-A, TPOCS-RS, and TPOCS-A coding team was comprised of three female clinical psychology doctoral students (33% Asian American; 66% White). One coder served on both teams. Coders were blind to the treatment condition (i.e., modular, standard, or usual care) and coded sessions in a random order. Principal investigators trained coding teams separately over about a three-month period until they were considered certified (i.e., reached item-level inter-rater reliability of ICC[2,2] $\geq$ .60).

Similar to the current study, coders who coded the aforementioned observer-rated measures read and discussed the relevant scoring manuals. Principal investigators also reviewed coded sessions with the coders to ensure mutual understanding of the codes. Then, coders coded sessions independently and participated in regular meetings in which results of the practice coding were discussed. Lastly, coders started coding sessions for the certification phase, during which they were required to reach an "good" (Cicchetti, 1994) level of inter-rater reliability across 32 recordings (ICC[2,2] $\geq$ .60). Psychotherapy tapes were recorded as audio and video files.

After coders were considered "certified", they started independently coding randomly assigned sessions. Principal investigators and coders met regularly throughout the independent coding phase to discuss difficulties in coding. Inter-rater reliability was also continually assessed throughout coding to prevent coder drift (Margolin et al., 1998). If items fell below "good" inter-rater reliability (i.e., ICC[2,2] < .60; Cicchetti, 1994), additional training was provided (e.g.,

review of the coding manual, discussion regarding coding discrepancies, and/or group coding of problematic items).

Sampling of treatment sessions. Except for the intake and termination sessions for each client, all available sessions were selected from each client for coding and randomly assigned to coders. The final sample of sessions coded with the TPOCS-RS, TPOCS-A, CBAY-A, and CBAY-C consisted of 796 coded sessions (244 in the modular condition, 359 in standard condition, and 193 sessions in usual care).

Results

A multi-trait multi-method (MTMM; Campbell & Fiske, 1959) approach was used to examine score convergent and discriminant validity. Consistent with MTMM, measures of the same trait should converge, whereas measures of different traits should not. Within the MTMM approach, a multidimensional treatment integrity measurement model (Hogue, 2002; Hogue et al., 1996; McLeod et al., 2013; Sutherland et al., 2013) that maps onto the MTMM framework was used to evaluate construct validity of SDM-YMH scores. The multidimensional model was designed for this purpose via the assessment of SDM practices, protocol adherence, competence, and differentiation (see Table 5). This model was designed to differentiate between different components of treatment delivery. The model also included the alliance (youth-clinician, caregiver-clinician) and youth clinical outcomes, which allowed for examining the degree to which SDM practices predicted care quality.

Preliminary Analyses: Youth and Caregiver Items

Preliminary Data Inspection

Primary data analyses were preceded by an examination of the distributional properties of SDM-YMH item and subscale scores (see below). There was no missing SDM-YMH data. The

aforementioned goals of sampling were not completely met. That is, caregivers were present in 43% of tapes (as opposed to the targeted 50%), while youth clients were present in 86%. However, for all but two cases, four tapes were able to be extracted from each phase of treatment. There were slightly more sessions sampled from the middle of treatment than the end (48% of sessions were from the beginning of treatment, 29% from the middle, and 23% from the end). Moreover, caregivers' attendance in the included tapes were as follows: 48% of sessions from the beginning phase of treatment involved a caregiver, 41% of sessions from the middle of treatment, and 35% from the end of treatment. The attendance of youth clients was as follows: 81% of sessions from the beginning phase of treatment involved a youth client, 88% of sessions from the middle of treatment, and 94% from the end of treatment). Lastly, 96% of sessions contained at least some coded observations of SDM practices directed at either the youth or caregiver.

Only one significant difference regarding demographic data between groups (Modular, Standard) emerged; that is, clinicians in the Standard condition were older than those in the Modular condition (see Tables 2-4 above). Moreover, six clients who participated in the anxiety arm of the Child STEPs Multisite Trial were not included in the current study, as they did not have more than two sessions. Youth participants who were excluded from analyses did not differ from the current sample in any of the key demographic or clinical variables (see Table 7).

Table 7. *Comparisons between Current Sample and Parent Sample*

Variable	**M (SD) or %**		
Youth	**Included** (N = 38)	**Excluded** (N = 6)	***t* or χ^2 value (p-value)**
Age	9.84 (1.65)	11.00 (2.35)	1.41 (.17)
Sex			
Female	47.4	16.7	1.99 (.16)
Race/Ethnicity			5.73 (.33)
Caucasian	60.5	16.7	
African American	5.3	.00	
Asian American	2.6	16.7	
Latinx	2.6	.00	
Multiracial	26.3	33.3	
Other	2.6	0.0	
Not Reported	0.0	33.3	
CBCL (pre) scores			
Total	64.58 (8.73)	60.17 (10.27)	-1.13 (.27)
Internalizing	69.82 (7.81)	65.67 (11.27)	-1.14 (.26)
Externalizing	57.34 (11.45)	50.83 (11.20)	-1.30 (.20)
Anxiety	69.74 (7.34)	63.67 (10.33)	-1.78 (.08)

Note. CBCL = Child and Behavioral Checklist

Table 8

Shared Decision Making Practices in Youth Mental Health Scale Experimental Items: Descriptive Data and Inter-rater Reliability

Item	N	Range	Min.	Max.	Mean(SD)	Skewness	Kurtosis	ICC(2,2) or κ
Was there an opportunity for SDM?	149	1.00	0.00	1.00	0.97(0.13)	-4.77	20.98	.17 (κ)
Initiate client	140	4.00	1.00	5.00	1.42(0.73)	2.21	5.95	.27
Responsive client	51	6.00	1.00	7.00	3.07(1.90)	0.79	-0.62	-.37
Initiate caregiver	134	4.00	1.00	5.00	1.44(0.84)	1.92	3.19	-.02
Responsive caregiver	36	6.00	1.00	7.00	3.75(1.76)	0.21	-0.71	.67

Note. Was there an opportunity for SDM? (Yes or No); Initiate client = To what extent were SDM conversations initiated by the client?; Responsive client = When the client attempted to initiate SDM, how responsive was the clinician?; Initiate caregiver = To what extent Were SDM conversations initiated by the caregiver?; Responsive caregiver = When the caregiver attempted to initiate SDM, how responsive was the clinician?

Inter-Rater Reliability

Inter-rater reliability is the extent to which observers agree on the scores they assign when coding participants' performance on an instrument (Kazdin, 2016). Inter-rater reliability was calculated for each item by estimating intraclass correlation coefficients (ICCs) based on a two-way random effects model, average measures (ICC[2,2]s; Cicchetti, 1994; Shrout & Fleiss, 1979). Per Cicchetti (1994)'s guidelines, ICCs below .40 were considered *poor*, between .40 and .59 were considered *fair*, between .60 and .74 were considered *good*, and .75 and above were considered *excellent*. ICCs were calculated for the full sample. If an item had poor reliability (ICC < .40), it was considered for exclusion from further analyses. It was hypothesized that SDM-YMH item subscale scores would demonstrate at least "good" inter-rater reliability (ICC $\geq$ .60; Cicchetti, 1994).

Youth Items

Inter-rater reliability was a key factor in determining which SDM-YMH items to include or exclude from Youth and Caregiver subscales. Items that demonstrated "poor" (Cicchetti, 1994) inter-rater reliability were considered for exclusion. Per hypothesis 1, it was expected that SDM-YMH Youth items would demonstrate "good" to "excellent" inter-rater reliability (Cicchetti, 1994; Gärtner et al., 2018).

SDM-YMH Youth item inter-rater reliability ranged from ICC(2,2) = -.02 to .61 (*M* = .30 ICC, *SD* = .26; see Table 9 below). None of the ICCs(2,2) of the 13 Youth items fell within the "excellent" range, one item fell within the "good" range, three fell within the "fair" range, and five items (Elicit Goals for Therapy [*n* = 150, 95% CI (-.30—.32), ICC(2,2) = .06], Discuss Pros/Cons [*n* = 150, 95% CI (-.40—.27), ICC(2,2) = -.01], Mutual Agreement [*n* = 150, 95% CI (-.10—.43), ICC(2,2) = .20], Check/Clarify Understanding [*n* = 150, 95% CI (-.03—.46),

ICC(2,2) = .25], and Arrange Follow-Up [*n* = 150, 95% CI (-.41—.26), ICC(2,2) = -.02] fell within the "poor" range (Cicchetti, 1994). For four items, ICCs could not be calculated because there was no variance in the data (Decision Identification, Define Roles, Equipoise, Present Evidence). Items were excluded from analyses if the ICC(2,2) was below .40 and the 95% confidence interval included zero (Cicchetti, 1994; Rousson et al., 2003; Tan & Tan, 2010).

Caregiver Items

Similar to the SDM-YMH Youth subscale, SDM-YMH Caregiver items that demonstrated "poor" (Cicchetti, 1994) inter-rater reliability were considered for exclusion. Per hypothesis 1, it was expected that SDM-YMH Caregiver items would demonstrate "good" to "excellent" inter-rater reliability (Cicchetti, 1994; Gärtner et al., 2018). SDM-YMH Caregiver item inter-rater reliability ranged from ICC(2,2) = -.02 to .73 (*M* = .36, *SD* = .26; see Table 10). None of the ICCs(2,2) of the 13 items fell within the "excellent" range, three items fell within the "good" range, three fell within the "fair" range, and six items (Decision Identification [*n* = 150, 95% CI (-.01 — .47), ICC(2,2) = .28], Define Roles [*n* = 150, 95% CI (-.28 — .33), ICC(2,2) = .07], Present Evidence [*n* = 150, 95% CI (.11 — .53), ICC(2,2) = .36], Discuss Pros/Cons [*n* = 150, 95% CI (-.41 — .26), ICC(2,2) = -.02], Make/Explicitly Defer Decision [*n* = 150, 95% CI (-.70 — .44), ICC(2,2) = .23], and Arrange Follow-Up [*n* = 150, 95% CI (-.39 — .27), ICC(2,2) = -.01] fell within the "poor" range (Cicchetti, 1994). The ICC could not be calculated for one of the items (Equipoise) due to no variance in the data. Items were excluded from analyses if the ICC(2,2) was below .40 and the 95% confidence interval included zero (Cicchetti, 1994; Rousson et al., 2003; Tan & Tan, 2010).

Table 9

Shared Decision Making Practices in Youth Mental Health Scale (Youth) Item and Subscale Descriptive Data and Inter-rater Reliability

Item	N	Range	Min.	Max.	Mean(*SD*)	Skew.	Kurt.	ICC(2,2)
1. Elicit Goals for Therapy	150	2.00	1.00	3.00	1.06(0.25)	5.30	33.01	.06
2. Define/Explain Problem**	150	5.50	1.00	6.50	1.72(1.17)	2.04	3.98	.61
3. Decision Identification	150	0.50	1.00	1.50	1.00(0.06)	8.57	72.45	NV
4. Define Roles	150	1.00	1.00	2.00	1.01(0.09)	9.81	100.50	NV
5. Present Options**	150	4.00	1.00	5.00	1.80(0.88)	1.46	2.05	.49
6. Equipoise	150	0.00	1.00	1.00	1.00(0.00)	--	--	NV
7. Present Evidence	150	2.00	1.00	3.00	1.02(0.17)	11.34	132.95	NV
8. Patient Values/Preferences**	150	5.50	1.00	6.50	2.08(1.15)	1.25	1.25	.56
9. Discuss Pros/Cons	150	1.00	1.00	2.00	1.01(0.09)	9.81	100.50	-.01
10. Mutual Agreement	150	1.50	1.00	2.50	1.06(0.23)	4.41	21.21	.20
11. Make/Explicitly Defer Decision*	150	3.00	1.00	4.00	1.10(0.34)	5.21	35.66	.53
12. Check/Clarify Understanding	150	3.00	1.00	4.00	1.33(0.67)	2.25	4.57	.25
13. Arrange Follow Up	150	0.50	1.00	1.50	1.02(0.11)	4.34	17.08	-.02
SDM-YMH Youth Total Subscale	150	3.38	1.00	4.38	1.68(0.63)	1.21	1.60	.64
SDM-YMH Youth Short Subscale	150	3.50	1.00	4.50	1.87(0.79)	1.02	0.54	.62

Note: NV = No Variance; SDM-YMH = Shared Decision Making Practices in Youth Mental Health Scale; Min = Minimum; Max = Maximum; SD = Standard Deviation; Skew = Skewness; Kurt = Kurtosis; ICC = Intraclass Correlation Coefficient.

*Included in Total subscale **Included in Short and Total subscales

Table 10

Shared Decision Making Practices in Youth Mental Health Scale (Caregiver) Item and Subscale Descriptive Data and Inter-rater Reliability

Item	N	Range	Min.	Max.	Mean(*SD*)	Skew.	Kurt.	ICC(2,2)
1. Elicit Goals for Therapy*	150	3.50	1.00	4.50	1.12(0.45)	4.94	28.20	.43
2. Define/Explain Problem**	150	6.00	1.00	7.00	1.56(1.17)	2.69	7.59	.67
3. Decision Identification	150	1.00	1.00	2.00	1.04(0.14)	4.14	18.24	.28
4. Define Roles	150	2.00	1.00	3.00	1.05(0.24)	5.86	37.65	.07
5. Present Options**	150	3.50	1.00	4.50	1.37(0.77)	2.53	5.95	.73
6. Equipoise	150	0.00	1.00	1.00	1.00(0.00)	--	--	NV
7. Present Evidence*	150	1.50	1.00	2.50	1.05(0.22)	4.63	21.40	.36
8. Patient Values/Preferences**	150	4.00	1.00	5.00	1.33(0.69)	2.73	8.07	.53
9. Discuss Pros/Cons	150	2.00	1.00	3.00	1.03(0.22)	7.33	56.09	-.02
10. Mutual Agreement*	150	1.50	1.00	2.50	1.03(0.18)	6.50	44.56	.62
11. Make/Explicitly Defer Decision	150	2.50	1.00	3.50	1.08(0.30)	4.96	31.25	.23
12. Check/Clarify Understanding*	150	3.00	1.00	4.00	1.19(0.49)	3.44	13.53	.48
13. Arrange Follow Up	150	2.00	1.00	3.00	1.03(0.18)	9.26	96.48	-.01
SDM-YMH Caregiver Total Subscale	150	2.14	1.00	3.14	1.23(.41)	2.12	5.10	.78
SDM-YMH Caregiver Short Subscale	150	3.67	1.00	4.67	1.42(0.74)	2.25	5.35	.81

Note. NV = No Variance; SDM-YMH = Shared Decision Making Practices in Youth Mental Health Scale; Min = Minimum; Max = Maximum; SD = Standard Deviation; Skew = Skewness; Kurt = Kurtosis; ICC = Intraclass Correlation Coefficient.

*Included in Total subscale **Included in Short and Total subscales

Descriptives

The mean, standard deviation, range, skewness, and kurtosis were examined for SDM-YMH items and subscale scores (Youth, Caregiver; see Tables 8 and 9). Items with a range of less than three were deemed problematic, due to restricted range (e.g., McLeod et al., 2018). Seven SDM-YMH Youth item scores had a range less than three: Elicit Goals for Therapy (range = 2.00), Decision Identification (range = 0.50), Define Roles (range = 1.00), Equipoise (range = 0.00), Present Evidence (range = 2.00), Discuss Pros and Cons (range = 1.00), Mutual Agreement (range = 1.50), Arrange Follow-Up (range = 0.50). Eight SDM-YMH Caregiver item scores also had a range less than three: Decision Identification (range = 1.00), Define Roles (range = 2.00), Equipoise (range = 0.00), Present Evidence (range = 1.50), Discuss Pros/Cons (range = 2.00), Mutual Agreement (range = 1.50), Make or Explicitly Defer Decision (range = 2.50), Arrange Follow-Up (range = 2.00).

Skewness and kurtosis values were considered problematic if they fell outside the range of -2 to 2 (George & Mallery, 2016). The skewness values of Youth item scores ranged from 1.25 to 11.34; whereas the skewness values for Caregiver item scores ranged from 2.53 to 9.26. Similarly, the kurtosis values of Youth item scores ranged from 1.25 to 132.95, and the kurtosis values for Caregiver item scores ranged from 5.95 to 96.48 (see Tables 8 and 9 below). The skewed and kurtotic nature of these data are likely due to the low base rate of SDM practices observed in the study. That is, Child STEPs Multisite Trial clinicians were not explicitly instructed to implement SDM practices, which resulted in fewer observable SDM practices in these tapes (i.e., low base rate). The low base rate contributed to the restricted range, which also increased the non-normality of the data as well as made it more difficult to establish inter-rater reliability (see *Limitations* for further detail).

In addition to the SDM-YMH items, several items were included to gauge the potential for SDM practices occurring in session. The reason being that if there was not an opportunity to implement SDM practices, it would be unlikely for clinicians to do so. Thus, coders answered the following items for each session they coded: (a) Was there an opportunity for the clinician to utilize SDM during this session (yes, no)?, (b) To what extent were SDM conversations initiated by the client?, (c) When the client attempted to initiate SDM, how responsive was the clinician?, (d) To what extent were SDM conversations initiated by the caregiver?, (e) When the caregiver attempted to initiate SDM, how responsive was the clinician? Items 2-5 were rated on a scale of 1 (*not at all*) to 7 (*extremely*). However, these data cannot be interpreted due to poor inter-rater reliability (see Table 8 below).

Inter-Item Correlations

Youth Items

After averaging items across coders 1 and 2 to reduce measurement error by removing differences between coders (Lambert & Hill, 1994; McLeod et al., 2018; Southam-Gerow et al., 2016), SDM-YMH Youth items were correlated with one another to assess for statistical redundancy (*r*'s > .7; Kline, 1979; see Table 11). Only two items were statistically redundant (Present Options and Patient Values/Preferences; $r = .77$; *r*'s > .7, Kline, 1979). This redundancy is likely a product of the two SDM practices often co-occurring. That is, clinicians in this sample frequently presented options to youth clients and immediately followed up with a question regarding their preference (e.g., "Here are options A and B, which would you prefer?"). While these two practices often co-occurred in the current sample, they are thought to be theoretically distinct. Therefore, they were both retained as separate items in analyses.

Table 11

Inter-Item Correlations: SDM-YMH Youth Subscale

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Elicit Goals for Therapy	1	.19*	-.03	.20*	.21**	--	.50**	.29**	.12	.31**	.49**	.13	.14	.31**
2. Define/ Explain Problem		1	-.02	.18*	.08	--	.27**	.17*	.18*	.06	.08	.50**	.08	.61**
3. Decision Identi-fication			1	-.01	.13	--	-.01	.14	-.01	-.03	.22**	.12	-.03	.11
4. Define Roles				1	.19*	--	.43**	.22**	-.01	-.03	.13	.03	.32**	.27**
5. Present Options					1	--	.00	.77**	-.08	.32**	.47**	.02	.05	.78**
6. Equipoise						1	--	--	--	--	--	--	--	--
7. Present Evidence							1	.13	.21*	-.03	.09	.22**	-.02	.20*
8. Patient Values/Pref-erences								1	-.02	.22**	.43**	.15	.12	.85**
9. Discuss Pros/Cons									1	-.03	-.03	.28**	-.02	.05
10. Mutual Agreement										1	.50**	-.05	.00	.25**
11. Make/ Explicitly Defer Decision											1	.01	.07	.42**

12. Check/ Clarify Under-standing	1	.06	.32**
13. Arrange Follow Up		1	.12
14. SDM-YMH Youth			1

Note. N = 150, SDM -YMH Youth = Shared Decision Making Practices in Youth Mental Health (SDM-YMH) Youth Subscale
*Correlation is significant at the 0.05 level (2-tailed);
** Correlation is significant at the 0.01 level (2-tailed)

Caregiver Items

Again, SDM-YMH Caregiver items were averaged across coders 1 and 2. Then, SDM-YMH items were correlated with one another to examine statistical redundancy (see Table 12). All items were statistically distinct (r's < .7; Kline, 1979).

Construct Validity: SDM-YMH Item Scores

Construct validity is the degree to which scores on a measure represent the construct they are purported to assess. Construct validity cannot be determined by only one correlation between scores on two instruments. Instead, construct validity is established by examining patterns of associations with scores on similar instruments, which involves assessing different types of validity (Kazdin, 2016). This study considered convergent and discriminant validity as key psychometric properties that could support the construct validity of SDM-YMH scores. Convergent validity refers to the degree to which scores on instruments that measure similar constructs are associated with one another. Conversely, discriminant validity refers to the extent to which instruments designed to measure distinct constructs produce scores that are not strongly associated with each other (Kazdin, 2016).

Per hypothesis 2, it was expected that SDM-YMH items and subscales would demonstrate greater shared variance with theoretically similar constructs (convergent) and less shared variance with theoretically dissimilar constructs (discriminant; Campbell & Fiske, 1959; Foster & Cone, 1995; Kazdin, 2016). In these analyses, the nested structure does not inflate Type I error, as the analyses do not rely on standard analytic models (e.g., general linear model or GLM; Hedeker & Gibbons, 2006; Zucker, 1990). Following standard practice (Carroll, 2000; Hogue et al., 2008; McLeod et al., 2015; McLeod et al., 2018; Southam-Gerow et al., 2016), analyses were conducted at the session level (N = 150).

Table 12

Inter-Item Correlations: SDM-YMH Caregiver Subscale

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Elicit Goals for Therapy	1	.21**	.35**	.18*	.46**	--	.18*	.42**	.06	.29**	.19*	.29**	.13	.40**
2. Define/Explain Problem		1	.29**	.32**	.51**	--	.34**	.49**	.06	.19*	.40**	.43**	.03	.86**
3. Decision Identification			1	.19*	.48**	--	.26**	.47**	.17*	.09	.40**	.26**	.22**	.47**
4. Define Roles				1	.23**	--	.48**	.41**	-.03	.36**	.54**	.00	.27**	.38**
5. Present Options					1	--	.09	.67**	.22**	.29**	.59**	.41**	.11	.83**
6. Equipoise						1	--	--	--	--	--	--	--	--
7. Present Evidence							1	.41**	-.03	.13	.37**	.16	-.03	.34**
8. Patient Values/Preferences								1	.21**	.33**	.60**	.51**	.04	.81**
9. Discuss Pros/Cons									1	-.03	.21**	.58**	.02	.17*
10. Mutual Agreement										1	.30**	.03	-.03	.30**
11. Make/ Explicitly Defer Decision											1	.29**	.27**	.59**
12. Check/Clarify Understanding												1	.00	.53**
13. Arrange Follow Up													1	.07

14. SDM-YMH Caregiver	

Note. N = 150, SDM -YMH Caregiver = Shared Decision Making Practices in Youth Mental Health (SDM-YMH) Caregiver Subscale;
*Correlation is significant at the 0.05 level (2-tailed);
** Correlation is significant at the 0.01 level (2-tailed)

First, Pearson correlations were used to assess the convergent validity of SDM-YMH items scores. Although some of the data were non-normal, Pearson correlations were selected for these analyses to make comparisons to similar studies more feasible, as many other studies (e.g., Elwyn, 2005; Simon, 2006) utilized Pearson correlations (as opposed to Spearman). It was expected that SDM-YMH item scores would evidence the strongest associations with TPOCS-RS Goal Setting and Agenda Setting item scores ("large" correlations: $r \geq .36$; Rosenthal & Rosnow, 1984) but not at a level that is redundant (r's $\leq .7$; Kline, 1979).

Second, the SDM-YMH item scores' discriminant validity was assessed by correlating scores on the SDM-YMH and scores on treatment integrity and process measures. It was expected that SDM-YMH items scores would evidence "medium" correlations ($.24 \leq r < .36$; Rosenthal & Rosnow, 1984) with TPOCS-A Youth scores, and "small" correlations ($.10 \leq r \leq .23$; Rosenthal & Rosnow, 1984) with scores on the: (1) CBAY-A Model, (2) CBAY-C Model, and (3) TPOCS-RS Psychodynamic Item and Family Subscale. It was also hypothesized that the strength of the correlations for the discriminant analyses would be significantly less strong (closer to 0) than analyses for convergent validity. Fisher's r-to-z transformation was used to determine statistical significance (Silver & Dunlap, 1987).

Youth Item Scores

Correlations were examined between SDM-YMH Youth items with at least "fair" (Cicchetti, 1994) inter-rater reliability (i.e., those being considered for inclusion in analyses) and item/subscale scores of theoretically similar (e.g., TPOCS-RS Goal Setting) and dissimilar constructs (e.g., TPOCS-RS Psychodynamic Item). See Table 13 for detail.

Table 13

Construct Validity: Shared Decision Making Practices in Youth Mental Health Scale: Youth Items

SDM-YMH Item	TPOCS-RS Goal Setting	TPOCS-RS Session Agenda	TPOCS-A Youth	CBAY-A	CBAY-C	TPOCS-RS Family	TPOCS-RS Psycho-dynamic
Define/Explain Problem	.13	.02	-.17	.002	.03	-.08	.22*
N	150	150	129	150	122	150	96
Present Options	-.13	.16*	.002	.14	.18	-.22*	.11
N	150	150	129	150	122	150	96
Patient Values/Preferences	-.06	.27**	-.04	.08	.14	-.19*	.08
N	150	150	129	150	122	150	96
Make/Explicitly Defer Decision *N*	.13 150	.16 150	-.003 129	.01 150	.17 122	-.06 150	.07 96

Note. SDM-YMH = SDM-YMH= Shared Decision Making Practices in Youth Mental Health, TPOCS-RS Goal Setting = Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale (TPOCS-RS) Goal Setting Item, TPOCS-RS Session Agenda = TPOCS-RS Session Agenda Item, CBAY-A = Cognitive Behavioral Therapy Adherence Scale for Youth Anxiety Model Subscale, CBAY-C = CBT for Anxiety in Youth Competence Scale Model Subscale, TPOCS-A Youth = Therapy Process Observational Coding System for Child Psychotherapy—Alliance, Youth Subscale TPOCS-RS Family = TPOCS-RS Family Subscale, Psychodynamic = TPOCS-RS Psychodynamic Item.

*Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

Define/Explain Problem. The correlations between the Define/Explain Problem item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), youth alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from .-.17 to .22 and were "small" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Define/Explain Problem scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .08 (*SD* = .08) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Define/Explain Problem item and the TPOCS-A was .17 and small in magnitude. The absolute value of the correlations between the Define/Explain Problem item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .08 (*SD* = .10) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Define/Explain Problem and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant (z = -0.75, p = .45) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 0.00, p = 1.00). Moreover, contrasts revealed that the correlation between the Define/Explain Problem item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 0.67, p = .50). Overall, this pattern does not support the construct validity of Define/Explain Problem (Youth) scores.

Present Options. The correlations between the Present Options item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), youth alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS

Psychodynamic) ranged from -.22 to .18 and were "small" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Present Options scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .15 (*SD* = .02) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Present Options item and the TPOCS-A was .002 and small in magnitude. The absolute value of the correlations between the Present Options item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .16 (*SD* = .05) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Present Options and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 1.23$, $p = .22$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.08$, $p = .94$). Moreover, contrasts revealed that the correlation between the Present Options item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 1.17$, $p = .24$). Overall, this pattern does not support the construct validity of Present Options (Youth) scores.

Patient Values/Preferences. The correlations between the Patient Values/Preferences item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), youth alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.19 to .27 and were "small" to "medium" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Patient Values/Preferences scores and scores on TPOCS-RS Goal Setting

and Session Agenda items was .17 (*SD* = .15) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Patient Values/Preferences item and the TPOCS-A was .04 and small in magnitude. The absolute value of the correlations between the Patient Values/Preferences item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .12 (*SD* = .05) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Patient Values/Preferences and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 1.08$, $p = .28$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.39$, $p = .70$). Moreover, contrasts revealed that the correlation between the Patient Values/Preferences item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.59$, $p = .56$). Overall, this pattern does not support the construct validity of Patient Values/Preferences (Youth) scores.

Make/Explicitly Defer Decision. The correlations between the Make/Explicitly Defer Decision item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), youth alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.003 to .17 and were "small" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Make/Explicitly Defer Decision scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .15 (*SD* = .02) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Make/Explicitly Defer Decision item and the TPOCS-A was .003 and small in magnitude. The absolute value of the correlations between the

Make/Explicitly Defer Decision item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .08 (*SD* = .07) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Make/Explicitly Defer Decision and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant (z = 1.22, p = .22) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 0.54, p = .59). Moreover, contrasts revealed that the correlation between the Make/Explicitly Defer Decision item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 0.56, p = .58). Overall, this pattern does not support the construct validity of Make/Explicitly Defer Decision (Youth) scores.

Caregiver Item Scores

Similarly, correlations were examined between SDM-YMH Caregiver item scores with at least "fair" (Cicchetti, 1994) inter-rater reliability (i.e., those being considered for inclusion in analyses) and item/subscale scores of theoretically similar (e.g., TPOCS-RS Goal Setting) and dissimilar constructs (e.g., TPOCS-RS Psychodynamic). See Table 14 for detail.

Elicit Goals for Therapy. The correlations between the Elicit Goals for Therapy item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.20 to .52 and were "small" to "large" in magnitude (Rosenthal & Rosnow, 1984). The mean of the absolute value of the inter-item correlations between Elicit Goals for Therapy scores and scores on TPOCS-RS Goal Setting and Session Agenda items was

Table 14

Construct Validity: Shared Decision Making Practices in Youth Mental Health Scale: Caregiver Items

SDM-YMH Item	TPOCS-RS Goal Setting	TPOCS-RS Session Agenda	TPOCS-A Caregiver	CBAY-A	CBAY-C	TPOCS-RS Family	TPOCS-RS Psychodynamic
Elicit Goals for Therapy	.09	.11	.23	-.20	-.06	.52**	.06
N	150	150	65	150	122	150	96
Define/Explain Problem	.54**	-.08	.02	-.14	-.13	.54**	.01
N	150	150	65	150	122	150	96
Present Options	.32**	.07	.22	-.27**	.01	.75**	-.08
N	150	150	65	150	122	150	96
Present Evidence	.24**	.02	-.08	-.04	.06	.20**	.08
N	150	150	65	150	122	150	96
Patient Values/Preferences	.43**	.16*	.10	-.26**	-.01	.73**	-.06
N	150	150	65	150	122	150	96
Mutual Agreement	.19*	.01	.07	-.16	-.001	.30**	-.05
N	150	150	65	150	122	150	96
Check/Clarify Understanding	.31**	.12	-.06	-.23**	-.17	.44**	-.06
N	150	150	65	150	122	150	96

Note. SDM-YMH= Shared Decision Making Practices in Youth Mental Health, TPOCS-RS Goal Setting = Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale (TPOCS-RS) Goal Setting Item, TPOCS-RS Session Agenda = TPOCS-RS Session Agenda Item, CBAY-A = Cognitive Behavioral Therapy Adherence Scale for Youth Anxiety Model Subscale, CBAY-C = CBT for Anxiety in Youth Competence Scale Model Subscale, TPOCS-A Caregiver = Therapy Process Observational Coding System for Child Psychotherapy—Alliance, Caregiver Subscale TPOCS-RS Family = TPOCS-RS Family Subscale, Psychodynamic = TPOCS-RS Psychodynamic Item.

*Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

.10 (*SD* = .01) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Elicit Goals for Therapy item and the TPOCS-A was .23 and small in magnitude. The absolute value of the correlations between the Elicit Goals for Therapy item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .21 (*SD* = .19) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Elicit Goals for Therapy and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 0.88$, $p = .38$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.85$, $p = .40$). Moreover, contrasts revealed that the correlation between the Elicit Goals for Therapy item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.13$, $p = .90$). Overall, this pattern does not support the construct validity of Elicit Goals for Therapy (Caregiver) scores.

Define/Explain Problem. The correlations between the Define/Explain Problem item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.14 to .54 and were "small" to "large" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Define/Explain Problem scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .31 (*SD* = .33) and medium (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Define/Explain Problem item and the TPOCS-A was .02 and small in magnitude. The absolute value of the correlations between the Define/Explain Problem

item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .21 (*SD* = .23) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Define/Explain Problem and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was significant ($z = 1.98$, $p < .05$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.81$, $p = .42$). Moreover, contrasts revealed that the correlation between the Define/Explain Problem item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 1.18$, $p = .24$). Overall, this pattern does not support the construct validity of Define/Explain Problem (Caregiver) scores.

Present Options. The correlations between the Present Options item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.27 to .75 and were "small" to "large" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Present Options scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .20 (*SD* = .18) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Present Options item and the TPOCS-A was .22 and small in magnitude. The absolute value of the correlations between the Present Options item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .28 (*SD* = .33) and medium in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Present Options and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 0.14$, $p = .89$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.64$, $p = .52$). Moreover, contrasts revealed that the correlation between the Present Options item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.39$, $p = .70$). Overall, this pattern does not support the construct validity of Present Options (Caregiver) scores.

Present Evidence. The correlations between the Present Evidence item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.04 to .24 and were "small" to "medium" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Present Evidence scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .11 (*SD* = .18) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Present Evidence item and the TPOCS-A was .08 and small in magnitude. The absolute value of the correlations between the Present Evidence item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .10 (*SD* = .07) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Present Evidence and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 0.20$, $p = .84$)

and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.08$, $p = .93$). Moreover, contrasts revealed that the correlation between the Present Evidence item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.12$, $p = .90$). Overall, this pattern does not support the construct validity of Present Evidence (Caregiver) scores.

Patient Values/Preferences. The correlations between the Patient Values/Preferences item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.26 to .73 and were "small" to "large" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Patient Values/Preferences scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .30 ($SD = .19$) and medium (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Patient Values/Preferences item and the TPOCS-A was .10 and small in magnitude. The absolute value of the correlations between the Patient Values/Preferences item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .27 ($SD = .33$) and medium in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Patient Values/Preferences and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 1.38$, $p = .17$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.25$, $p = .80$). Moreover, contrasts revealed that the correlation between the Patient Values/Preferences item and the TPOCS-A and mean of the

CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 1.08, p = .28). Overall, this pattern does not support the construct validity of Patient Values/Preferences (Caregiver) scores.

Mutual Agreement. The correlations between the Mutual Agreement item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.16 to .30 and were "small" to "medium" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Mutual Agreement scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .10 (*SD* = .13) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Mutual Agreement item and the TPOCS-A was .07 and small in magnitude. The absolute value of the correlations between the Mutual Agreement item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .13 (*SD* = .13) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Mutual Agreement and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant (z = 0.20, p = .84) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 0.23, p = .81). Moreover, contrasts revealed that the correlation between the Mutual Agreement item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant (z = 0.37, p = .71). Overall, this pattern does not support the construct validity of Mutual Agreement (Caregiver) scores.

Check/Clarify Understanding. The correlations between the Check/Clarify Understanding item and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.23 to .44 and were "small" to "large" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between Check/Clarify Understanding scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .22 (*SD* = .13) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the Check/Clarify Understanding item and the TPOCS-A was .06 and small in magnitude. The absolute value of the correlations between the Check/Clarify Understanding item and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .23 (*SD* = .16) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the Check/Clarify Understanding and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 1.08$, $p = .28$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.08$, $p = .94$). Moreover, contrasts revealed that the correlation between the Check/Clarify Understanding item and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 1.06$, $p = .29$). Overall, this pattern does not support the construct validity of Check/Clarify Understanding (Caregiver) scores.

Subscale Generation

Several factors contributed to the generation of SDM-YMH subscales. First, items that demonstrated "poor" inter-rater reliability (ICC[2,2]s < .4; Cicchetti, 1994) and had 95% confidence intervals that included zero (Rousson et al., 2003; Tan & Tan, 201) were excluded from analyses. Then, inter-item correlations were then examined for statistical redundancy (*r*'s > .7 were considered redundant; Kline, 1979). Although two Youth item scores (Present Options, Patient Values/Preferences) were redundant, they were both retained in analyses due to being conceptually distinct and integral to SDM (see *Inter-Item Correlations* section above). Next, construct validity was examined for each item with at least "fair" inter-rater reliability (ICC[2,2]s > .4; Cicchetti, 1994). However, construct validity was poor for the majority of SDM-YMH Youth- and Caregiver-focused items. Thus, this information was not prioritized with regards to decisions about including or excluding items from the subscales.

As shown in Tables 8 and 9, items that demonstrated at least "fair" inter-rater reliability (ICC[2,2]s > .4; Cicchetti, 1994) varied across Youth and Caregiver subscales. Therefore, two sets of subscales were generated. First, the SDM-YMH Youth/Caregiver Total subscales were generated to include all items that demonstrated at least "fair" inter rater reliability (ICC[2,2]s > .4; Cicchetti, 1994). That is, the *SDM-YMH Youth Total* subscale included the following items: Define/Explain Problem, Present Options, Patient Values/Preferences, and Make/Explicitly Defer Decision. The *SDM-YMH Caregiver Total* subscale included the following items: Elicit Goals for Therapy, Define/Explain Problem, Present Options, Present Evidence, Patient Values/Preferences, Mutual Understanding, and Check/Clarify Understanding. Then, two more subscales were generated to include identical items across both Youth and Caregiver subscales, as SDM practices were originally hypothesized to be the same when directed at youth and their

caregivers. Both *SDM-YMH Youth Short* and *SDM-YMH Caregiver Short* subscales included the following items: Define/Explain Problem, Present Options, and Patient Values/Preferences, as these items demonstrated at least fair inter-rater reliability (ICC[2,2]s > .4; Cicchetti, 1994) when directed at youth and caregivers (see Tables 8 and 9).

Item scores within each subscale were then averaged to generate subscale scores. An average score was expected to best capture SDM practice delivery, as SDM practice elements are not intended to be used in isolation. For example, clinicians utilizing SDM techniques would likely not focus solely on discussing the pros and cons of a treatment option in one session and exclusively on exploring family values in the next. Instead, it is more probable that clinicians would utilize multiple SDM practices within a given session to engage clients and caregivers in the SDM process throughout the course of psychotherapy. The Youth Total subscale scores (*M* = 1.68, *SD* = 0.63) had a range of 3.38, and the Caregiver Total subscale scores (*M* =1.23, *SD* = 0.41) had a range of 2.14. The Youth Short subscale scores (*M* = 1.87, *SD* = 0.79) had a range of 3.50, and the Caregiver Short subscale scores (*M* =1.42, *SD* = 0.74) had a range of 3.67. Additionally, skewness and kurtosis values were less problematic for subscale scores as compared to item scores. The Youth Total subscale (Skewness = 1.21; Kurtosis = 1.60) and Youth Short subscale (Skewness = 1.02; Kurtosis = 0.54) scores were more normal than Caregiver Total (Skewness = 2.12; Kurtosis = 5.10) and Short subscale scores (Skewness = 2.25; Kurtosis = 5.35). Inter-rater reliability for Youth Short subscale scores (items: Define/Explain Problem, Present Options, Patient Values/Preferences) was in the "good" range (ICC[2,2] = .62), and inter-rater reliability for Youth Total subscale scores was also in the "good" range (ICC[2,2] = .64), See Table 9 for descriptive data for SDM-YMH Youth scores. Inter-rater reliability for Caregiver Short subscale scores (items: Define/Explain Problem, Present Options, Patient

Values/Preferences) was in the "excellent" range (ICC[2,2] = .81), and inter-rater reliability for Caregiver Total subscale scores was also in the "excellent" range (ICC[2,2] = .78). See Table 10 for descriptive data of SDM-YMH Caregiver scores.

Additionally, SDM-YMH Youth and Caregiver subscale scores were compared across the Modular and Standard conditions to test for study condition effects on SDM-YMH scores. There was no significant effect of study condition on SDM-YMH Youth Total subscale scores, $t(148) = 1.90$, $p = .06$, despite clinicians administering slightly more SDM practices to youth in the Modular condition ($M = 1.79$, $SD = 0.65$) as compared to those in the Standard condition ($M = 1.59$, $SD = 0.61$). There was also no significant effect of study condition on SDM-YMH Youth Short subscale scores, $t(148) = 1.82$, $p = .07$, despite clinicians administering slightly more SDM practices to youth in the Modular condition ($M = 2.00$, $SD = 0.79$) as compared to those in the Standard condition ($M = 1.76$, $SD = 0.78$). Similarly, there was no significant effect of study condition on SDM-YMH Caregiver Total subscale scores, $t(148) = 1.21$, $p = .23$, despite clinicians administering slightly more SDM practices to caregivers in the Modular condition ($M = 1.28$, $SD = 0.41$) as compared to those in the Standard condition ($M = 1.20$, $SD = 0.41$). There was also no significant effect of study condition on SDM-YMH Caregiver Short subscale scores, $t(148) = 1.43$, $p = .15$, despite clinicians administering slightly more SDM practices to caregivers in the Modular condition ($M = 1.52$, $SD = 0.74$) as compared to those in the Standard condition ($M = 1.34$, $SD = 0.73$). Therefore, analyses proceeded with the combined sample of Modular and Standard conditions. This approach is also similar to previous studies that have used the same sample (e.g., Cecilione et al., 2021).

Construct Validity: SDM-YMH Subscale Scores

The same procedures used to assess the construct validity of the SDM-YMH item scores were also used to examine the construct validity of SDM-YMH subscale scores. That is, Pearson correlations were used to assess convergent validity of SDM-YMH subscale scores. It was expected that SDM-YMH subscale scores would evidence the strongest associations with TPOCS-RS Goal Setting and Session Agenda item scores ("large" correlations: $r > .36$; Rosenthal & Rosnow, 1984) but not at a level that is redundant (r's $\leq .7$; Kline, 1979). Then, discriminant validity of the SDM-YMH subscale scores was assessed by correlating scores on the SDM-YMH and treatment integrity and process measures. It was expected that SDM-YMH subscale scores would evidence "medium" correlations ($.24 \leq r < .36$; Rosenthal & Rosnow, 1984) with TPOCS-A scores, and "small" correlations ($.10 \leq r \leq .23$; Rosenthal & Rosnow, 1984) with scores on the: (a) CBAY-A Model, (b) CBAY-C Model, and (c) the TPOCS-RS Psychodynamic and Family subscales. It was also hypothesized that the strength of the correlations for the discriminant analyses would be significantly less strong (closer to 0) than analyses for convergent validity. Fisher's r-to-z transformation was used to determine statistical significance (Silver & Dunlap, 1987).

SDM-YMH Youth Total Subscale Scores

The correlations between the SDM-YMH Youth Total subscale and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.21 to .21 and were "small" in magnitude (Rosenthal & Rosnow, 1984). The absolute value of the mean inter-item of the correlations between SDM-YMH Youth Total subscale scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .11

(*SD* = .14) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the SDM-YMH Youth Total subscale and the TPOCS-A was .10 and small in magnitude. The absolute value of the correlations between the SDM-YMH Youth Total subscale and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .16 (*SD* = .05) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the SDM-YMH Youth Total subscale and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 0.08$, $p = .94$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.38$, $p = .70$). Moreover, contrasts revealed that the correlation between the SDM-YMH Youth Total subscale and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.45$, $p = .65$). Overall, this pattern does not support the construct validity of SDM-YMH Youth Total subscale scores. The restricted ranges of these subscale scores as well as poorer than anticipated inter-rater reliability made establishing construct validity especially difficult for the Caregiver Total scores (e.g., Foster & Cone, 1995).

SDM-YMH Youth Short Subscale Scores

The correlations between the SDM-YMH Youth Short subscale and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.21 to .20 and were "small" in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between SDM-YMH Youth Short subscale scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .11

Table 15

Construct Validity: Shared Decision Making Practices in Youth Mental Health Scale Subscale Scores

	1	2	3	4	5	6	7	8	9	10	11	12
1. SDM Youth (Total)	1	.99**	-.25**	-.25**	.01	.21*	.10	.16	-.10	-.26*	.17	-.21*
95% CI		[.99,.99]	[-.39,-.09]	[-.40,-.10]	[-.15,.17]	[.05,.36]	[-.07,.24]	[-.02,.33]	[-.27,.07]	[-.47,-.02]	[-.03,.36]	[-.35,-.05]
N	150	150	150	150	150	150	150	122	129	65	96	150
2. SDM Youth (Short)		1	-.25**	-.26**	-.01	.20*	.09	.15	-.11	-.26*	.17	-.21**
95% CI			[-.40,-.10]	[-.40,-.10]	[-.15,.17]	[.04,.35]	[-.07,.25]	[-.03,.32]	[-.28,.06]	[-.47,-.01]	[-.03,.36]	[-.36,-.05]
N		150	150	150	150	150	150	122	129	65	96	150
3. SDM Caregiver (Total)			1	.97**	.51**	.07	-.27**	-.09	-.23**	.13	-.02	.79**
95% CI				[.96,.98]	[.38,.62]	[-.10,.22]	[-.42,-.12]	[-.26,.09]	[-.39,-.06]	[-.12,.36]	[-.22,.18]	[.72,.84]
N			150	150	150	150	150	122	129	65	96	150
4. SDM Caregiver (Short)				1	.53**	.03	-.25**	-.07	-.25**	.13	-.03	.78**
95% CI					[.41,.64]	[-.13,.19]	[-.39,-.09]	[-.25,.11]	[-.40,-.08]	[-.12,.36]	[-.23,.17]	[.71,.83]
N				150	150	150	150	122	129	65	96	150
5. TPOCS-RS Goal Setting					1	.16*	-.12	-.004	-.17	-.13	-.08	.31**
95% CI						[.00,.32]	[-.28,.04]	[-.18,.17]	[-.33,.01]	[-.36,.12]	[-.28,.12]	[.16,.45]
N					150	150	150	122	129	65	96	150
6. TPOCS-RS Session Agenda						1	.19*	.31**	.04	-.24	-.08	.08
95% CI							[.03,.34]	[.14,.46]	[-.13,.21]	[-.46,.00]	[-.28,.12]	[-.08,.24]
N						150	150	122	129	65	96	150
7. CBAY-A							1	.57**	.26**	-.26*	-.08	-.39**
95% CI								[.44,.68]	[.10,.42]	[-.47,-.02]	[-.28,.12]	[-.51,.24]
N							150	122	129	65	96	150

8. CBAY-C	1	.30**	-.20	-.08	.03
95% CI		[.12,.46]	[-.46,.10]	[-.30,.15]	[-.15,.20]
N	122	110	45	78	122
9. TPOCS-A Youth		1	.23	.02	-.15
95% CI			[-.09,.48]	[-.20,.24]	[-.31,.03]
N		129	45	81	129
10. TPOCS-A Caregiver			1	-.01	.38**
95% CI				[-.31,.29]	[.15,.57]
N			65	43	65
11. TPOCS-RS Psychodynamic				1	-.04
95% CI					[-.24,.16]
N				96	96
12. TPOCS-RS Family					1
95% CI					
N					150

Note. SDM-YMH= Shared Decision Making Practices in Youth Mental Health, TPOCS-RS = Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale; CBAY-A = Cognitive Behavioral Therapy Adherence Scale for Youth Anxiety Model Subscale, CBAY-C = CBT for Anxiety in Youth Competence Scale Model Subscale, TPOCS-A Youth = Therapy Process Observational Coding System for Child Psychotherapy—Alliance, Youth Subscale TPOCS-A Caregiver = Therapy Process Observational Coding System for Child Psychotherapy—Alliance, Caregiver Subscale.

*Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

(*SD* = .14) and small (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the SDM-YMH Youth Short subscale and the TPOCS-A was .11 and small in magnitude. The absolute value of the correlations between the SDM-YMH Youth Short subscale and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .16 (*SD* = .05) and small in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the SDM-YMH Youth Short subscale and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 0.01$, $p = 1.00$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.38$, $p = .70$). Moreover, contrasts revealed that the correlation between the SDM-YMH Youth Short subscale and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.37$, $p = .71$). Overall, this pattern does not support the construct validity of SDM-YMH Youth Short subscale scores. This pattern closely resembles that of Youth Total scores, likely in large part due to the statistically redundant nature of these subscale scores ($r = .99$, $p < .001$).

SDM-YMH Caregiver Total Subscale Scores

The correlations between the SDM-YMH Caregiver Total subscale and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.27 to .79 and were “small” to “large” in magnitude (Rosenthal & Rosnow, 1984). The mean inter-item of the absolute value of the correlation between SDM-YMH Caregiver Total subscale scores and scores on TPOCS-RS Goal Setting and Session

Agenda items was .29 (*SD* = .31) and medium (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the SDM-YMH Caregiver Total subscale and the TPOCS-A was .13 and small in magnitude. The absolute value of the correlations between the SDM-YMH Caregiver Total subscale and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .29 (*SD* = .35) and medium in magnitude (Rosenthal & Rosnow, 1984). Follow up contrasts revealed that mean of the absolute value of the correlations between the SDM-YMH Caregiver Total subscale and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 1.11$, $p = .27$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.00$, $p = 1.00$). Moreover, contrasts revealed that the correlation between the SDM-YMH Caregiver Total subscale and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 1.02$, $p = .31$). Overall, this pattern does not support the construct validity of SDM-YMH Caregiver Total subscale scores. Again, the restricted ranges of these subscale scores and poorer than anticipated inter-rater reliability made it difficult to evaluate the construct validity of the Caregiver Total scores (e.g., Foster & Cone, 1995). Similarly, the high correlation between scores on Caregiver Total and TPOCS-RS Family subscale scores made it difficult to evaluate construct validity (see *Discussion* for further detail).

SDM-YMH Caregiver Short Subscale Scores

The correlations between the SDM-YMH Caregiver Short subscale and the SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting), caregiver alliance (TPOCS-A), competence (CBAY-C), and treatment process (TPOCS-RS Family, TPOCS-RS Psychodynamic) ranged from -.25 to .78 and were “small” to “large” in magnitude (Rosenthal &

Rosnow, 1984). The mean inter-item of the absolute value of the correlation between SDM-YMH Caregiver Short subscale scores and scores on TPOCS-RS Goal Setting and Session Agenda items was .28 (*SD* = .35) and medium (Rosenthal & Rosnow, 1984). The absolute value of the correlation between scores on the SDM-YMH Caregiver Short subscale and the TPOCS-A was .13 and small in magnitude. The absolute value of the correlations between the SDM-YMH Caregiver Short subscale and the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was .28 (*SD* = .35) and medium in magnitude (Rosenthal & Rosnow, 1984).

Follow up contrasts revealed that mean of the absolute value of the correlations between the SDM-YMH Caregiver Short subscale and the item scores most similar to SDM practices (TPOCS-RS Goal Setting, TPOCS-RS Agenda Setting) and (a) the TPOCS-A was not significant ($z = 1.04$, $p = .30$) and (b) the mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.00$, $p = 1.00$). Moreover, contrasts revealed that the correlation between the SDM-YMH Caregiver Short subscale and the TPOCS-A and mean of the CBAY-A, CBAY-C, TPOCS-RS Family, and TPOCS-RS Psychodynamic was not significant ($z = 0.96$, $p = .34$). Overall, this pattern does not support the construct validity of SDM-YMH Caregiver Short subscale scores. This pattern closely resembles that of Caregiver Total scores, likely in large part due to high correlation between the subscale scores ($r = .97$, $p < .001$).

Predictive Validity

An a priori power analysis was conducted using G*Power3 (Faul et al., 2007). To calculate power, the potential effect of nesting must be considered. Thus, the sample size must be adjusted by an inflation factor (IF) of 1+(c-1)r, where c is the average cluster size and r is the ICC (Donner, Birkett, & Buck, 1981). This formula was used to produce adjusted sample size numbers (i.e., reductions in sample size of 38) by dividing the sample size (*N*=38 clients) by the

IF. Power was then calculated to test for effects under two conditions (Donner et al., 1981): observed ICC of .03, (IF=1.045, adjusted n = 36) and observed ICC of .01 (IF=1.015; adjusted n = 37).

Thus, power was calculated for medium and large effects for the current study. For predictive validity analyses, the predictors included SDM-YMH, treatment condition, time, and interactions. At a conservatively estimated magnitude of ICC=.01 there was adequate power to detect medium to large effects (.82 and .93, respectively). At an estimated magnitude of ICC = .03, there was adequate power to detect medium to large effects (.84 and .95, respectively). Results indicated that a total sample of 152 therapy tapes was sufficient to detect medium to large effects. As aforementioned, only 150 tapes were included due to availability of tapes in the parent studies (Child STEPs Multisite Trial; Weisz et al., 2012; TIMS; TIMS; R01 MH1086539; PIs McLeod, Southam-Gerow).

Predictive validity refers to the association of a score on a measure at one point in time with performance on another measure at some point in the future (Kazdin, 2016). Higher SDM-YMH subscale scores (Youth, Caregiver) were hypothesized to predict the slope of change in (3a) youth-clinician alliance; (3b) caregiver-clinician alliance, (3c) youth-reported internalizing symptoms, and (3d) caregiver-reported youth internalizing symptoms. (Of note, SDM-YMH Youth Total and Caregiver Total were utilized in analyses, as they were not statistically distinct from the Youth Short (r = .99; r's > .7; Kline, 1979) and Caregiver Short (r = .97; r's > .7; Kline, 1979) subscales, respectively (see Table 15). The relation between SDM practices, the alliance, and clinical outcomes was assessed using multilevel modeling techniques in SPSS (IBM, Corp., 2020), as the dependent variables (DVs) were longitudinal. For hypotheses 3a and 3b, the DVs were the TPOCS-A (Youth, Caregiver) collected at each session. For hypotheses 3c and 3d, the

DVs were the WPC Internalizing scales (Youth, Caregiver) collected at baseline and weekly during treatment.

Longitudinal multilevel modeling estimates the association between independent (IV) and dependent variables (DV) when there is a correlation between observations as well as examines trends over time. This model was chosen as it accounted for the nested structure of the data (i.e., nesting of clients under therapists—each client only saw one therapist, but one study therapist may have seen more than one client. For each DV (e.g., youth-clinician alliance, youth self-reported internalizing symptoms), session number and therapist number were entered as subjects, and client number was entered as a repeating factor. Repeated covariance type was set to Scaled Identity because this type of structure matrix allows for the modeling of a set of repeated measures that are likely independent and of equal variance. Lastly, predictors were entered simultaneously, including: SDM subscale (either Youth Total or Caregiver Total; see Subscale Generation above for description of score generation), time (days in treatment, calculated by subtracting the date of the first session from the current session date—i.e., time started at 0), and condition (i.e., Modular or Standard).

Youth Subscale Scores

First, overall youth-reported alliance scores did not change significantly over the course of treatment ($\beta = 0.01$, $t = 1.39$, $p = .17$); this was true for both treatment groups ($\beta = 0.16$, $t = 0.27$, $p = .79$). SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) did not significantly affect the rate of change (slope) of youth-clinician alliance (TPOCS-A Youth subscale scores) over the course of treatment ($\beta = -0.002$, $t = -1.47$, $p = .14$). Similarly, SDM practices directed towards youth clients (SDM-YMH Youth subscale scores) did not significantly affect the rate of change of youth-clinician alliance (TPOCS-A Youth subscale

scores) regardless of what study condition youth were assigned (β = -0.16, *t* = -0.51, *p* = .61). Likewise, there was no significant effect of SDM practices directed towards youth clients (SDM-YMH Youth subscale scores) on the rate of change of youth-clinician alliance (TPOCS-A Youth subscale scores) when accounting for both condition and time in treatment (β = -0.003, *t* = 0.80, *p* = .43). In short, all two-way and three-way interactions between condition, days, and SDM were not significant, suggesting that the rate of change did not differ according to either condition or initial SDM-YMH Youth Total score. See Table 16.

Similarly, overall youth-reported internalizing scores did not change significantly over the course of treatment (β = -0.001, *t* = -0.10, *p* = .92). However, there was a significant effect of condition on youth-reported internalizing symptoms (WPC scores), such that youth in the Modular condition demonstrated a greater rate of decrease in their internalizing symptoms as compared to youth in the Standard condition (β = 3.38, *t* = 2.06, *p* = .04). SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) did not significantly affect the rate of change (slope) of youth-reported internalizing symptoms (WPC scores) over the course of treatment (β = 0.001, *t* = 0.12, *p* = .91). Similarly, SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) did not significantly affect the rate of change of youth-reported internalizing symptoms (WPC scores) regardless of what study condition youth were assigned (β = -0.47, *t* = -0.54, *p* = .59). Likewise, there was no significant effect of SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) on the rate of change of youth-reported internalizing symptoms (WPC scores) when accounting for both condition and time in treatment (β = -0.003, *t* = -0.27, *p* = .79). See Table 17.

Additionally, overall caregiver-reported youth internalizing scores did not change significantly over the course of treatment (β = 0.01, *t* = 0.44, *p* = .66). Unlike youth-reported

symptom change, there was not a significant effect of condition on caregiver-reported youth internalizing symptoms (WPC scores) (β = 1.56, *t* = 2.14, *p* = .47). SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) did not significantly affect the rate of change (slope) of caregiver-reported youth internalizing symptoms (WPC scores) over the course of treatment (β = -0.01, *t* = -1.20, *p* = .23). Similarly, SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) did not significantly affect the rate of change of caregiver-reported youth internalizing symptoms (WPC scores) regardless of what study condition youth were assigned (β = -0.58, *t* = -0.51, *p* = .61). Likewise, there was no significant effect of SDM practices directed towards youth clients (SDM-YMH Youth Total subscale scores) on the rate of change of caregiver-reported youth internalizing symptoms (WPC scores) when accounting for both condition and time in treatment (β = 0.001, *t* = 0.10, *p* = .93). All two-way and three-way interactions between condition, days, and SDM were not significant, suggesting that the rate of change did not differ according to either condition or initial SDM-YMH Youth Total score. See Table 18.

Caregiver Subscale Scores

First, overall caregiver-reported alliance scores did not change significantly over the course of treatment (β = 0.01, *t* = 1.84, *p* = .07); this was true for both conditions (β = -0.01 *t* = -1.88, *p* = .07). SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) did not significantly affect the rate of change (slope) of caregiver-clinician alliance (TPOCS-A Caregiver scores) over the course of treatment (β = -0.004, *t* = -1.77, *p* = .08). Similarly, SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) did not significantly affect the rate of change of caregiver-clinician alliance (TPOCS-A Caregiver scores) regardless of what study condition clients were assigned

Table 16

Shared Decision-Making in Youth Mental Health (Youth) Scores Predicting Youth-Clinician Alliance

Fixed effects						
Parameter	*B*	*SE*	*df*	*t*	*p*	95% CI
Intercept	3.24	0.63	0.00	5.16	1.00	[-1.45, 7.92]
Condition	0.16	0.61	121	0.27	.79	[-1.04, 1.37]
Days	0.01	0.004	121	1.39	.17	[-0.002, 0.01]
SDM	0.11	0.23	121	0.50	.62	[-0.33, 0.56]
Condition * Days	-0.01	0.01	121	-0.86	.39	[-0.02, 0.01]
Condition * SDM	-0.16	0.32	121	-0.51	.61	[-0.79, 0.47]
Days * SDM	-0.002	0.002	121	-1.47	.14	[-0.01, 0.001]
Condition * Days * SDM	0.003	0.003	121	0.80	.43	[-0.004, 0.009]

Note. Condition = Study Condition (Standard = 0; Modular = 1); Days = Days in Treatment (Time starts at 0); SDM = SDM-YMH Youth Total Subscale score

Table 17

Shared Decision-Making in Youth Mental Health (Youth) Scores Predicting Youth-Reported Internalizing Symptoms

Fixed effects						
Parameter	*B*	*SE*	*df*	*t*	*p*	95% CI
Intercept	1.81	1.40	494.15	1.30	0.20	[-0.93, 4.56]
Condition	3.38	1.64	131	2.06	0.04	[0.14, 6.61]
Days	-0.001	0.01	131	-0.10	0.92	[-0.02, 0.02]
SDM	-0.19	0.61	131	-0.31	0.76	[-1.41, 1.02]
Condition * Days	-0.01	0.02	131	-0.48	0.64	[-0.05, 0.03]
Condition * SDM	-0.47	0.88	131	-0.54	0.59	[-2.22, 1.27]
Days * SDM	0.001	0.01	131	0.12	0.91	[-0.01, 0.01]
Condition * Days * SDM	-0.003	0.01	131	-0.27	0.79	[-0.02, 0.02]

Note. Condition = Study Condition (Standard = 0; Modular = 1); Days = Days in Treatment (Time starts at 0); SDM = SDM-YMH Youth Total Subscale score

(β = -0.06, *t* = -0.23, *p* = .82). Likewise, there was no significant effect of SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) on the rate of change of caregiver-clinician alliance (TPOCS-A Caregiver scores) when accounting for both condition and time in treatment (β = 0.01, *t* = 1.73, *p* = .09). Yet, this association was in the predicted direction, such that higher instances of SDM practices directed towards caregivers was associated with increased caregiver-clinician alliance scores over the course of treatment, albeit not at a statically significant level. In short, all two-way and three-way interactions between condition, days, and SDM were not significant, suggesting that the rate of change did not differ according to either condition or initial SDM-YMH Caregiver Total score. See Table 19.

Again, overall caregiver-reported internalizing scores did not change significantly over the course of treatment (β = -0.02, *t* = -1.31, *p* = .19); this was true for both conditions (β = 2.40 *t* = 0.95, *p* = .35). SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) did not significantly affect the rate of change (slope) of caregiver-reported youth internalizing symptoms (WPC scores) over the course of treatment (β = 0.01, *t* = 0.91, *p* = .37). This association was in the predicted direction (i.e., higher SDM practices predicting greater clinical outcomes), however not at a level of statistical significance. Similarly, SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) did not significantly affect the rate of change of caregiver-reported youth internalizing symptoms (WPC scores) regardless of what study condition youth clients were assigned (β = -1.20, *t* = -0.61, *p* = .54). Likewise, there was no significant effect of SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) on the rate of change of caregiver-reported youth internalizing symptoms (WPC scores) when accounting for both condition and time in treatment (β = 0.25, *t* = 0.90, *p* = .37). Yet, this association was in the predicted direction, such that higher

Table 18

Shared Decision-Making in Youth Mental Health (Youth) Scores Predicting Caregiver-Reported Youth Internalizing Symptoms

	Fixed effects					
Parameter	*B*	*SE*	*df*	*t*	*p*	95% CI
Intercept	2.59	2.83	2887.96	.92	.36	[2.96, 8.14]
Condition	1.56	2.14	130	.73	.47	[-2.67, 5.79]
Days	0.01	0.01	130	0.44	.66	[-0.02, 0.03]
SDM	1.33	0.80	130	1.66	.10	[-0.25, 2.92]
Condition * Days	-0.001	0.03	130	-0.05	.96	[-0.05, 0.05]
Condition * SDM	-0.58	1.15	130	-0.51	.61	[-2.86, 1.69]
Days * SDM	-0.01	0.01	130	-1.20	.23	[-0.02, 0.01]
Condition * Days * SDM	0.001	0.01	130	0.10	.93	[-0.03, 0.03]

Note. Condition = Study Condition (Standard = 0; Modular = 1); Days = Days in Treatment (Time starts at 0); SDM = SDM-YMH Youth Total Subscale score

instances of SDM practices directed towards caregivers was associated with increased caregiver-reported youth outcomes over the course of treatment, yet not at a statically significant level. Again, all two-way and three-way interactions between condition, days, and SDM were not significant, suggesting that the rate of change did not differ according to either condition or initial SDM-YMH Caregiver Total score. See Table 20.

Overall youth-reported internalizing scores did not change significantly over the course of treatment (β = -0.004, t = -0.3, p = .76); this was true for both conditions (β = 0.004 t = 0.14, p = .89). SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) did not significantly affect the rate of change (slope) of youth-reported internalizing symptoms (WPC scores) over the course of treatment (β = 0.003, t = 0.30, p = .77). This association was in the predicted direction (i.e., higher SDM practices predicting greater clinical outcomes), however not at a level of statistical significance. Similarly, SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) did not significantly affect the rate of change of youth-reported internalizing symptoms (WPC scores) regardless of what study condition youth clients were assigned (β = 0.24, t = 0.16, p = .87). Likewise, there was no significant effect of SDM practices directed towards caregivers (SDM-YMH Caregiver Total subscale scores) on the rate of change of youth-reported internalizing symptoms (WPC scores) when accounting for both condition and time in treatment (β = -0.01, t = -0.63, p = .53). Again, all two-way and three-way interactions between condition, days, and SDM were not significant, suggesting that the rate of change did not differ according to either condition or initial SDM-YMH Caregiver Total score. See Table 21.

Table 19

Shared Decision-Making in Youth Mental Health (Caregiver) Scores Predicting Caregiver-Clinician Alliance

Fixed effects						
Parameter	*B*	*SE*	*df*	*t*	*p*	95% CI
Intercept	3.11	0.25	57	12.39	<.001	[2.60, 3.61]
Condition	0.25	0.36	57	0.68	.50	[-0.48, 0.97]
Days	0.01	0.003	57	1.84	.07	[-0.001, 0.01]
SDM	0.18	0.16	57	1.11	.27	[-0.14, 0.50]
Condition * Days	-0.01	0.01	57	-1.88	.07	[-0.02, 0.001]
Condition * SDM	-0.06	0.24	57	-0.23	.82	[-0.54, 0.43]
Days * SDM	-0.004	0.002	57	-1.77	.08	[-0.01, 0.001]
Condition * Days * SDM	0.01	0.004	57	1.73	.09	[-0.001, 0.01]

Note. Condition = Study Condition (Standard = 0; Modular = 1); Days = Days in Treatment (Time starts at 0); SDM = SDM-YMH Caregiver Total Subscale score

Table 20

Shared Decision-Making in Youth Mental Health (Caregiver) Scores Predicting Caregiver-Reported Youth Internalizing Symptoms

Fixed effects						
Parameter	*B*	*SE*	*df*	*t*	*p*	95% CI
Intercept	5.39	1.53	130	3.53	<.001	[2.36, 8.41]
Condition	2.40	2.54	130	0.95	.35	[-2.61, 7.42]
Days	-0.02	0.02	130	-1.31	.19	[-0.06, 0.01]
SDM	-0.61	1.21	130	-0.51	.61	[-2.99, 1.77]
Condition * Days	-0.03	0.03	130	-0.92	.36	[-0.10, 0.04]
Condition * SDM	-1.20	1.95	130	-0.61	.54	[-5.07, 2.67]
Days * SDM	0.01	0.01	130	0.91	.37	[-0.01, 0.04]
Condition * Days * SDM	0.25	0.03	130	0.90	.37	[-0.03, 0.08]

Note. Condition = Study Condition (Standard = 0; Modular = 1); Days = Days in Treatment (Time starts at 0); SDM = SDM-YMH Caregiver Total Subscale score

Table 21

Shared Decision-Making in Youth Mental Health (Caregiver) Scores Predicting Youth-Reported Internalizing Symptoms

Fixed effects						
Parameter	*B*	*SE*	*df*	*t*	*p*	95% CI
Intercept	0.81	3.70	0.00	0.22	1.00	[-8.23, 9.84]
Condition	2.03	1.94	131	1.05	.30	[-1.81, 5.87]
Days	-0.004	0.01	131	-0.30	.76	[-0.03, 0.02]
SDM	0.60	0.93	131	0.65	.52	[-1.23, 2.43]
Condition * Days	0.004	0.03	131	0.14	.89	[-0.05, 0.06]
Condition * SDM	0.24	1.50	131	0.16	.87	[-2.72, 3.20]
Days * SDM	0.003	0.01	131	0.30	.77	[-0.02, 0.02]
Condition * Days * SDM	-0.01	0.02	131	-0.63	.53	[-0.06, 0.03]

Note. Condition = Study Condition (Standard = 0; Modular = 1); Days = Days in Treatment (Time starts at 0); SDM = SDM-YMH Caregiver Total Subscale score

Discussion

SDM has been posited as a potential method by which clinicians can adapt psychotherapy treatments, including EBPs, to increase their fit with individual families, thereby potentially improving the alliance and clinical outcomes (Langer & Jensen-Doss, 2018; Loh et al., 2006; Street et al., 2008). Yet, there is a dearth of reliable and valid measures of SDM practices, especially with regard to youth psychotherapy. Thus, the goal of the present study was to develop the SDM-YMH, an observer-rated measure of SDM practices for youth psychotherapy, and examine key psychometric properties, including inter-rater reliability, construct validity, and predictive validity. This project incorporated SDM practices established by previous empirical studies (e.g., Alegría et al., 2018; Aoki et al., 2019) and theoretical works (e.g., Clayman & Makoul, 2009; Langer & Jensen-Doss, 2018) to develop the item content of the SDM-YMH items. Overall, eight experts in the field of SDM agreed that 13 of the originally proposed 14 items were representative core SDM practices (see Table 1 above). While 31% of SDM-YMH Youth items and 54% of SDM-YMH Caregiver items were able to be coded reliably at a level considered "good" (i.e., ICC $\geq$ .60, Cicchetti, 1994), the construct validity of the SDM-YMH item and subscale scores as well as the predictive validity of the SDM-YMH subscale scores are not supported at this time.

Content Validity

One of most promising findings of the current project was the strong evidence to support the content validity of the SDM-YMH items. Specifically, 13 of the original 14 items were retained in the final version of the measure due to favorable content validity ratios (i.e., CVRs of at least zero). That SDM-YMH items demonstrated acceptable content validity based on expert review increases the potential usability of the scale in future studies (e.g., to examine

psychometric properties further) as well as increases the interpretability of the current study's findings. In other words, findings from reliability and validity analyses can be interpreted with a reasonable degree of confidence that the items and their definitions were representative of SDM practices applicable for youth psychotherapy.

Reliability

It was anticipated that SDM-YMH items and subscale scores would demonstrate at least "good" inter-rater reliability (ICCs ≥ .60; Cicchetti, 1994; Gärtner et al., 2018). As aforementioned, this was partially supported. That is, one of the SDM-YMH Youth items (Define/Explain Problem) and several SDM-YMH Caregiver items (Define/Explain Problem, Present Options, Mutual Agreement) met this criteria. Additionally, the final SDM-YMH subscales (Youth Total, Youth Short, Caregiver Total, Caregiver Short) demonstrated "good" to "excellent" inter-rater reliability. Overall, inter-rater reliability was stronger for Caregiver item and subscales scores than for Youth item and subscales scores. This is likely due to clinicians more explicitly involving caregivers in treatment decisions throughout the therapy tapes coded as part of the current study. Thus, SDM practices were more easily identified and coded when clinicians interacted with caregivers as opposed to youth clients. In general, the lower than hypothesized inter-rater reliability is likely due to there being fewer opportunities than anticipated to observe SDM practices in these therapy tapes (see below for further explanation). One of the primary reasons being that the Child STEPs Multisite Trial clinicians were not purposefully implementing an SDM intervention, which contributed to the low base rate of SDM practices observed in the current study. The low base rate of SDM practices also contributed to the largely restricted range of most SDM-YMH items as well as the non-normality of the data (Hallgren, 2012). Similarly, there were several SDM-YMH items that were never or rarely coded

with this sample (e.g., Equipoise). The low base rate of SDM practices in the included tapes also likely contributed to difficulties reaching acceptable inter-rater reliability for many of the SDM-YMH items, especially the Youth items. With regards to coding, about half of the included tapes were audio recordings, which made it difficult to discern to whom the SDM practice was directed (i.e., to youth or caregiver). It is difficult to discern whether the coding manual (containing the item definitions, exemplars, and distinctions) was insufficient in outlining SDM practices, given that the coders did not have an abundance of opportunities to code SDM practices in the included tapes. In other words, the lower than anticipated inter-rater reliability was likely due to a lack of opportunities for coders to code SDM practices in the included tapes. However, future works with this measure should revisit item exemplars and definitions if inter-rater reliability is still poor when used with a sample of tapes in which clinicians were explicitly instructed to use SDM practices in youth psychotherapy.

These findings are similar to previous examinations of the psychometric properties of other SDM measures. For example, Alegría et al. (2021) also utilized an observer-rated measure of SDM in their effectiveness trial, examining the effects of SDM on clinical outcomes for adult patients in outpatient community settings. The current study's measure demonstrated slightly stronger inter-rater reliability at the subscale level (Youth Total: ICC[2,2] = .64; Youth Short: ICC[2,2] = .62; Caregiver Total: ICC[2,2] = .78; Caregiver Short: ICC[2,2] = .81) than the observer-rated measure that Alegría et al. (2021) used (subscale ICC[2,2] = .53; "fair"; Cicchetti, 1994). However, a primary difference between this study and Alegría et al. (2021)'s was that Alegría et al. (2021)'s sample was comprised of adults receiving psychotherapy, whereas the current study employed a youth population. Observational coding of SDM practices (e.g., OPTION-5; Barr et al., 2015) was also utilized in the only known RCT examining the effects of

an SDM intervention on alliance and clinical outcomes in youth psychotherapy (Langer et al., 2022). While Langer et al. (2022)'s findings suggested that clinicians were adherent to the protocol, inter-rater reliability data were not available for this study's measures (i.e., OPTION-5). Yet, in other studies that also utilized the OPTION and OPTION-5, the subscale scores have demonstrated similar inter-rater reliability (e.g., OPTION-5 subscale scores: ICC[2,2] = .67—70; Barr et al., 2015; OPTION total score: ICC[2,2] = .77; Elwyn et al., 2005) as compared to the SDM-YMH subscale scores in the current study.

Notably, the SDM-YMH Youth Total subscale only included four of the 13 items due to poor item-level inter-rater reliability. Similarly, due to poor item-level inter-rater reliability, the SDM-YMH Caregiver Total subscale only included seven of the 13 items. Had all items been included in the SDM-YMH subscales, the inter-rater reliability still would have been "good" for the Youth Total subscale (ICC[2,2] = .64) but "good" for the Caregiver Total subscale (ICC[2,2] = .73; Cicchetti, 1994). Yet, this is likely due to several items having no variance between coders 1 and 2, since they were not coded frequently (e.g., Decision Identification) or at all (e.g., Equipoise). Thus, there is preliminary evidence to suggest that the inter-rater reliability of the SDM-YMH Youth Total/Short and SDM-YMH Caregiver Total/Short subscales (but not items) is similar to the inter-rater reliability of subscale scores on established measures (e.g., OPTION-5; Barr et al., 2015). However, whether this pattern would remain true if more SDM-YMH items were used more frequently and included in subscale scores is unclear.

Construct Validity

It was also anticipated that SDM-YMH item and subscale scores would demonstrate greater shared variance with theoretically similar constructs (convergent validity) and less shared variance with theoretically dissimilar constructs (discriminant validity; Campbell & Fiske, 1959;

Foster & Cone, 1995; Kazdin, 2016). More specifically, it was anticipated that SDM-YMH scores would be most strongly associated with scores on TPOCS-RS Goal Setting and Session Agenda items (i.e., items assessing the degree to which clinicians engage families in discussions regarding long- and short-term treatment goals). Then, SDM-YMH item and subscale scores were expected to be less strongly associated with scores on measures of alliance (TPOCS-A). Lastly, SDM-YMH scores were expected to be least strongly associated with scores on measures of treatment integrity (CBAY-A, CBAY-C, TPOCS-RS Psychodynamic and Family). However, findings from the current study did not reflect this pattern and thus did not support the construct validity of SDM-YMH scores.

SDM-YMH Youth item (Define/Explain Problem, Present Options, Patient Values/Preferences, and Make/Explicitly Defer Decision) and subscale (Total/Short) scores were hypothesized to be correlated with scores on theoretically similar measures (e.g., TPOCS-RS Goal Setting) at a "large" magnitude (Rosenthal & Rosnow, 1984). However, this pattern was not observed; thus, convergent validity was not supported for any SDM-YMH item or subscale scores. However, discriminant validity was supported in that SDM-YMH item (Define/Explain Problem, Present Options, Patient Values/Preferences, and Make/Explicitly Defer Decision) and subscale scores (Total/Short) were correlated with scores on measures of theoretically dissimilar constructs (e.g., TPOCS-RS Family) at a "small" magnitude (Rosenthal & Rosnow, 1984).

However, the associations observed amongst SDM-YMH Youth item scores (Define/Explain Problem, Present Options, Patient Values/Preferences, and Make/Explicitly Defer Decision) and scores on measures of theoretically similar constructs (e.g., TPOCS-RS Goal Setting) did not statistically differ from the associations between SDM-YMH Youth items and scores on measures of theoretically dissimilar constructs (e.g., TPOCS-RS Psychodynamic).

That is, the hypothesized pattern of correlations was not observed, and SDM-YMH Youth item scores were no more similar to scores on theoretically similar subscales than those on theoretically dissimilar subscales. The same pattern was observed of SDM-YMH Youth subscales (Total and Short); such that SDM-YMH Youth item scores were no more similar to scores on theoretically similar subscales (e.g., TPOCS-RS Goal Setting) than they were to scores on theoretically dissimilar subscales (e.g., TPOCS-RS Family). Thus, construct validity was not supported for the SDM-YMH Youth item and subscale scores.

Similarly, SDM-YMH Caregiver item (Elicits Goals for Therapy, Define/Explain Problem, Present Options, Present Evidence, Patient Values/Preferences, Mutual Agreement, Check/Clarify Understanding) and subscale (Total/Short) scores were hypothesized to be correlated with scores on theoretically similar measures (e.g., TPOCS-RS Goal Setting) at a "large" magnitude (Rosenthal & Rosnow, 1984). This pattern was partially observed for two items (Define/Explain Problem and Patient Values/Preferences) and subscales (Total/Short). Thus, convergent validity was partially supported for some SDM-YMH Caregiver item and subscale scores.

However, discriminant validity was not supported for the majority of SDM-YMH Caregiver item and subscale scores. That is, scores on several Caregiver items (Elicits Goals for Therapy, Define/Explain Problem, Present Options, Patient Values/Preferences, Check/Clarify Understanding) and subscales (Total and Short) were correlated with scores on a theoretically dissimilar measure (i.e., TPOCS-RS Family) at a "large" magnitude, not the hypothesized "small" magnitude. In several instances, SDM-YMH Caregiver scores were statistically redundant with TPOCS-RS Family scores (i.e., Present Options, Patient Values/Preferences, Total subscale, and Short subscale).

Similarly, the construct validity of SDM-YMH Caregiver item and subscale scores was not supported. The associations observed amongst SDM-YMH Caregiver item scores (Elicits Goals for Therapy, Define/Explain Problem, Present Options, Present Evidence, Patient Values/Preferences, Mutual Agreement, Check/Clarify Understanding) and scores on measures of theoretically similar constructs (e.g., TPOCS-RS Goal Setting) did not statistically differ from the associations between SDM-YMH Caregiver items and scores on measures of theoretically dissimilar constructs (e.g., TPOCS-RS Psychodynamic). Thus, the construct validity of these items was not supported. The same pattern was observed of SDM-YMH Caregiver subscales (Total and Short). That is, SDM-YMH Caregiver item scores were no more similar to scores on theoretically similar subscales than they were to scores on theoretically dissimilar subscales.

One important caveat to the aforementioned should be noted. That is, the construct validity of the SDM-YMH Youth and Caregiver subscales as originally designed was unable to be assessed in the current study. Since several items were removed from both subscales due to poor inter-rater reliability, the subscales as examined in the current study do not represent the full item list of each subscale (only a subset of items). Future research should work to establish inter-rater reliability for all SDM-YMH items and then reassess the construct validity of each item and subscale. It is important to examine the construct validity of the subscales with all items included, especially as content validity ratios were favorable for all 13 SDM-YMH items, suggesting that they are all at least somewhat representative of SDM in youth psychotherapy.

The present study's findings are similar to the overall trend in the SDM literature, such that many studies have either struggled to demonstrate adequate psychometric properties (e.g., construct validity) of their SDM measures' scores (e.g., Simon et al., 2006; Stacey et al., 2008) or have omitted the analysis of construct validity altogether (e.g., Clayman et al., 2012). One

possible explanation for why the current study struggled to demonstrate construct validity for the SDM-YMH item and subscale scores is that there were no other measures of SDM practices available by which to compare the SDM-YMH scores. That is, in an ideal situation, per MTMM (Campbell & Fiske, 1959), an established observer-rated measure of SDM practices (e.g., OPTION-5; Elwyn et al., 2013) as well as a clinician-report measure of SDM practices (e.g., SDM-Q-Doc, Scholl et al., 2012) would have been included in the current study. This would have allowed for comparisons to measures that have already been established to assess SDM practices instead of measures that included only some aspects of SDM (i.e., TPOCS-RS Goal Setting and Session Agenda).

Predictive Validity

Moreover, it was expected that SDM-YMH scores would be predictors of change in the alliance and youth clinical outcomes throughout treatment. Specifically, higher observed SDM practices were hypothesized to predict a stronger alliance (youth-clinician, caregiver-clinician; Langer & Jensen-Doss, 2018) and greater improvement in youth clinical (internalizing) outcomes (Ackerman & Hilsenroth, 2003; Fjermestad et al., 2020; Langer & Jensen-Doss, 2018). However, predictive validity of the SDM-YMH scores was not supported in the present study, as neither Youth nor Caregiver SDM-YMH scores predicted improvements in alliance (youth-clinician or caregiver-clinician) or clinical outcomes (youth or caregiver-reported youth internalizing symptoms) throughout treatment. While some findings were trending in the anticipated direction (e.g., higher instances of SDM practices directed towards caregivers was associated with increased caregiver-clinician alliance scores over the course of treatment), these findings did not rise to the level of statistical significance. It is likely that the poor construct

validity of SDM-YMH subscales, low incidence of SDM practices present in the current study, as well as the non-normality of the data contributed to this null finding.

The present study's difficulty to demonstrate an effect of SDM practices on alliance and clinical outcomes is similar to Langer et al. (2022)'s findings, such that they also did not find evidence for higher instances of SDM increasing youth-therapist alliance or youth clinical outcomes. One important distinction between Langer et al. (2022)'s study and the present one is that the present study included a sample that was adequately powered to test the relationship between SDM, the alliance, and outcomes. Whereas Langer et al. (2022)'s study was a pilot study and thus underpowered to detect this association. Thus, the current study was powered to detect a significant association between SDM practices, the alliance, and clinical outcomes if such an association were present in the current data and if construct validity was supported. Another important distinction is that Langer et al. (2022)'s study involved clinicians intentionally implementing a specific SDM intervention, whereas this study involved coding SDM practices that were more "naturally occurring". That is, clinicians in the tapes coded as part of the current study were not instructed to deliver SDM practices. It could be that the low incidence of SDM practices in the current study partially contributed to the difficulties in establishing stronger psychometric properties, including predictive validity. Yet, as both this study and Langer et al. (2022)'s failed to demonstrate a statistically significant association between SDM practices, alliance, and outcomes, it is also possible that such an association does not exist. However, future research is needed to examine this question further.

Limitations

There are several limitations of this study that may have influenced the current findings. The first and potentially most impactful limitation of this study were the therapy tapes that were

coded. As aforementioned, the Child STEPs Multisite Trial was an effectiveness trial that was conducted in community mental health centers in Massachusetts and Hawaii. These therapy sessions were conducted by clinicians implementing one of two manualized cognitive-behavioral treatments in the early to mid-2000s. Thus, clinicians in these tapes were not explicitly instructed to implement SDM practices. As the integration of SDM into youth psychotherapy is relatively new, it would have been unlikely for Child STEPs Multisite Trial clinicians to organically implement a significant amount of SDM practices. Additionally, even in the Modular condition in which clinicians were encouraged to adapt the treatment plan to best fit clients' needs when appropriate, such adaptations were largely clinician- and supervisor-driven. It is possible that a large amount of the adaptations to these clients' treatment plans happened in supervision meetings and not in session with clients and their families. Moreover, while there was an exerted effort made to distinguish between SDM practices and clinician behaviors consistent with a collaborative style (e.g., for rapport building) when developing the SDM-YMH items and codebook, easily identifying the SDM practices as written was difficult to achieve while coding the Child STEPs Multisite tapes. For example, many exemplars for SDM-YMH items were highly specific and occurred only briefly during these tapes in large part because discussing treatment decisions with clients and families was not a designated component of either EBP implemented as a part of the Child STEPs Multisite Trial. Therefore, the limits of what was codable were expanded slightly. For example, many codable clinician behaviors in this study included single-phrase utterances in the context of discussing treatment decisions (e.g., clinician saying, "I'd like to choose something more manageable for you and your family" would have been coded under Patient/Values Preferences). Coding such small components made it more difficult for signal detection while coding. Using these tapes also required coders to correctly

identify what was a “treatment decision” versus behavior consistent with collaborative style. Lastly, coding such small behavioral components likely increased the chance of coders accidentally coding behaviors that only approximated SDM practices, as the line between treatment decision and collaborative style was especially difficult to determine in a treatment with so few formal treatment planning discussions.

Additionally, the Child STEPs Multisite Trial sample only included youth clients aged eight to 13. It is likely that teenagers would play a more active role in their own psychotherapy treatment planning. For example, most typically developing children under the age of 12 do not have mastery over abstract thought (Piaget, 1964), which would be beneficial to fully understand many decision points throughout the course of psychotherapy. Similarly, most typically developing children struggle to forgo immediate comfort (e.g., avoiding anxiety-provoking stimuli) in service of a future goal (e.g., feeling less anxious), especially if working towards said goal required significant discomfort (e.g., participating in exposures; Atance, 2008; Mischel et al., 1989). On the contrary, teenage clients would be further along in developing their ability to plan for the future, delay gratification, and critically think about complex subjects (e.g., Blakemore & Choudhury, 2006; Nurmi, 1991). Likewise, child and young adolescent clients would likely require more assistance from their caregivers to make difficult decisions (e.g., treatment decisions); whereas older adolescent and teenage clients would likely be more strongly motivated to individuate from their caregivers and practice making their own decisions (e.g., Erikson, 1994). While it is still possible to deliver SDM practices to younger children (with appropriate developmental adaptations), directing these practices towards younger children is likely less common, especially before the study of SDM in youth psychotherapy began (i.e., when Child STEPs Multisite Trial tapes were recorded). Therefore, this younger sample likely

contributed to the lower base rate of SDM practices, as well as precluded the examination of differences amongst children, adolescents, and teenagers with regards to SDM practices in youth psychotherapy. Additionally, clinicians were not instructed to implement SDM practices with caregivers either in the current study. The fewer opportunities to observe clinicians implementing SDM practices with caregivers also likely contributed to the lower inter-rater reliability of the SDM-YMH Caregiver items.

Lastly, the coding assignment was a limitation of the current study. That is, an ideal coding assignment would have been accomplished using a balanced incomplete block design (Fleiss, 1981). This would have allowed for the inter-rater reliability of the item and subscale scores of the SDM-YMH to be estimated based off of the group's coding. Instead, due to availability of coders, the inter-rater reliability for SDM-YMH scores was estimated by comparing the codes of most coders to those from a single coder (i.e., the first author).

Future Directions

One of the most important next steps following this study is to further examine the psychometric properties of the SDM-YMH. The following recommendations are offered for future psychometric evaluation of the SDM-YMH. First, pilot coding with a sample of therapy tapes from an efficacy trial of an SDM intervention would help refine some of the definitions and exemplars of the SDM practices in the SDM-YMH codebook. These therapy tapes would likely increase the opportunities for coders to observe more explicit implementations of SDM practices. After SDM-YMH item definitions and exemplars are further clarified, coder training should take place with an independent review of the codebook and stronger exemplar tapes. For example, it will be advantageous for coders to be trained on clearer examples of SDM practices (e.g., tapes in which SDM practice extensiveness was high—i.e., a 6 or 7/7) so that they have a stronger

understanding of which clinician behaviors are codable (i.e., improve signal detection). Similarly, clearer exemplar tapes of low and medium SDM practice extensiveness would help coders better differentiate SDM practice "dose" and potentially improve inter-rater reliability. Similarly, this may improve the normality of the data, as coders will be better equipped to utilize the full range of the scale. Moreover, it would be helpful to establish a coding team of individuals who were able to meet consistently together and who have prior coding experience to reduce variation in coder expertise. More evenly splitting coding assignments amongst coders (as opposed to the coding design of the current study in which all codes were compared to coder 1) would also be useful in establishing the reliability of the SDM-YMH without relying too heavily on any given coder, thus improving the generalizability of the SDM-YMH's psychometric properties. As aforementioned, including measures more similarly related to SDM (e.g., SDM-Q-Doc, Scholl et al., 2012) would improve the ability to establish the convergent validity of the SDM-YMH scores. After these steps are taken, predictive validity should again be examined to determine whether SDM-YMH scores can predict alliance and clinical outcomes in youth psychotherapy.

After further establishing the psychometric properties of the SDM-YMH, it would also be helpful to examine the following. First, it is important to examine the current study's hypotheses with a different population. As aforementioned, the current study's youth sample was limited to children ages eight to 13. It is likely that older adolescents and teenagers would take a more active role in their own psychotherapy treatment plans. Thus, it is important to examine whether SDM increases the effectiveness of evidence-based psychotherapy in this population. Moreover, it would be helpful to replicate the present study with more current therapy tapes. As aforementioned, utilizing SDM practices in youth psychotherapy is a relatively new concept.

Therefore, it is possible that the base rate of SDM practices being implemented would be higher in therapy sessions conducted more recently than those included in this study. Similarly, using the SDM-YMH to examine sessions in which clinicians were explicitly instructed to include SDM practices would likely increase the base rate of observed SDM practices, providing a richer opportunity to examine the psychometric properties of the SDM-YMH.

Additionally, examining SDM practices in settings other than community mental health centers would help inform evidence-based care more broadly, especially as psychotherapy expands further into interdisciplinary spaces (e.g., integrated primary care). Examining how SDM practices influence the effectiveness of evidence-based care in these settings would likely be distinct from community mental health settings, as clinicians in interdisciplinary settings often collaborate with medical professionals more frequently. Therefore, there is an increased opportunity to integrate medical interventions into psychotherapy and include more providers (e.g., psychiatrists, nurses, physicians) in decision-making. Lastly, in some interdisciplinary settings, like integrated primary care, psychotherapy may be shorter and more targeted than in community mental health settings. Therefore, SDM may play a different role in treatment planning, as there may be a more urgent need to focus treatment more quickly and decide what to address and what to omit.

Conclusion

Despite the largely null findings and aforementioned limitations, the measure developed as part of this project, the SDM-YMH, can potentially be used in future work regarding SDM practices in youth psychotherapy. As aforementioned, the therapy tapes used to code the SDM-YMH items were not an ideal fit, as they did not contain an abundance of SDM practices to observe and code. Thus, it is possible that the SDM-YMH will perform stronger

psychometrically if used to code therapy tapes from a different sample. Although this project did not support the construct validity or predictive validity of SDM-YMH scores in this sample, this project demonstrated modest support for the score reliability of a few SDM-YMH items (e.g., Patient Values/Preferences) as well as the content validity for the majority of items. The continued examination of the psychometric properties of the SDM-YMH is warranted, especially given the content validity established for the majority of items as well as the remaining gap in the SDM field for an observed-rater measure specific to youth psychotherapy. It is important for further research to continue studying SDM, especially as it has the potential to help providers effectively adapt evidence-based mental health treatments to best serve the youth and families who receive them so that the short-term negative effects of mental illness can be assuaged, and the long-term negative effects can be prevented.

References

Aarons, G. A., Wells, R. S., Zagursky, K., Fettes, D. L., & Palinkas, L. A. (2009). Implementing Evidence-Based Practice in Community Mental Health Agencies: A Multiple Stakeholder Analysis. *American Journal of Public Health*, *99*(11), 2087–2095. https://doi.org/10.2105/AJPH.2009.161711

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*.

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review*, *23*(1), 1–33. https://doi.org/10.1016/S0272-7358(02)00146-0

Aggarwal, N. K., Pieh, M. C., Dixon, L., Guarnaccia, P., Alegría, M., & Lewis-Fernández, R. (2016). Clinician descriptions of communication strategies to improve treatment engagement by racial/ethnic minorities in mental health services: A systematic review. *Patient Education and Counseling*, *99*(2), 198–209. https://doi.org/10.1016/j.pec.2015.09.002

Alegría, M., Nakash, O., Johnson, K., Ault-Brutus, A., Carson, N., Fillbrunn, M., Wang, Y., Cheng, A., Harris, T., Polo, A., Lincoln, A., Freeman, E., Bostdorf, B., Rosenbaum, M., Epelbaum, C., LaRoche, M., Okpokwasili-Johnson, E., Carrasco, M., & Shrout, P. E. (2018). Effectiveness of the DECIDE Interventions on Shared Decision Making and Perceived Quality of Care in Behavioral Health With Multicultural Patients: A Randomized Clinical Trial. *JAMA Psychiatry*, *75*(4), 325-335. https://doi.org/10.1001/jamapsychiatry.2017.4585

Allen, J. D., Linnan, L. A., & Emmons, K. M. (2012). Fidelity and its relationship to

implementation effectiveness, adaptation, and dissemination. In *Dissemination and implementation research in health: Translating science to practice* (R. C. Brownson, G. A. Colditz & E. K. Proctor, pp. 281–304). New York: Oxford University Press.

American Educational Research Association. (2014). *The Standards for Educational and Psychological Testing*. https://www.apa.org/science/programs/testing/standards

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th Text Revision).

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling, 12*(3), 411-434.

Atance, C. M. (2008). Future thinking in young children. *Current Directions in Psychological Science, 17*(4), 295-298.

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development, 47*(1), 79-86. https://doi.org/10.1177/0748175613513808

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*(5), 493–500. https://doi.org/10.1037/0735-7028.38.5.493

Barkley, R. (2013). *Defiant children: A clinician's manual for assessment and parent training*. Guilford press.

Barry, C. T., & Pickard, J. D. (2008). Developmental issues. In *Handbook of assessment, case conceptualization, and treatment Volume II: Children and Adolescents* (M. Hersen&D. Reitman, pp. 76–101). John Wiley & Sons.

Baumann, A., Cabassa, L. J., & Stirman, S. W. (2017). Adaptation in dissemination and implementation science. In *Dissemination and Implementation Research in Health: Translating Science to Practice* (Vol. 2, pp. 286–300).

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy, 16, 252–260.

Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*(3-4), 296-312.

Blumenthal-Barby, J., Opel, D. J., Dickert, N. W., Kramer, D. B., Tucker Edmonds, B., Ladin, K., Peek, M. E., Peppercorn, J., & Tilburt, J. (2019). Potential Unintended Consequences Of Recent Shared Decision Making Policy Initiatives. *Health Affairs*, *38*(11), 1876–1881. https://doi.org/10.1377/hlthaff.2019.00243

Brown, R. C., Southam-Gerow, M. A., McLeod, B. D., Wheat, E. B., Tully, C. B., Reise, S. P., Kendall, P. C., & Weisz, J. R. (2018). The global therapist competence scale for youth psychosocial treatment: Development and initial validation. *Journal of Clinical Psychology*, *74*(4), 649–664. https://doi.org/10.1002/jclp.22537

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

Carroll, K. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*(3), 225–238. https://doi.org/10.1016/S0376-8716(99)00049-6

CDC. (2020, June 15). *Data and Statistics on Children's Mental Health*. from https://www.cdc.gov/childrensmentalhealth/data.html

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*, *52*(6), 1395-1407. https://doi.org/10.1016/j.beth.2021.03.009

Cecilione Herbst, J. L. (2022). Adapting Current Evidence-Based Practices for Promoting Shared Decision Making to Youth Psychotherapy: A Descriptive Review. Unpublished manuscript.

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.

Chambers, D. A., & Norton, W. E. (2016). The Adaptome. *American Journal of Preventive Medicine*, *51*(4), S124–S131. https://doi.org/10.1016/j.amepre.2016.05.011

Cheng, H., Hayes, D., Edbrooke-Childs, J., Martin, K., Chapman, L., & Wolpert, M. (2017). What approaches for promoting shared decision-making are used in child mental health? A scoping review. *Clinical Psychology & Psychotherapy*, *24*(6), 1495–1511. https://doi.org/10.1002/cpp.2106

Chiu, A. W., McLeod, B. D., Har, K., & Wood, J. J. (2009). Child-therapist alliance and clinical outcomes in cognitive behavioral therapy for child anxiety disorders. *Journal of Child Psychology and Psychiatry*, *50*(6), 751–758. https://doi.org/10.1111/j.1469-7610.2008.01996.x

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., Phillips, L., Ward, A., Lynch, R., Trent, L., Smith, R. L., Okamura, K., & Starace, N. (2011). Evidence-Based Treatments for Children and Adolescents: An Updated Review

of Indicators of Efficacy and Effectiveness: Evidence-based treatments for children and adolescents. *Clinical Psychology: Science and Practice*, *18*(2), 154–172. https://doi.org/10.1111/j.1468-2850.2011.01247.x

Chorpita, B. F., & Weisz, J. R. (2005). *MATCH-ADC: A Modular Approach to Treatment for Children with Anxiety, Depression, and Conduct Problems*.

Clark, L. A., & Watson, D. (1995). *Constructing validity: Basic issues in objective scale development*. *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Clayman, M. L., & Makoul, G. (2009). Conceptual variation and iteration in shared decision-making: The need for clarity. In A. Edwards & G. Elwyn (Eds.), Shared decision-making in health care (2nd ed., pp. 109–116). Oxford, UK: Oxford University Press.

Clayman, M. L., Makoul, G., Harper, M. M., Koby, D. G., & Williams, A. R. (2012). Development of a shared decision making coding system for analysis of patient–healthcare provider encounters. *Patient Education and Counseling*, *88*(3), 367-372.

Cochrane. (2021). *RoB 2: A revised Cochrane risk-of-bias tool for randomized trials*. RoB 2: A revised Cochrane risk-of-bias tool for randomized trials | Cochrane Bias. Retrieved September 14, 2021, from https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials

Connor-Smith, J. K., & Weisz, J. R. (2003). Applying Treatment Outcome Research in Clinical Practice: Techniques for Adapting Interventions to the Real World: *Commissioned Review: Applying Treatment Outcome Research*. *Child and Adolescent Mental Health*, *8*(1), 3–10. https://doi.org/10.1111/1475-3588.00038

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cummings, J. R., Case, B. G., Ji, X., & Marcus, S. C. (2016). Availability of Youth Services in U.S. Mental Health Treatment Facilities. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(5), 717–727. https://doi.org/10.1007/s10488-015-0685-2

Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing Main Effects and Interactions in Latent Curve Analysis. *Psychological Methods*, *9*(2), 220–237. https://doi.org/10.1037/1082-989X.9.2.220

De las Cuevas, C., & Peñate, W. (2014). To what extent psychiatric patients feel involved in decision making about their mental health care? Relationships with socio-demographic, clinical, and psychological variables. *Acta Neuropsychiatrica*, *26*(6), 372-381. https://doi.org/10.1017/neu.2014.21

Donner, A., Birkett, N., & Buck, C. (1981). *Randomization by cluster: Sample size requirements and analysis*. *114*(6), 906–914.

Duncan, T. E., & Duncan, S. C. (2004). A Latent Growth Curve Modeling Approach to Pooled Interrupted Time Series Analyses. *Journal of Psychopathology and Behavioral Assessment*, *26*(4), 271–278. https://doi.org/10.1023/B:JOBA.0000045342.32739.2f

Durand, M.-A., Carpenter, L., Dolan, H., Bravo, P., Mann, M., Bunn, F., & Elwyn, G. (2014). Do Interventions Designed to Support Shared Decision-Making Reduce Health Inequalities? A Systematic Review and Meta-Analysis. *PLoS ONE*, *9*(4), e94670. https://doi.org/10.1371/journal.pone.0094670

Edwards, A., Elwyn, G., Hood, K., Robling, M., Atwell, C., Holmes-Rovner, M., ... & Russell, I. (2003). The development of COMRADE—a patient-based outcome measure to evaluate the effectiveness of risk communication and treatment decision making in

consultations. *Patient Education and Counseling*, *50*(3), 311-322. https://doi.org/10.1016/S0738-3991(03)00055-7

Eliacin, J., Salyers, M. P., Kukla, M., & Matthias, M. S. (2015). Patients' Understanding of Shared Decision Making in a Mental Health Setting. *Qualitative Health Research*, *25*(5), 668–678. https://doi.org/10.1177/1049732314551060

Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, *28*(7), 1167–1187. https://doi.org/10.1016/j.cpr.2008.04.002

Elwyn, G., Edwards, A., Kinnersley, P., & Grol, R. (2000). Shared decision making and the concept of equipoise: the competences of involving patients in healthcare choices. *British Journal of General Practice*, *50*(460), 892-899.

Elwyn, G., Edwards, A., Mowle, S., Wensing, M., Wilkinson, C., Kinnersley, P., & Grol, R. (2001). Measuring the involvement of patients in shared decision-making: A systematic review of instruments. *Patient Education and Counseling*, *43*(1), 5–22. https://doi.org/10.1016/S0738-3991(00)00149-X

Elwyn, G., Hutchings, H., Edwards, A., Rapport, F., Wensing, M., Cheung, W.-Y., & Grol, R. (2005). The OPTION scale: Measuring the extent that clinicians involve patients in decision-making tasks. *Health Expectations*, *8*(1), 34–42. https://doi.org/10.1111/j.1369-7625.2004.00311.x

Elwyn, G., Tsulukidze, M., Edwards, A., Légaré, F., & Newcombe, R. (2013). Using a 'talk' model of shared decision making to propose an observation-based measure: Observer OPTION5 Item. *Patient Education and Counseling*, *93*(2), 265-271. https://doi.org/10.1016/j.pec.2013.08.005

Erikson, E. H. (1994). Identity and the life cycle. WW Norton.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fjermestad, K., McLeod, B. D., Tully, C. B., & Liber, J. M. (2016). *Therapist Characteristics and Interventions* (S. Maltzman, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199739134.013.11

Fjermestad, K. W., Føreland, Ø., Oppedal, S. B., Sørensen, J. S., Vognild, Y. H., Gjestad, R., Öst, L.-G., Bjaastad, J. F., Shirk, S. S., & Wergeland, G. J. (2020). Therapist Alliance-Building Behaviors, Alliance, and Outcomes in Cognitive Behavioral Treatment for Youth Anxiety Disorders. *Journal of Clinical Child & Adolescent Psychology*, 1–14. https://doi.org/10.1080/15374416.2019.1683850

Fjermestad, K. W., McLeod, B. D., Heiervang, E. R., Havik, O. E., Öst, L.-G., & Haugland, B. S. M. (2012). Factor Structure and Validity of the Therapy Process Observational Coding System for Child Psychotherapy–Alliance Scale. *Journal of Clinical Child & Adolescent Psychology*, *41*(2), 246–254. https://doi.org/10.1080/15374416.2012.651999

Fleiss, J. L. (1981). Balanced Incomplete Block Designs for Inter-Rater Reliability Studies. *Applied Psychological Measurement*, *5*(1), 105–112. https://doi.org/10.1177/014662168100500115

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*(3), 248–260. https://doi.org/10.1037/1040-3590.7.3.248

Fryers, T., & Brugha, T. (2013). Childhood Determinants of Adult Psychiatric Disorder. *Clinical*

Practice & Epidemiology in Mental Health, *9*(1), 1–50. https://doi.org/10.2174/1745017901309010001

Fuller, S. M., Koester, K. A., Guinness, R. R., & Steward, W. T. (2017). Patients' Perceptions and Experiences of Shared Decision-Making in Primary HIV Care Clinics. *Journal of the Association of Nurses in AIDS Care*, *28*(1), 75–84. https://doi.org/10.1016/j.jana.2016.08.010

Gärtner, F. R., Bomhof-Roordink, H., Smith, I. P., Scholl, I., Stiggelbout, A. M., & Pieterse, A. H. (2018). The quality of instruments to assess the process of shared decision making: A systematic review. *PLOS ONE*, *13*(2), e0191747. https://doi.org/10.1371/journal.pone.0191747

George, D., & Mallery, P. (2016). Descriptive statistics. In *IBM SPSS Statistics 23 Step by Step* (pp. 126-134). Routledge.

Grim, K., Rosenberg, D., Svedberg, P., & Schön, U.-K. (2016). Shared decision-making in mental health care—A user perspective on decisional needs in community-based services. *International Journal of Qualitative Studies on Health and Well-Being*, *11*(1), 30563. https://doi.org/10.3402/qhw.v11.30563

Haine-Schlagel, R., & Walsh, N. E. (2015). A review of parent participation engagement in child and family mental health treatment. *Clinical Child and Family Psychology Review*, *18*(2), 133–150. https://doi.org/10.1007/s10567-015-0182-x

Hale, K. L., Wallace, D. D., Blanco-Durán, D., Annis, I. E., Guzman, L. E., García, S. J., Pérez Jolles, M., Sleath, B. L., Elwyn, G., Stein, G. L., & Thomas, K. C. (2020). Conversations between Latina mothers and their child's mental health provider: An observational study

of shared decision-making regarding pediatric patient mental health needs. *Patient Education and Counseling*, *103*(1), 96–102. https://doi.org/10.1016/j.pec.2019.08.013

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23. https://doi.org/10.20982/tqmp.08.1.p023

Hawley, K. M., & Weisz, J. R. (2005). Youth Versus Parent Working Alliance in Usual Clinical Care: Distinctive Associations With Retention, Satisfaction, and Treatment Outcome. *Journal of Clinical Child & Adolescent Psychology*, *34*(1), 117–128. https://doi.org/10.1207/s15374424jccp3401_11

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. John Wiley & Sons.

Hedeker, D., Gibbons, R. D., & Flay, B. R. (1994). *Random-effects regression models for clustered data with an example from smoking prevention research. 62*(4), 757–765.

Heisler, M., Vijan, S., Anderson, R. M., Ubel, P. A., Bernstein, S. J., & Hofer, T. P. (2003). When do patients and their physicians agree on diabetes treatment goals and strategies, and what difference does it make? *Journal of General Internal Medicine*, *18*(11), 893–902. https://doi.org/10.1046/j.1525-1497.2003.21132.x

Hofstra, M. B., Van Der Ende, J., & Verhulst, F. C. (2002). Child and Adolescent Problems Predict DSM-IV Disorders in Adulthood: A 14-Year Follow-up of a Dutch Epidemiological Sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(2), 182–189. https://doi.org/10.1097/00004583-200202000-00012

Hogue, A. (2002). Adherence process research on developmental interventions: Filling in the middle. *New Directions for Child and Adolescent Development*, *2002*(98), 67–74. https://doi.org/10.1002/cd.59

Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C., Inclan, J., Reiner, R. H., & Liddle, H. A. (2008). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment*, *35*(2), 137–147. https://doi.org/10.1016/j.jsat.2007.09.002

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. https://doi.org/10.1007/s10488-014-0548-2

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. https://doi.org/10.1037/0033-3204.33.2.332

IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

Karver, M. S., De Nadai, A. S., Monahan, M., & Shirk, S. R. (2018). Meta-analysis of the prospective relation between alliance and outcome in child and adolescent psychotherapy. *Psychotherapy*, *55*(4), 341-355. https://doi.org/10.1037/pst0000176

Kazdin, A. E. (2016). *Research Design in Clinical Psychology* (5th ed.). Pearson.

Kendall, P. C. (1994). Treating anxiety disorders in children: results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *62*(1), 100-110. https://doi.org/10.1037//0022-006x.62.1.100

Kendall, P. C., & Hedtke, K. (2006). *Cognitive-behavioral therapy for anxious children: Therapist Manual.* (3rd ed.). Workbook Publishing.

Kendall, P. C., Hudson, J. L., Gosch, E., Flannery-Schroeder, E., & Suveg, C. (2008). Cognitive-behavioral therapy for anxiety disordered youth: A randomized clinical trial evaluating child and family modalities. *Journal of Consulting and Clinical Psychology*, *76*(2), 282–297. https://doi.org/10.1037/0022-006X.76.2.282

Kline, P. (1979). *Psychometrics and psychology*. Academic Press.

Knapp, C., Huang, I. C., Madden, V., Vadaparampil, S., Quinn, G., & Shenkman, E. (2009). An evaluation of two decision-making scales for children with life-limiting illnesses. *Palliative Medicine*, *23*(6), 518-525. https://doi.org/10.1177/0269216309104892

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In *Handbook of psychotherapy and behavior change* (pp. 72–113). John Wiley & Sons.

Langer, D. A., Holly, L. E., Wills, C. E., Tompson, M. C., & Chorpita, B. F. (2022). Shared decision-making for youth psychotherapy: A preliminary randomized clinical trial on facilitating personalized treatment. *Journal of Consulting and Clinical Psychology*, *90*(1), 29–38. https://doi.org/10.1037/ccp0000702

Langer, D. A., & Jensen-Doss, A. (2018). Shared decision-Making in youth mental health care: Using the evidence to plan treatments collaboratively. *Journal of Clinical Child & Adolescent Psychology*, *47*(5), 821–831. https://doi.org/10.1080/15374416.2016.1247358

Liber, J. M., McLeod, B. D., Van Widenfelt, B. M., Goedhart, A. W., van der Leeden, A. J. M., Utens, E. M. W. J., & Treffers, P. D. A. (2010). Examining the relation between the therapeutic alliance, treatment adherence, and outcome of cognitive behavioral therapy for children with anxiety disorders. *Behavior Therapy*, *41*(2), 172–186. https://doi.org/10.1016/j.beth.2009.02.003

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202.

Liverpool, S., Pereira, B., Hayes, D., Wolpert, M., & Edbrooke-Childs, J. (2020). A scoping review and assessment of essential elements of shared decision-making of parent-involved interventions in child and adolescent mental health. *European Child & Adolescent Psychiatry*, *30*, 1319–1338. https://doi.org/10.1007/s00787-020-01530-7

Loh, A., Leonhart, R., Wills, C. E., Simon, D., & Härter, M. (2007). The impact of patient participation on adherence and clinical outcome in primary care of depression. *Patient Education and Counseling*, *65*(1), 69–78. https://doi.org/10.1016/j.pec.2006.05.007

Makoul, G., & Clayman, M. L. (2006). An integrative model of shared decision making in medical encounters. *Patient Education and Counseling*, *60*(3), 301-312. https://doi.org/10.1016/j.pec.2005.06.010

Mandelblatt, J., Kreling, B., Figeuriedo, M., & Feng, S. (2006). What Is the Impact of Shared Decision Making on Treatment and Outcomes for Older Women With Breast Cancer? *Journal of Clinical Oncology*, *24*(30), 4908–4913. https://doi.org/10.1200/JCO.2006.07.1159

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. https://doi.org/10.1023/a:1022608117322

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. https://doi.org/10.1037/pas0000037

Marques, L., Valentine, S. E., Kaysen, D., Mackintosh, M.-A., Dixon De Silva, L. E., Ahles, E. M., Youn, S. J., Shtasel, D. L., Simon, N. M., & Wiltsey-Stirman, S. (2019). Provider fidelity and modifications to cognitive processing therapy in a diverse community health clinic: Associations with clinical change. *Journal of Consulting and Clinical Psychology*, *87*(4), 357–369. https://doi.org/10.1037/ccp0000384

McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review*, *31*(4), 603–616. https://doi.org/10.1016/j.cpr.2011.02.001

McLeod, B. D., Islam, N., & Wheat, E. (2013). Designing, conducting, and evaluating therapy process research. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 142–164). Oxford University Press.

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. https://doi.org/10.1037/pas0000037

McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Hogue, A., Kendall, P. C., & Weisz, J. R. (2019). Benchmarking treatment adherence and therapist competence in individual cognitive-behavioral treatment for youth anxiety disorders. *Journal of Clinical Child & Adolescent Psychology*, *48*(sup1), S234–S246. https://doi.org/10.1080/15374416.2017.1381914

McLeod, B. D., Southam-Gerow, M. A., & Kendall, P. C. (2017). Observer, youth, and therapist perspectives on the alliance in cognitive behavioral treatment for youth anxiety. *Psychological Assessment*, *29*(12), 1550–1555. https://doi.org/10.1037/pas0000465

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. https://doi.org/10.1080/15374416.2016.1253018

McLeod, B. D., & Weisz, J. R. (2005). The Therapy Process Observational Coding System-Alliance Scale: Measure characteristics and prediction of outcome in usual clinical practice. *Journal of Consulting and Clinical Psychology*, *73*(2), 323–333. https://doi.org/10.1037/0022-006X.73.2.323

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. https://doi.org/10.1080/15374411003691750

McNeish, D., & Stapleton, L. M. (2016). Modeling Clustered Data with Very Few Clusters. *Multivariate Behavioral Research*, *51*(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change* (3rd ed.). Guilford Press.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*(4907), 933-938.

NAMI. (2019, September). *Mental Health By the Numbers*. https://www.nami.org/mhstats

Nicolas, G., & Schwartz, B. (2012). *Culture first: Lessons learned about the importance of the*

cultural adaptation of cognitive behavior treatment interventions for Black Caribbean youth.

Normand, S.-L. T., McNeil, B. J., Peterson, L. E., & Palmer, R. H. (1998). Eliciting expert opinion using the Delphi technique: Identifying performance indicators for cardiovascular disease. *International Journal for Quality in Health Care*, *10*(3), 247–260. https://doi.org/10.1093/intqhc/10.3.247

Norton, E. C., Bieler, G. S., Ennett, S. T., & Zarkin, G. A. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations. *Journal of Consulting and Clinical Psychology*, *64*(5), 919-926. https://doi.org/10.1037/0022-006X.64.5.919

Nurmi, J. E. (1991). How do adolescents see their future? A review of the development of future orientation and planning. *Developmental Review*, *11*(1), 1-59.

Okoli, C., Brough, G., Allan, B., Castellanos, E., Young, B., Eremin, A., Corbelli, G. M., Mc Britton, M., Muchenje, M., Van de Velde, N., & de los Rios, P. (2020). Shared decision making between patients and healthcare providers and its association with favorable health outcomes among people living with HIV. *AIDS and Behavior*, *25*(5), 1384-1395. https://doi.org/10.1007/s10461-020-02973-4

Piaget, J. (1964). Cognitive Development in Children: Development and Learning. *Journal of Research in Science Teaching*, *2*, 176-186.

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. https://doi.org/10.1037/0022-006X.75.6.829

Rousson, V., Gasser, T., & Seifert, B. (2003). Confidence intervals for intraclass correlation in inter-rater reliability. *Scandinavian Journal of Statistics*, *30*(3), 617-624.

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

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

Schiffman, J., Becker, K. D., & Daleiden, E. L. (2006). Evidence-based services in a statewide public mental health system: Do the services fit the problems? *Journal of Clinical Child & Adolescent Psychology*, *35*(1), 13–19. https://doi.org/10.1207/s15374424jccp3501_2

Schrager, S., Ovsepyan, V., & Burnside, E. (2020). Breast cancer screening in older women: The importance of shared decision making. *The Journal of the American Board of Family Medicine*, *33*(3), 473–480. https://doi.org/10.3122/jabfm.2020.03.190380

Sepucha, K. R., Atlas, S. J., Chang, Y., Freiberg, A., Malchau, H., Mangla, M., Rubash, H., Simmons, L. H., & Cha, T. (2018). Informed, patient-centered decisions associated with better health outcomes in orthopedics: Prospective cohort study. *Medical Decision Making*, *38*(8), 1018–1026. https://doi.org/10.1177/0272989X18801308

Shay, L. A., & Lafata, J. E. (2015). Where is the evidence? A systematic review of shared decision making and patient outcomes. *Medical Decision Making*, *35*(1), 114–131. https://doi.org/10.1177/0272989X14551638

Shirk, S. R., & Saiz, C. C. (1992). Clinical, empirical, and developmental perspectives on the

therapeutic relationship in child psychotherapy. *Development and Psychopathology*, *4*(4), 713-728. https://doi.org/10.1017/S0954579400004946

Shrout, P. E., & Fleiss, J. L. (1979). *Intraclass correlations: Uses in assessing rater reliability*. *86*(2), 420–428. https://doi.org/10.1037//0033-2909.86.2.420

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: should Fisher's z transformation be used?. *Journal of Applied Psychology*, *72*(1), 146-148.

Simmons, M. B., Batchelor, S., Dimopoulos-Bick, T., & Howe, D. (2017). The Choice Project: Peer workers promoting shared decision making at a youth mental health service. *Psychiatric Services*, *68*(8), 764–770. https://doi.org/10.1176/appi.ps.201600388

Simon, D., Schorr, G., Wirtz, M., Vodermaier, A., Caspari, C., Neuner, B., Spies, C., Krones, T., Keller, H., Edwards, A., Loh, A., & Härter, M. (2006). Development and first validation of the shared decision-making questionnaire (SDM-Q). *Patient Education and Counseling*, *63*(3), 319–327. https://doi.org/10.1016/j.pec.2006.04.012

Slade, M. (2017). Implementing shared decision making in routine mental health care: Forum – shared decision making in mental health care. *World Psychiatry*, *16*(2), 146–153. https://doi.org/10.1002/wps.20412

Southam-Gerow, M. A., McLeod, B. D., Arnold, C. C., Rodríguez, A., Cox, J. R., Reise, S. P., Bonifay, W. E., Weisz, J. R., & Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive–behavioral therapy for child anxiety. *Psychological Assessment*, *28*(1), 70–80. https://doi.org/10.1037/pas0000141

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. https://doi.org/10.1016/j.jaac.2010.06.009

Stiggelbout, A. M., Van der Weijden, T., De Wit, M. P., Frosch, D., Légaré, F., Montori, V. M., ... & Elwyn, G. (2012). Shared decision making: Really putting patients at the centre of healthcare. *BMJ*, *344*. https://doi.org/10.1136/bmj.e256

Stirman Wiltsey, S. W., Miller, C. J., Toder, K., & Calloway, A. (2013). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, *8*(1), 65-77. https://doi.org/10.1186/1748-5908-8-65

Stacey, D., Taljaard, M., Drake, E. R., & O'Connor, A. M. (2008). Audit and feedback using the brief Decision Support Analysis Tool (DSAT-10) to evaluate nurse–standardized patient encounters. *Patient Education and Counseling*, *73*(3), 519-525.

Stirman Wiltsey, S., A Gutner, C., Crits-Christoph, P., Edmunds, J., Evans, A. C., & Beidas, R. S. (2015). Relationships between clinician-level attributes and fidelity-consistent and fidelity-inconsistent modifications to an evidence-based psychotherapy. *Implementation Science*, *10*(1), 115. https://doi.org/10.1186/s13012-015-0308-z

Street, R. L., Makoul, G., Arora, N. K., & Epstein, R. M. (2009). How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Patient Education and Counseling*, *74*(3), 295–301. https://doi.org/10.1016/j.pec.2008.11.015

Sutherland, K. S., McLeod, B. D., Conroy, M. A., & Cox, J. R. (2013). Measuring implementation of evidence-based programs targeting young children at risk for emotional/behavioral disorders: Conceptual issues and recommendations. *Journal of Early Intervention*, *35*(2), 129–149. https://doi.org/10.1177/1053815113515025

Tan, S. H., & Tan, S. B. (2010). The correct interpretation of confidence intervals. *Proceedings of Singapore Healthcare*, *19*(3), 276-278.

Ware, N. C., Tugenberg, T., & Dickey, B. (2003). Ethnography and measurement in mental health: Qualitative validation of a measure of continuity of care (CONNECT). *Qualitative Health Research*, *13*(10), 1393–1406. https://doi.org/10.1177/1049732303258316

Weisz, J. R. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*, *69*(3), 274. https://doi.org/10.1001/archgenpsychiatry.2011.147

Weisz, J. R., & Hawley, K. M. (2002). Developmental factors in the treatment on adolescents. *Journal of Consulting and Clinical Psychology*, *70*(1), 21–43. https://doi.org/10.1037/0022-006X.70.1.21

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: Multilevel meta-analysis. *JAMA Psychiatry*, *70*(7), 750. https://doi.org/10.1001/jamapsychiatry.2013.1176

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., Jensen-Doss, A., Hawley, K. M., Krumholz Marchette, L. S., Chu, B. C., Weersing, V. R., & Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist*, *72*(2), 79–117. https://doi.org/10.1037/a0040360

Weisz, J. R., Weersing, V. R., Valeri, S. M., & McCarty, C. A. (1999). *Therapist's Manual*

PASCET: Primary and secondary control enhancement training program.

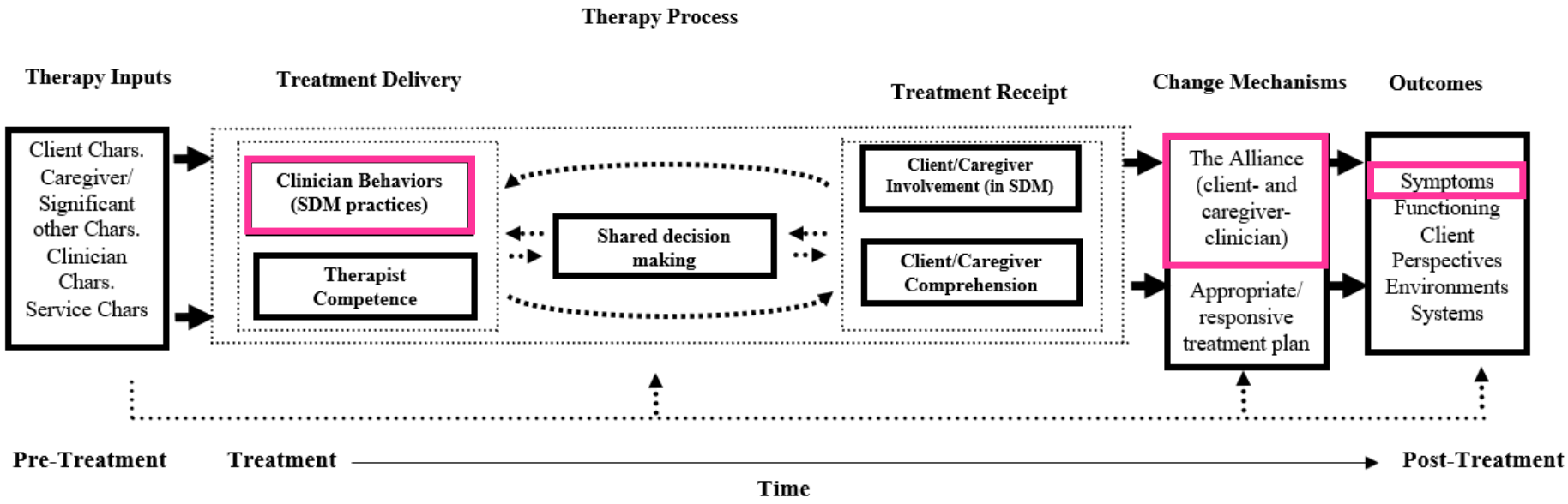
Whitley, R. (2009). The implications of race and ethnicity for shared decision-making. *Psychiatric Rehabilitation Journal*, *32*(3), 227-230. https://doi.org/10.2975/32.3.2009.227.230

Zucker, D. M. (1990). An analysis of variance pitfall: The fixed effects analysis in a nested design. *Educational and Psychological Measurement*, *50*(4), 731-738.

Appendix

Figure 1

Proposed Pathway of SDM Practices' Effect on Clinical Outcomes.



Note. *Treatment delivery* refers to clinicians' actions and skills involved in implementing interventions. *Therapy process* refers to interactives processes between clinicians and clients that occur during psychotherapy sessions. *Treatment receipt* refers to actions and comprehension of the clients and caregivers throughout treatment. The SDM-YMH is designed to capture one contributing factor of SDM (i.e., clinician behaviors that promote SDM—*SDM practices*). SDM is a collaborative process that involves both clinician and client/caregiver input. However, the current study and measure focus solely on clinician behaviors. SDM practices are posited to increase clinical outcomes for youth receiving psychotherapy via the alliance.

Table A

Demographic Data for Included Studies

First Author	Year	Sample Demographic Information	Tx Type	Provider Information	Setting	Location	Primary Presenting Problem
Alegría	2018	N = 312; Age range: 18-80 (*M*=49); 35.9% White, 10.9% Black, 42% Latinx, 11.2% Asian; 32.1% male; *M* income: $20k	MH	MH providers, nurses	13 outpatient behavioral health clinics	Boston, MA	Varied (e.g., anxiety, depression)
Aoki	2019	N=88; Age range: 20+ (*M*=22); Racial/ethnic identities of participants: U; 55% male; *M* income: U	MH	MH providers, nurses	Waseda University counseling clinic	Tokyo, Japan	Major Depressive Disorder
Hamann	2006	N=107; Age range: 18-65 (*M*=37); Racial/ethnic identities of participants: U; 52% male: *M* income: U	MH	Nurses/Physicians	12 acute psychiatric wards	Munich, Germany	Schizophrenia or schizophreniform
Hamann	2020	N=161; Age range: 18-65 (*M*=42); Racial/ethnic identities of participants: U; 50% male; *M* income: U	MH	MH providers, physicians, nurses	12 acute psychiatric wards	Germany	Schizophrenia spectrum disorders
Ishii	2017	N=24; Age range: 16-65 (*M*=39); Racial/ethnic identities of participants: U; 68% male: *M* income: U	MH	MH providers, physicians, nurses	Acute psychiatric ward of Numazu Chuo Hospital	Shizuoka, Japan	Schizophrenia spectrum disorder
Joosten	2009	N=220; Age range: 29-51 (*M*=40); 93% Dutch; 71% male; *M* income: U	SU	Social workers, nurses	Addiction treatment centers	Netherlands	Substance use
Joosten	2011	N=212; Age range: 32-52 (*M*=42); 97% Dutch; 70% male; *M* income: U	SU	Social workers, nurses	Addiction treatment centers	Netherlands	Substance use
Langer	2022	N = 40; 67% White; Age range: 7-15 (*M*=10.95); 55% female; Household income: 80% <$60,000	MH	MH providers	Specialty clinic	USA	Anxiety
Lovell	2018	N=497; Age range: 18-65+ (*M*=45); 86% White, 14% U; 39% male; *M* income: U	MH	MH providers, nurses	Community MH clinics	England	Severe mental illness (e.g., psychosis)
Metz	2018	N=200; Age range: unknown (*M*=38); Racial/ethnic identities of participants: U; 34% male; *M* income: U	MH	MH providers	Outpatient MH clinics	Netherlands	Anxiety, depression, personality disorders
Metz	2019	N=186; Age range: 18-83 (*M*=47.2); Racial/ethnic identities of participants: U; 40.3% male; *M* income: U	MH	MH providers	Specialist mental healthcare	Netherlands	Varied
Mott	2014	N=27; Age range: 22-47 (*M*=29.3); 70% White, 30% U; 85% male; *M*income: $27k	MH	MH providers, Trainees	VA PTSD clinic	USA	PTSD
Westermann	2013	N=94; Age range: 2-12 (*M*=6.5); Racial/ethnic identities of participants: U; 30% male; *M* income: U	MH	MH Providers	Two centers for youth MH care	South Netherlands	Varied (e.g., ADHD, GAD)

Yama-guchi	2017	N=56; Age range: 20+ (*M*=38.5); Racial/ethnic identities of participants: U; 60% male; *M* income: U	MH	Paraprofessionals, MH providers, physicians	Outpatient psychiatric clinic & psychiatric hospital	Tokyo, Japan	Varied (e.g., depression,)

Note. Tx = Treatment; MH = mental health; SU = substance use; U= Unknown

Table B

Study Information

First Author	Year	RoB 2 Score	Power Analysis	Adherence to SDM Assessed?	Measurement of Outcome Variables & Psychometric Properties
Alegría	2018	Low risk of bias	N = 300 based on power of 80% (Cohen *d* = .30) to 90% (*d* = .35)	Yes (observer-rated)	Observer-rated SDM (ICC = .53) Patient-report SDM (no psychometrics) Provider-report SDM (no psychometrics)
Aoki	2019	Low risk of bias	Considering a significance level of 5% (2-sided) and β = .8, n = 27 per arm was needed to detect a mean difference of 15	No	Patient-report SDM (no psychometrics) Depression symptom severity (no psychometrics)
Hamann	2006	Low risk of bias	None included.	No	Patient-report SDM (no psychometrics) Provider-reported alliance (no psychometrics)
Hamann	2020	Low risk of bias	Considering a significance level of 5% (2-sided) and β = .8, n = 23 per arm (12 arms; N = 276) was needed to detect a mean difference of 15	Yes (supervision)	Patient-reported SDM (no psychometrics) Patient-reported alliance (no psychometrics) Provider-reported alliance (no psychometrics)
Ishii	2017	Low risk of bias	None included. Reported that the trial did not have adequate statistical power to detect a difference in outcomes.	Yes (supervision)	Patient-report satisfaction (no psychometrics) Patient-report of attitude towards drugs (no psychometrics) Treatment continuation (review of records)
Joosten	2009	Some concerns; randomization based first availability; more patients missing from tx condition at follow-up	Based on β = .8, to detect a significant difference (*p* = .05; 2-sided), n = 77 were needed for each study condition. Considering 20% dropout, n = 97 per condition was needed.	Yes (self-report from providers delivering intervention)	Patient-report symptom severity (no psychometrics) Patient-report symptom-drug use (no psychometrics) Patient-report quality of life (no psychometrics)
Joosten	2011	Some concerns; outcome measures do not map onto SDM definition	None included. Reported that inconclusive results were due to small sample size.	Yes (self-report from providers delivering intervention)	Patient-report interpersonal behavior between provider and patient (α = .83) Provider-report interpersonal behavior between provider and patient (α = .83)
Langer	2022	Low risk of bias	Underpowered	Yes	
Lovell	2018	Low risk of bias	Considering a significance level of 5% (2-sided) and β = .8, N = 480 (n = 20 per arm) was needed to detect an effect size of .4	Yes (supervision)	Patient-report of health care climate (no psychometrics) Patient-report involvement (no psychometrics) Patient-report alliance and engagement (no psychometrics)
Metz	2018	Low risk of bias	Considering a significance	Yes (supervision)	Patient-report of decisional conflict (α = .95)

			level of 5% (2-sided) and β = .8, n= 65 patients per arm was needed (N = 130).		Patient-report of item: `to what extent do you agree with the decision taken?' (no psychometrics) Provider-report of item: `to what extent do you agree with the decision taken?' (no psychometrics) Patient-report of Participation (α = .90) Patient-report of SDM (α = .95) Provider-report of SDM (α = .86) Patient-report of working alliance (α = .97)
Metz	2019	Low risk of bias	Considering a significance level of 5% (2-sided) and β = .8, n = 65 per arm was needed to detect an effect size of *d* = .5. After correcting for clustering, intra-cluster correlation coefficient of .05 was used; n = 136 per arm was needed (N = 272).	Yes (supervision)	Patient-report of decisional conflict (no psychometrics) Patient-report of item: 'to what extent do you agree with the decision taken?') (no psychometrics) Provider-report of item: 'to what extent do you agree with the decision taken?' (no psychometrics) Patient-report of working alliance (α = .93) Patient-report of SDM (α = .91) Provider-report of SDM (α = .91)
Mott	2014	Some concerns; follow up data from n=5; inappropriate method of assessing outcome	None included. Reported that study was underpowered.	No	Treatment engagement (measured attendance via chart review) (k = .68 – 1)
Westermann	2013	Low risk of bias	Considering a significance level of 5% (2-sided) and β = .8, n = 37 per arm as needed (N = 74).	No	Caregiver-report of decisional conflict (no psychometrics) Caregiver-report of acceptance of decision (no psychometrics)
Yamaguchi	2017	Low risk of bias	None included. Reported that study was underpowered.	Yes (observer-rated)	Observer-rated SDM (k during training = .85) Provider-report of alliance (α =.9) Patient-report of alliance (α = .85) Clinical outcome of drug symptoms (no psychometrics)

Note. RoB 2 score = Risk of Bias Tool Version 2; Tx = treatment

Table C

SDM Practices and their Demonstrated Effectiveness

First Author	Year	Provider-Delivered SDM Practices	Source of Practices	Link to Outcomes
Alegría	2018	1. Works to establish a shared agenda with patient 2. Provider sticks to agenda while still being flexible to patient 3. Elicits patient's goals for the visit 4. Involves the patient in decisions 5. Involves patient in discussion of treatment options/possibilities 6. Explores patient's acceptance of possible treatment(s) 7. Involves the patient in explicitly made treatment decisions 8. Patient and provider agree in problem formulation 9. Elicits patient's understanding of the problem/symptoms 10. Uses patient's frame of reference when explaining problems 11. Gives hope to patient about recovery 12. Offers explanations about treatment possibilities 13. Explores barriers about treatment possibilities 14. Explores the patient's background in non-judgmental manner 15. Uses situational and external explanations to describe the patient's behavior or symptoms (i.e., not putting blame on patient) 16. Collects enough information about symptoms 17. Listens attentively 18. Gives patient time and space to present the problem 19. Is attentive and respectful when changing subjects 20. Provides illness management education 21. Helps patient to formulate questions 22. Offers Praise 23. Creates an open, inviting, and non-judgmental atmosphere	Codes used to assess adherence; described in online supplemental materials (Alegría et al., 2018b)	Based on intention to treat analyses, the intervention significantly increased observer-rated SDM (d = .29), but not patient- or provider-rated SDM.
Aoki	2019	1. Informs patient of diagnosis and treatment options (with pros and cons, chosen for patient's situation and lifestyle) 2. Gives patient decision aid booklet, comprising general information about mood disorders and treatment options. 3. Discusses the treatment options (using decision aid booklet, answering questions, and encouraging the patient to state opinions regarding treatment options.) 4. Clarifies the patient's understanding and starts discussions on topics that depended on the patient's understanding, 5. Discusses treatment options and decides on the treatment, which is in line with the preferences indicated by the patient's values	Primary article's text; study protocol in English unavailable	SDM intervention significantly increased patient-reported SDM (no effect size reported), but not clinical outcomes.
Hamann	2006	1. Decision aid covers the pros and cons of treatment options and psychoeducation	Primary article's text;	Intervention led to greater perceived

		2. Helps patients through these books and asks patients to write down their experiences with previous medication and to indicate preferences regarding different treatment options 3. Nurses answer any questions of the patients and encouraged them to state any point of view contrary to that of the doctor 4. Doctor and patient reach agreement on further treatment according to the preferences indicated by the patient in the booklet	study protocol unavailable	involvement in medical decisions early in treatment; this difference was not present at time of discharge. No differences in alliance were found. (No effect sizes reported).
Hamann	2020	1. Identifies current situation: (a) "life or death", (b) preference-sensitive decisions, or (c) "best choice" decisions 2. Determines which SDM strategy to use: (a) "classical way" (e.g., describing options, communicating risks and benefits, and identifying patient preferences) for preference-sensitive decisions or (b) Harvard negotiation model (i.e., a method for negotiating mutually satisfactory agreements) 3. Negotiation process includes strategies such as: (a) taking time-outs, (b) identifying one's own interests (that is, the option that is seen as the best choice by the psychiatrist), (c) developing an alternative plan (identifying other potential solutions if the patient does not accept the best-choice option) 4. Uses MI throughout to avoid conflict 5. Assists patients to identify or develop their individual preferences 6. Creates a participatory atmosphere to help motivate, empower, and enable patients to participate in SDM 7. Provides psychoeducation to increase self-esteem and health literacy	Study protocol (Hamann et al., 2017)	Intervention led to increased patient-reported SDM and alliance scores. (No effect sizes reported).
Ishii	2017	1. Patient and at least three ward staff members discuss questionnaire results of patients' perceptions of treatment 2. Creates a comfortable atmosphere both for the patient and staff members 3. Patients and providers draft the care plan sheet in order to outline clearly what they have shared in the session.	Study protocol (Ishii et al., 2014)	No differences between SDM and usual care group found.
Joosten	2009	1. MI is offered in a structured way to explore and compare indicated treatment goals and to reach an agreement on these goals. 2. Both patient and clinician discuss the results of the Goals of Treatment questionnaire to reach a treatment agreement 3. Translates areas of problems into goals 4. Patients and providers write goals on cards; arrange by level of importance 5. Patient's treatment goals and expectations are explored and compared to the clinician's perception 6. Discuss agreements/differences between the clinician's and patient's perceptions 7. Based on this discussion, the treatment contract is completed	Primary article's text; study protocol unavailable	Intervention led to decreased psychiatric problems (d = .38) and drug use (d = .42). No differences in patient-reported quality of life, abstinence, primary substance use or substance dependence were found.

		8. Halfway through the treatment, goals and expectations are discussed and adapted based on patient progress		
Joosten	2011	Same practices as Joosten et al. (2009)	Primary article's text; study protocol unavailable	Intervention led to increased patient-reported autonomy, control, and extraverted behavior. (No effect sizes reported).
Langer	2022	1 . Introduce SDM 2. Practice collaborative decision making 3. Select treatment targets 4. Discuss treatment-related values 5. Introduce the evidence 6. Select treatment participants 7. Select treatment components 8. Plan symptom tracking and follow-up	Primary article's text	SDM intervention led to significantly greater involvement in the treatment planning for youth and caregivers. Caregivers in the SDM condition reported lower decisional conflict and decisional regret. No significant differences between conditions on treatment length, satisfaction with decisions, or engagement. No significant differences in clinical outcomes.
Lovell	2018	No discernible practices described	Study protocol (Bower et al., 2015)	No differences between SDM and usual care group.
Metz	2018	No discernible practices described	Study protocol (Metz et al., 2017)	Based on intention to treat analyses, intervention led to increase patient-reported SDM (d = .32), but not decreased decisional conflict.
Metz	2019	1. Establishes expectations about shared process 2. Discusses which role the patient desired in decision making	Study protocol (Metz et al., 2015)	Intervention led to increased provider-reported alliance (d =

		3. Connects with patient's wishes and goals. 'What does he/she want to achieve in treatment?' 4. Explains about Routine Outcome Monitoring (ROM) as an information source 5. Discusses options, advantages, and disadvantages, in a neutral manner 6. Weigh options: Weighs advantages and disadvantages 7. Shared Decision; Select together most appropriate option		.45) and agreement on decision (d = .45). No differences were found for patient-reported outcomes.
Mott	2014	1. "Choice talk": provider indicates that a choice exists and that the patient can have a role in treatment decisions 2. "Option talk": provider gives detailed information about benefits/risks, mechanisms, and effectiveness of treatments using decision-support tools (i.e., Patient Decision Aid or PDA) 3. "Decision talk": during which the patient and provider dialogue about preferences, eventually eliciting a decision 4. Treatment information consists of nontechnical, factual statements that described the intervention name, purpose, components, mechanisms, effectiveness, potential discomforts, client/therapist roles, and frequency 5. The PDA includes a comparison chart that summarizes the central aspects of each featured treatment and briefly describes alternative treatments, inviting patients to request further details	Primary article's text; study protocol unavailable	No differences in engagement were found, but intervention led to more patients choosing an EBP.
Wester-mann	2013	1. Summarizes information from the diagnostic phase 2. Describes treatment options (e.g., pros and cons, targets) 3. Attunes to patients' preferences, encourages partnership/joint visions/choices 4. Parents invited to note items they want to discuss 5. Uses retrospection to attune to parents and empower them, based on the parents' opinion of the referral, their earlier experiences with (mental) health care, the extent to which they wish to be informed, their preferred role in decision-making and points of interest 6. Uses this information, to establish a natural connection with their expectations and the efforts the parents and child have made during assessment in order to reduce stress and enhance cooperation 7. Dialogues with the parents about diagnostic formulation and treatment options 8. In everyday language an understandable and meaningful joint narrative is created 9. Schematically visualizes the narrative upon a whiteboard 10. Positive aspects of the child and the environment are included to broaden and balance the view and to use them later on as points of therapeutic application 11. Parents are invited to react 12. Combining professional ideas and parental conceptions a meaningful and clear picture emerges 13. After consensus is reached upon this strengths-difficulties synopsis, the treatment options are discussed and illustrated	Primary article's text; study protocol unavailable	Intervention led to less caregiver-reported decisional conflict (no effect sizes reported).

		14. The evidence, aims, uncertainties, benefits, or possible harms of the options proposed are pointed out 15. This dialogue helps to create a treatment plan 16. (Dis)agreements upon the strengths/limitation's analysis and recommended treatment are noted 17. Arrangements are made of who will inform the child (if not present)		
Yama-guchi	2017	1. Software program helps patients summarize recovery goals and personal strengths and then convey that information to their doctors 2. Peer support specialists in the decision support center help patients use software and share their personal recovery experiences 3. Doctors then proceed with their medical consultation according to the participant's condition and concerns 4. Doctors ask questions to discuss treatment/self-management behaviors 5. Determine treatment based on patients' preference and discussion 6. Patient and doctor determine treatment (e.g., medication type and timing/use of medication) and/or self-management behaviors for the next consultation	Online appendix (Yamaguchi et al., 2017b)	No differences between SDM and usual care group found.