ORIGINAL PAPER

Partnering for Success Symptom Monitoring Outcomes: Cognitive Behavioral Therapy Plus for Children and Youth in Child Welfare

Suzanne E. U. Kerns [1,5] · Jon D. Phillips [2] · Lucy Berliner [3] · Richard P. Barth [4]

Abstract
A substantial proportion of children and youth in the child welfare system have mental health concerns that warrant attention. While these youth are more likely to receive treatment in general, they are less likely to receive evidence-based treatments. Cognitive Behavioral Therapy Plus (CBT+) is a training approach in evidence-based treatments for anxiety, depression, behavior problems, and symptoms of traumatic stress. Included in the model is an emphasis on symptom and outcome monitoring. Though based on well-established evidence-based treatments, clinical effectiveness of CBT+ has not been evaluated for child welfare populations. Embedded within a broader initiative called Partnering for Success, the current study evaluated the impact of CBT+ training on treatment outcomes for the mental health symptoms of a racially diverse sample of 1281 children and youth in four different states. Symptom data were collected by community-based mental health therapists as part of standard CBT+ treatment adherence procedures. Multilevel linear growth modeling was used to evaluate symptom change over time. Significant improvements were observed across all four clinical targets, with a slight curvilinear relationship found for anxiety, depression, and behavior problems. This promising initial evidence suggests CBT+ is a viable training option for treatment of child welfare-involved children and youth with a range of mental health concerns.

Keywords Symptom monitoring · CBT · Child welfare · Mental health · Consultation

Highlights
- CBT+ is a promising training approach for therapists working with children and youth in the child welfare system.
- Including symptom monitoring as part of CBT+ treatment facilitates assessment of clinical change over time.
- Children and youth receiving CBT+ generally experienced improvements in mental health symptoms.
- While improvements in symptoms were observed for anxiety, depression, behavior problems and symptoms of traumatic stress, improvement trajectories differed.

✉ Suzanne E. U. Kerns
  Suzanne.Kerns@cuanschutz.edu

1 University of Denver, Center for Effective Interventions, Graduate School of Social Work, 2148 S. High Street, Craig Hall, Denver, CO 80208, USA

2 University of Connecticut, School of Social Work, 38 Prospect St, Hartford, CT 06103, USA

3 Harborview/UW Medicine, University of Washington, 325 9th Ave, MS 359947, Seattle, WA 98104, USA

4 University of Maryland, School of Social Work, 525 W. Redwood Street, Baltimore, MD 21201, USA

5 Present address: The Kempe Center at the University of Colorado-Anschutz, Aurora, CO, USA

Children and youth involved in child welfare systems (CWS) have a diverse set of needs, frequently including emotional and behavioral health-related concerns (Burns et al., 2004). The rates of emotional and behavioral health needs are typically higher in this population of children and youth compared to non-child welfare involved counterparts (Casanueva et al., 2011). Moreover, those CWS-involved children and youth placed in foster care have the highest rates of mental health-related concerns (Goemans et al., 2016; Landsverk et al., 2009; Lehmann et al., 2013). Untreated mental health concerns, and particularly behavioral concerns, are associated with placement disruptions and subsequent placement challenges (Chamberlain et al., 2006; James, 2004). Because of these reasons, in many

jurisdictions, CWS place an emphasis on improving child wellbeing and addressing emotional and behavioral healthcare needs. Indeed, promoting child wellbeing is one of the main goals of CWS and, therefore, child welfare workers are expected to identify and take steps to address the emotional and behavioral healthcare needs of the children in their care (US Department of Health and Human Services, 2019). Strategies to achieve this goal vary considerably across the United States. No singular strategy has emerged as the most effective (Hambrick et al., 2016; Landers et al., 2018).

Due to the relatively high base rates of mental health symptoms for children and youth in CWS, and particularly those in foster care (Kerker & Dore, 2006), some locales recommend that all children and youth receive mental health treatment. This approach may be problematic because, despite relatively high prevalence rates, many CWS-involved children and youth do not have observable symptoms (Bronsard et al., 2016). We know from experience that some agencies and courts send nearly every older child to mental health services (MHS). Requiring participation in MHS for nonsymptomatic children and youth may put an unnecessary burden on families. It may unnecessarily convey to children and their caregivers that what may be normal adjustment reactions to their circumstances is psychopathology requiring formal psychological intervention. Further, the MHS workforce is frequently lacking available professionals to meet the demands for clinical services (Thomas et al., 2009; Walker, et al., 2016). Thus, it is important to refer precisely those children and youth who actually have mental health needs. Once mental health needs are identified, child welfare professionals are often the gatekeepers connecting children and youth with therapeutic services (Dorsey et al., 2012).

Such services should be evidence-based and trauma-informed (Berliner et al., 2015; Kerns et al., 2016). However, although children and youth in CWS may be more likely than those in the general public to receive mental health services, the services they receive are less likely to be evidence-based (Burns et al., 2004; Horwitz et al., 2010). This may contribute to the lack of strong evidence that mental health services provided to CWS-involved children are effective in improving the wellbeing of these children (McCrae et al., 2010). Unique therapeutic challenges often arise during the provision of trauma-informed and evidence-based interventions for the child welfare population. This includes navigating regulations about who (e.g., foster parents, biological parents, siblings without child welfare involvement) is permitted to participate in treatment, especially when a youth is in foster care, and addressing unique circumstances of children and youth living in out of home treatment within the context of an evidence-based treatment protocol (Kerns et al., 2014).

Partnering for Success

Recognizing the need for cross-systems partnerships to address the emotional and behavioral healthcare needs of child welfare involved children and youth, the National Center for Evidence-Based Practice in Child Welfare developed the Partnering for Success (PfS) model (Barth et al., 2020). The PfS model aligns the activities of child welfare and mental health professionals to provide a seamless transition between screening, referral, treatment, and progress monitoring for child welfare involved children and youth. The four primary components of the model include (1) the creation of local implementation teams to develop organizational supports; (2) synchronous training of child welfare workers and mental health therapists on the goals and key components of the mental health intervention (Cognitive Behavioral Therapy Plus [CBT+], described below); (3) consultation and guidance related to implementation and clinical treatment; and (4) continuous progress monitoring of therapeutic progress. The entirety of the model is described elsewhere (Barth et al., 2020); the focus of the current study is evaluation of the clinical effectiveness of the mental health treatment provided by therapists in the four sites that implemented the PfS model.

The mental health treatment component selected for PfS is CBT+ (Dorsey et al., 2016). CBT+ is a training model in CBTs for four of the most common childhood disorders: anxiety, depression, trauma impact, and behavior problems. It was originally developed to provide a low cost and efficient approach for public mental health in Washington State. Though at the time that the PfS project launched there was not an evidence-base specific to CBT+, the Washington State Institute for Public Policy (WSIPP) (https://www.wsipp.wa.gov/), which assesses the evidentiary status and cost-benefits for EBTs in Washington state, considered the approach used by CBT+ to be consistent with the research base for general cognitive behavioral therapies and aligned with the concept of "treatment families" (i.e., groupings of interventions that have a common theoretical basis and use similar therapeutic strategies; as described in Chorpita et al., 2011).

For two of the clinical targets—anxiety and depression—CBT+ uses CBT-based treatment components because CBT is the well-established first line treatment for these conditions. To treat the third clinical target, behavior management, CBT+ includes CBT-based treatment components and behavioral management training for parents as the latter has been found to be highly effective in addressing disruptive behaviors in children. Trauma-Focused Cognitive Behavioral Therapy (TF-CBT; Cohen, Mannarino, & Deblinger, 2010), a well-researched and highly effective CBT for trauma-specific impact, was incorporated within CBT+ to treat the clinical target of trauma. In general,

CBT-based trauma-focused treatments are the first line for PTSD treatment (Bisson et al., 2019). It should be noted that after data were collected for the current study, CBT+ switched to a nonbranded version of CBT-based treatment for trauma because it can be more readily disseminated.

The CBT+ training model teaches therapists the core components of active treatments (e.g., explicit engagement strategies and measurement-based care). The model also includes the common CBT clinical components that appear in many CBTs (e.g., psychoeducation, coping skills, and cognitive restructuring) and the clinical components that are specific to each clinical condition (e.g., exposure for anxiety, behavioral activation for depression, trauma processing for symptoms of traumatic stress, and parent child relationship enhancement for behavior problems). CBT+ is not an explicitly transdiagnostic approach (as described by Dalgleish et al., 2020), meaning that, while CBT+ has the capabilities of addressing multiple treatment targets, it relies on diagnostic categorization as a basis for the selection and application of treatment components. Therapists are encouraged to identify a primary clinical target and follow a flowchart containing the common and specific clinical components that are matched to the clinical target. They are permitted to add modules/sessions when they are clinically indicated (Dorsey et al., 2016). CBT+ embraces the CBT principles of building collaborative and transparent therapist-client relationships, provision of skill-oriented sessions that are focused on the clinical target, homework to complete between sessions, and using standardized assessment tools to track progress.

CBT+ was informed by the common elements approach first elucidated by Chorpita, Daleiden, and Weisz (2005), which subsequently led to the development of MATCH-ADTC. MATCH-ADTC is a structured transdiagnostic intervention for anxiety, depression, trauma impact, and conduct problems. MATCH-ADTC has been shown to have good efficacy in clinical trials though a more recent trial indicates results may be influenced by study methodology (Chorpita et al., 2017; Weisz et al., 2020). See Dorsey, Berliner, Lyon, Pullmann, and Murray (2016) for more information about the evolution of CBT+. Though CBT+ was informed by MATCH-ADTC and is organized across the same treatment targets (anxiety, depression, PTSD, and behavior problems), CBT+ uses a modified organizational structure. Specifically, active treatment elements are organized by the clinical activities associated with engagement/ buy-in, addressing behaviors, addressing thoughts, and addressing feelings. CBT+ does not use the guided algorithm for treatment that is part of MATCH-ADTC and, instead, uses a simplified symptom monitoring approach to guide treatment decisions.

Within the CBT+ model, therapists start treatment with an assessment of clinical need. This assessment is informed by an initial clinical interview with the client and/or caregiver, information from the referral source, and client and/or caregiver responses to standardized mental health screening assessments. Therapists use the information obtained from these sources to identify the priority clinical target. CBT+ requires use of standardized measures as part of the evidence-based treat to target approach. The model does not definitively specify which measures the therapists are required to use as long as the measure is validated and specific to the clinical target. PfS requires therapists to administer the Pediatric Symptom Checklist-17 (PSC-17; Gardner et al., 1999), a caregiver completed general problem measure, plus at least one screening tool that measures symptoms of the treatment target (anxiety, depression, or trauma impact). The protocol permits flexibility across sites in regard to which tool is used for a given clinical target (e.g., anxiety, depression, or trauma) because agencies may already be using different tools and it is generally unnecessary to disrupt their routine screening procedures if validated tools are being used. For example, some sites may administer the Child and Adolescent Trauma Screen (CATS) while others may use the Child PTSD Symptom Scale (CPSS) to assess DSM-based symptoms of traumatic stress. Therapists are required to readminister the clinical target measure a minimum of twice following baseline using the measurement-based care approach. They are encouraged to measure progress even more frequently, which explains why some cases have more measures. Therapists record this information, along with the elements of the treatment delivered in each session, in a HIPAA-compliant web-based platform called the EBP Toolkit (https://www.ebptoolkit.com/). The EBP Toolkit entries are expected to be reviewed by supervisors and clinical consultants on a regular basis.

Training in CBT+ was adjusted for the PfS initiative. For most therapists, initial training consisted of (1) completion of the on-line TF-CBT training (https://tfcbt2.musc.edu/) and (2) an in-person 2- or 3-day intensive training conducted in conjunction with the child welfare training participants. The in-person training included integrated exercises designed to enhance collaboration and communication between child welfare professionals and mental health practitioners (Barth et al., 2020). Following training, therapists were required to participate in at least 9 out of 12 group-based bi-weekly consultation calls. Consultation calls (consisting of 8–10 trainees in each consultation group) were conducted by clinicians who are experts in the CBT+ model. The general agenda for each consultation call included: Roll call, prioritizing any urgent case-related needs (especially safety-related), updates from the previous consultation call, presentation of typically two cases from two different therapists, and selecting which therapist will present during the next call. Case presentations included

discussion and provision of feedback from consultants and other therapists on the call. Consultants recorded who participated in the calls and the case information discussed in the EBP Toolkit.

The CBT+ model offers a public domain option where a CBT-based trauma-focused training approach is substituted for TF-CBT. It has the potential to be feasibly scaled and integrated across child welfare and mental health jurisdictions. A major benefit of the model is the ability to train therapists on evidence-based strategies to address multiple prevalent disorder categories and the ability of the intervention to be flexible for those clients with comorbidities, which are especially common in child welfare settings (Dorsey et al., 2016). However, while preliminary studies indicate participation in CBT+ in the context of PfS is positively related to placement stability (Cosgrove, Lee, Greeno, & Horen, in press), thus far, the clinical effectiveness of the CBT+ training model has not been established. The present study aims to provide indicators of clinical effectiveness when CBT+ is delivered within the context of the PfS model. The primary research question is: Do children and youth who received CBT+ treatment from PfS trained therapists show reduced symptoms over time? The hypothesis is that clinical improvements will be observed across all clinical targets.

Methods

Human Subjects Approval

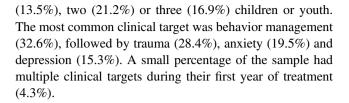This study received human subjects approval through the University of Maryland Institutional Review Board.

Participants

Service settings

The PfS model was implemented between 2013–2018 in four different sites, including (1) Baltimore, Maryland, (2) Brooklyn, New York, (3) two regions in Maine and (4) Tulsa (and surrounding area), Oklahoma. Regional child welfare jurisdictions recruited community-based mental health (CMH) agencies that served children and youth in the child welfare system. Those CMH agencies identified therapists to participate in the initiative.

Sample

Children and youth received CBT+ treatment ($N = 2285$) by one of 260 PfS-trained therapists. The average number of CBT+ clients documented per therapist was 8.79 ($SD = 13.74$), though most documented program use with one

(13.5%), two (21.2%) or three (16.9%) children or youth. The most common clinical target was behavior management (32.6%), followed by trauma (28.4%), anxiety (19.5%) and depression (15.3%). A small percentage of the sample had multiple clinical targets during their first year of treatment (4.3%).

Analytic sample To be included in the analytic sample, children and youth had to be assigned a clinical target. Those receiving multiple clinical targets were removed from the sample given that this study examined outcomes associated with target-specific treatments. A baseline measurement was required to be administered within three weeks of initiating treatment and at least one subsequent measurement had to be administered at least two weeks post-baseline. No restrictions were placed on the analytic sample for the timing of mid-treatment administrations of measures, which varied by client and therapist. All administrations had to be collected by the same respondent (usually the biological mother) to reduce bias associated with using scores from different respondents when measuring change in symptoms (Romano et al., 2018; Van der Meer et al., 2008; Weems et al., 2011). There were 1458 children and youth with scores on the measures meeting the criteria (64% of the full sample). Approximately 12% had missing values for one or more variables used in the overall statistical model, resulting in a final analytic sample of 1281 children and youth (56% of the full sample). Relative to children and youth removed from the sample, those in the analytic sample were slightly older on average ($M = 11.16$ years, $SD = 3.10$ vs. $M = 10.65$ years, $SD = 3.58$; $t(1353) = 3.28$, $p = 0.001$), and more likely to have a clinical target of behavior management (42.4% compared to 22.2%; $X^2[3] = 112.33$, $p < 0.001$) or receive treatment in Baltimore (94.8% compared to 84.5%; $X^2[3] = 70.19$, $p < 0.001$). See Table 1 for complete demographic information for the analytic sample.

Procedure

For each child and youth, therapists recorded demographic characteristics, the clinical target, dates of treatment sessions, the clinical elements of CBT+ delivered in each session, and responses to items on the measures in the EBP Toolkit. Therapists were required to administer a measure appropriate for the child or youth's clinical target at baseline, at case closure, and periodically during treatment. Baseline scores were obtained prior to initiating active treatment or within the first session or two. As specified in the CBT+ training model, there were no hard and fast stipulations about the frequency of ongoing progress monitoring. Therapists were expected to recollect assessment data at least monthly.

Table 1 Participant demographic characteristics

	All (n = 1281) n (%)	Behavior (n = 543) n (%)	Anxiety (n = 205) n (%)	Depression (n = 208) n (%)	Trauma (n = 325) n (%)
Age (years)	M = 11.17 SD = 3.10	M = 9.76 SD = 2.92	M = 11.76 SD = 2.59	M = 12.99 SD = 2.60	M = 11.99 SD = 2.96
Gender					
Female	608 (47.5)	179 (33.0)	127 (62.0)	123 (59.1)	179 (55.1)
Male	673 (52.5)	364 (67.0)	78 (38.1)	85 (40.9)	146 (44.9)
Race/Ethnicity					
African Amer	473 (36.9)	221 (40.7)	47 (22.9)	72 (34.6)	133 (40.9)
White	573 (44.7)	212 (39.0)	124 (60.5)	91 (43.8)	146 (44.9)
Latinx	77 (6.0)	28 (5.2)	13 (6.3)	16 (7.7)	20 (6.2)
Multiple	131 (10.2)	74 (13.6)	16 (7.8)	22 (10.6)	19 (5.9)
Other	27 (2.1)	8 (1.5)	5 (2.4)	7 (3.4)	7 (2.2)
Site					
Baltimore	1214 (94.8)	519 (95.6)	191 (93.2)	196 (94.2)	308 (94.8)
Maine	22 (1.7)	9 (1.7)	7 (3.4)	5 (2.4)	1 (0.3)
Oklahoma	28 (2.2)	5 (0.9)	5 (2.4)	2 (1.0)	16 (4.9)
New York	17 (1.3)	10 (1.8)	2 (1.0)	5 (2.4)	–
Baseline score on measure	–	M = 7.10 SD = 3.93	M = 3.60 SD = 2.18	M = 10.59 SD = 6.83	M = 25.78 SD = 12.88

Notes: Measures used were PSC-17 externalizing scale (≥7 = clinical range) for behavior management, SCARED anxiety – self report (≥3 = clinical range) for anxiety, MFQ – self (≥11 = clinical range) report for depression and CATS – self report (≥15 = clinical range) for trauma.

Measures

Outcome variables

Changes in mental health symptoms were measured by standardized and validated mental health screening tools. The measures used by therapists varied by clinical target (i.e., behavior, anxiety, depression, and trauma) and, in some sites, therapists could select one of several measures appropriate for a clinical target. For example, therapists could use the General Anxiety Disorder-9 (GAD-7) or the Screen for Child Anxiety Related Disorders (SCARED-Anxiety) to measure symptoms of anxiety. However, preliminary analyses revealed only one measure per clinical target that satisfied the sampling criteria described above (i.e., baseline score collected within three weeks, follow-up score collected at least two weeks post-baseline, baseline and follow-up scores collected by same responder, and a minimum of 100 children and youth had scores meeting the aforementioned criteria). Only measures with baseline and follow-up scores for 100 or more children and youth were analyzed to obtain unbiased estimates when using multi-level modeling (Maas & Hox, 2005; McNeish & Stapleton, 2016).

Clinical target: behavior management The externalizing subscale of the Pediatric Symptom Checklist-17 (PSC-17; Gardner et al., 1999) was used to measure behavioral problems among children and youth (ages 4–17) with a clinical target of behavior management. Although scores on the attention subscale of the PSC-17 met criteria for inclusion (n = 191) only the externalizing subscale was used because it is a more direct measure of behavioral problems. The full PSC-17 is a 17-item scale that contains three subscales, externalizing, internalizing, and attention. The measure has been shown to have adequate reliability and validity (Gardner et al., 2007), though there may be some variability in the measure's convergent validity between different reporters for child welfare samples (Parker et al., 2019).

The externalizing subscale contains seven items and is completed by a primary caregiver (adult). Responses are on a 3-point frequency scale (0 = Never, 1 = Sometimes, or 2 = Often). Total scores for the subscale can range from zero to 14, and scores greater than or equal to seven are in the clinical range. The omega total coefficient (ω)—with a polychoric covariance matrix—was used to assess inter-item reliability because it has less rigid and more realistic assumptions than the more commonly known Cronbach's alpha (McNeish, 2018). The internal consistency in this study ($\omega = 0.90$) was very good (DeVellis, 2003).

Clinical target: anxiety The anxiety subscale of the Screen for Childhood Anxiety Related Emotional Disorders (SCARED) Brief Assessment (Birmaher et al., 1999) was

used to measure symptoms for children and youth (ages 8–17) with a clinical target of anxiety. The other measures clinicians used to assess anxiety had sample sizes ($n = 27$ or less), well below the cutoff ($n = 100$) established in this study for evaluating change in symptoms over time. The anxiety subscale contains five items and is completed by the youth and/or the primary caregiver. Responses are on a 3-point frequency scale (0 = Not true or hardly ever true, 1 = Sometimes true or somewhat true, or 2 = very true or often true). Total scores for the subscale can range from zero to 10 and scores greater than or equal to three are in the clinical range. Internal consistency for the anxiety subscale in this study ($\omega = 0.72$) was respectable (DeVellis, 2003).

Clinical target: depression The Moods and Feelings Questionnaire – self report (MFQ) was used to measure symptoms among children and youth (ages 8–17) with a clinical target of depression. Two measures of depression were not examined because the numbers of children and youth with follow-up scores ($n = 32$ and $n = 20$) were below the cutoff of 100. The MFQ has previously been shown to have adequate reliability and validity (Wood et al., 1995) and its internal consistency in the present study ($\omega = 0.96$) was very good (DeVellis, 2003). The 13-item questionnaire asks respondents if they have felt or acted in various ways in the past two weeks (0 = Not true, 1 = Sometimes true, 2 = True). Total scores can range from zero to 26, scores greater than or equal to 11 are in the clinical range.

Clinical target: traumatic stress The Child and Adolescent Trauma Screen – self report (CATS) was used to measure symptoms of traumatic stress among children and youth (ages 7–17) with a clinical target of trauma. The other instruments used to measure trauma symptoms had 66 observations or less and thus did not meet inclusion criteria. The CATS scale contains 20 items and respondents report how frequently various feelings, thoughts, or behaviors have bothered them during the prior two weeks (0 = Never, 1 = Once in a while, 2 = Half the time, and 3 = Almost always). Total scores can range from zero to 60 and the clinical cutoff is 15. The CATS has previously been shown to have adequate reliability and validity (Sachser et al., 2017) and its internal consistency in the present study was very good ($\omega = 0.94$; DeVellis, 2003).

Treatment duration

A continuous variable measured the *number of months* that a child or youth was in treatment at the time the measure was administered. Quadratic and cubic terms for this variable were calculated to test for nonlinear relationships.

Child demographics

A continuous variable was used to represent the child's *age* in years at the start of treatment. *Race and ethnicity* were included in the model with dummy variables indicating if the child or youth was white, African American, Latinx/ Hispanic, or multiracial/other race. Dichotomous variables were used to represent *gender* (0 = male, 1 = female) and site (0 = Maine, New York, and Oklahoma, 1 = Baltimore). Children and youth from New York, Maine, and Oklahoma, were combined because of the smaller proportion of respondents in these categories (see Table 1).

Analytic Approach

One of the challenges in the present study was accounting for the fact that follow-up measurements were not obtained at fixed time intervals (e.g., every three months or every six months). Rather, the mental health screening instruments were administered when therapists deemed it clinically warranted, as is typical in nonexperimental clinical settings. Multilevel modeling was chosen to test for significant changes in symptomology over time because it allows for unbalanced longitudinal data where the number of measurement occasions and the time between measurement occasions varies across subjects (Snijders & Bosker, 2012). In a multilevel model with longitudinal data, the measurements occasions (level 1) are nested within the individual (level 2). The model can test whether there is a significant change over time for all subjects overall (a fixed effect), as well as whether the amount of change varies significantly across subjects (a random effect).

A multilevel linear growth model was constructed for each clinical target using measurements obtained within the first year of treatment. A two-step model-building process was followed for each clinical target. In step one, a random intercept model was tested that included level 1 variables (i.e., scores on the mental health screening instruments and treatment duration) and level 2 variables (i.e., covariates) as fixed effects. To test for nonlinear changes in symptomology, quadratic and cubic terms for treatment duration were sequentially entered into the model and only included in subsequent models if they were statistically significant. The analysis was conducted with robust standard errors because the small number of observations from some sites made it difficult to assess whether the assumption of homoskedasticity (that the variance is equal across sites) was met. A sensitivity analysis showed that the results are similar with and without robust standard errors. In step two, a random effect for treatment duration was tested by allowing its slope to vary across children and youth, and comparing model fit of the restricted model (i.e., fixed effect model) and unrestricted model (i.e., random effect model). An

adjusted likelihood ratio test (Snijders & Bosker, 2012) was used to test for a significant random effect where the critical value for the difference in the deviances of the restricted and the unrestricted models is tested using a chi-bar distribution. Diagnostic tests (e.g., histograms and scatterplots of residuals) indicated that each multilevel met these assumptions: There were no outlying residuals that changed the parameter estimates, the residuals were normally distributed with means close to zero (means ranged from 0.00 to 0.04), and the residuals were equally distributed across values of the predictor variables (i.e., homoscedastic).

Post-hoc analyses were conducted due to observed differences in racial and ethnic distribution of children and youth across clinical targets. First, a chi-square test of independence was used to test if clinical target was associated with race and ethnicity. Then, the adjusted standardized residuals for each cell (i.e., the difference between the observed and expected frequencies) were calculated and compared to the critical value for a two-tailed z-test ($z = 1.96$, $p = 0.05$).

All analyses were conducted in Stata (version 15).

Missing Data

Multilevel modeling approaches assume data are missing at random (MAR) or missing completely at random (MCAR; Snijders & Bosker, 2012). The extent of missing values was minimal (<10%) for most variables. Notably, age was missing for 11.2% of children and youth who had a baseline score for their clinical target. For each clinical target, t-tests and chi-square tests of independence were conducted to test (1) if the likelihood of a variable having a missing value was related to other variables in the model and (2) if there were differences between children and youth with a baseline score only compared to those with at least one follow-up score. Out of 64 comparisons, only children and youth with a clinical target of behavior management were more likely to have a follow-up score on the PSC-17 externalizing scale if they were between the ages of 3 and 7. No other significant differences were found. Based on these findings, the data were assumed to be MAR.

Results

Demographic Characteristics

The demographic characteristics of children and youth in the final models for each clinical target are provided in Table 1. Children and youth in the overall analytic sample were 11 years old on average, yet children and youth receiving treatment for behavior management were slightly younger ($M = 9.76$ years) relative children and youth in

treatment for anxiety ($M = 11.76$ years), depression ($M = 12.99$ years) or trauma ($M = 11.99$). Although the analytic sample was almost balanced regarding gender (47.5% female and 52.5% male), children and youth receiving treatment for behavior management were predominantly male (67.0%), while children and youth in treatment for anxiety or depression tended to be female (62.0% and 59.1%, respectively). The overall analytic sample was predominantly white (44.7%) or African American (36.9%). Notably, most children and youth receiving treatment for anxiety were white (60.5%) while most children and youth with a clinical target of behavior management were non-white (61.0%). Over 90% of the children and youth in the overall analytic sample and within each clinical target grouping were from the Baltimore site.

Average baseline scores measures of behavior management ($M = 7.33$, $SD = 3.48$) and anxiety ($M = 3.68$, $SD = 2.23$) were slightly above (<1 point) their respective clinical cutoffs. The average baseline score for children and youth in treatment for depression ($M = 9.87$, $SD = 6.69$) was slightly below (1.13 points) the clinical range. The average baseline score for children in treatment for trauma ($M = 25.78$, $SD = 13.24$) was well above (10 points) the clinical cutoff.

Results of the Multilevel Models

The results of the final multilevel models for all clinical targets are presented in Table 2.

Behavior management

Children and youth with a clinical target of behavior management, showed a negative linear relationship ($b = -0.48$, $SE = 0.08$, $p < 0.001$) and a positive quadratic relationship ($b = 0.04$, $SE = 0.01$, $p < 0.001$) between treatment duration and scores on the PSC-17 externalizing scales. These findings indicate that children and youth experienced initial improvement in symptoms, although the amount of improvement was reduced with each month of treatment. The adjusted likelihood ratio tests did not detect any variation (i.e., random effects) for the linear (Δ Deviance = 0.186, $p = .789$) or the quadratic terms (Δ Deviance = 0.00, $p = 1.00$).

Anxiety

A quadratic relationship was found among children and youth with a clinical target of anxiety. More specifically, a negative linear relationship ($b = -0.30$, $SE = 0.06$, $p < 0.001$), and a positive quadratic relationship ($b = 0.02$, $SE = 0.01$, $p = 0.008$) was detected between treatment duration and scores on the SCARED anxiety – self report.

Table 2 Results of multilevel models testing for change in symptoms over time by clinical target.

	Behavior ($n = 543$)		Anxiety ($n = 205$)		Depression ($n = 208$)		Trauma ($n = 325$)	
	b	Robust SE	b	Robust SE	b	Robust SE	b	Robust SE
Time								
Months in treatment	−0.48***	0.08	−0.30***	0.06	−1.45***	0.21	−1.47***	0.18
Months in treatment^2	0.04***	0.01	0.02**	0.01	0.11***	0.02	–	–
Age (years)	−0.07	0.05	0.08	0.06	0.43**	0.16	0.84***	0.22
Gender (ref. = female)								
Male	−0.08	0.30	−0.83**	0.28	−3.05***	0.85	−4.48**	1.31
Race/Ethnicity (ref. = white)								
African Amer.	−0.09	0.31	−0.19	0.34	−0.31	1.03	−0.46	1.35
Latinx/Hispanic	−0.40	0.59	0.22	0.54	−0.34	1.25	1.25	3.03
Multiracial/other	0.08	0.43	0.63	0.41	0.08	1.07	3.39	2.41
Site (ref. = Baltimore)								
Maine	−0.99	1.06	0.39	0.72	−0.97	3.01	16.64***	1.43
Oklahoma	3.23*	1.11	0.23	0.85	0.29	1.56	0.23	3.30
NYC	−0.36	1.17	−0.35	0.75	−0.96	1.95	–	–

Note. *$p < .05$; **$p < .01$; ***$p < .001$

No significant variation was found for the linear (Δ Deviance = 4.00, $p = 0.090$) or the quadratic relationship (Δ Deviance = 2.26, $p = 0.232$).

Depression

The results suggest a quadratic relationship between treatment duration and scores on the MFQ – self report. As in the prior models, a negative linear relationship ($b = −1.45$, $SE = 0.20$, $p < 0.001$) and a positive quadratic relationship ($b = 0.11$, $SE = 0.02$, $p < 0.001$) was found. Neither the linear relationship (Δ Deviance = 1.24, $p = 0.401$) nor the quadratic relationship (Δ Deviance = 0.06, $p = 0.892$) varied significantly.

Trauma

Children and youth with a clinical target of trauma consistently showed decreased symptoms with each month of treatment. On average, scores on the CATS decreased 1.5 points each month they were in treatment ($b = −1.47$, $SE = 0.18$, $p < 0.001$). The magnitude of change did not vary significantly across children and youth (Δ Deviance = 3.49, $p = 0.098$).

Post-hoc analysis

Clinical target was associated with race and ethnicity in the overall sample of children and youth who received treatment in the PfS program, $X^2(12) = 72.70$, $p < 0.001$. African American children and youth were more likely than expected to have a clinical target of behavior management (40.4%; $z = 4.44$, $p < 0.001$) and less likely than expected to have a clinical target of anxiety (13.5%; $z = −5.67$, $p < 0.001$). Multiracial children were also more likely than expected to receive treatment for behavior management (44.9%; $z = 3.49$, $p < 0.001$) and were less likely to receive treatment for trauma (19.6%; $z = −3.39$, $p < 0.001$). While African American and multiracial children and youth were more likely than expected to have a clinical target of behavior management, white children and youth were less likely than expected to have a clinical target of behavior management (28.0%; $z = −5.39$, $p < 0.001$). White children and youth were also more likely than expected to receive treatment for anxiety ($z = 5.41$, $p < 0.001$).

Discussion

The current study evaluated change in clinical symptoms for children and youth receiving one of the four treatment protocols contained within CBT+ training model and delivered within the context of a broader initiative (PfS) to improve the working partnerships between mental health and child welfare professionals (Barth et al., 2020). This study demonstrated preliminary evidence of clinical improvements in anxiety, depression, behavior problems and symptoms of traumatic stress for children and youth, as evidenced by reduced symptoms observed over time using standardized and validated clinical measures. For children and youth in child welfare, a supportive component of effective cross-system collaboration and mental health

treatment is progress monitoring (Barth et al., 2020; Barth & Watrous, 2021). Progress monitoring provides valuable information about the focus of treatment and improvement trajectories. PfS provides a prototype approach for supporting such collaboration. The purpose of this study was to evaluate the clinical outcomes for children and youth who receive mental health treatment from therapists trained in CBT+, which includes progress monitoring. Future studies will examine the unique contributions of progress monitoring to cross-system collaboration.

Treatment protocols were flexible based on observed clinical need and delivered within an initiative that supported therapists to collaborate closely with child welfare professionals. In this study, substantial improvements were noted across all four target areas (anxiety, depression, behavior problems and post-traumatic stress) over the course of treatment. Curvilinear trajectories were observed for anxiety, depression and behavior problems, which indicated the rate of improved symptomology decreased over time. The coefficients for the quadratic terms were quite small in the sample and, therefore, caution is warranted when interpreting these findings. Traumatic symptoms had a more linear improvement trajectory.

Given the limited data available for the present study, the reason for the different trajectories was not able to be fully explored. However, there could be several potential explanations. One possible contributor is that PTSD, unlike other disorders, has a cause (e.g., the traumatic stressor); trauma focused models emphasize explicit validation and normalization. This active strategy may enhance client engagement by providing the rationale for treatment participation and reduce distress associated with stigma. There is also a natural recovery course for trauma-specific impact (e.g., Bell et al., 2015). The other disorders may be more vulnerable to relapse. For example, it is common that behavior problems reemerge when parents relax their use of behavior management approaches (Pritchard et al., 2014). The findings of this study implied quadratic relationships while prior research indicates cubic relationships is more common. In general, such curvilinear relationships are common in treatment research (e.g., Chu et al., 2013; Sunderland et al., 2012). Especially for anxiety-related disorders, there may be an initial spike in symptoms, (Chu et al., 2013). Further, the timing of symptom measurement in relation to the delivery of different treatment elements may influence the trajectory of symptom improvement. Certain CBT treatment elements are more related to change than others (Peris et al., 2015). Without a study comparison group and the ability to link symptom change to delivery of clinical elements, interpretation of treatment trajectories is limited. Nonetheless, it is promising that most youth demonstrated substantial improvement over time during treatment.

The most commonly delivered treatment was behavioral parenting for child behavior problems. One observation in these findings was that there was a higher relative percentage of males and African American children receiving treatment for behavior problems. This finding could be expected given previous research indicating higher prevalence rates of behavior problems for males (Zahn-Waxler, 1993). However, the proportion of children and youth of color being referred for behavioral treatment occurred despite little explanatory evidence that such youth in child welfare have substantial differences in externalizing behavior symptoms compared with internalizing symptoms (Casanueva et al., 2011; McCrae, 2009). Why are children of color more likely, then, to receive treatment for behavior problems? The answer appears complex and multifaceted. Within the child welfare system, it is possible that behavioral concerns are emphasized by referral agencies for children of color because they are more likely to experience a placement disruption due to behavior problems (Leathers, 2006). It is further possible that therapists are more likely to diagnose behavioral concerns for people of color (Garb, 1997).

Results of this study indicate that the CBT+ model has preliminary indications of promise for supporting the emotional and behavioral health needs of CWS-involved children and youth. Strengths of the model include (1) the ability to address multiple mental health concerns within one structure; (2) the ability to address co-morbid problems; and (3) the symptom monitoring structure, which aids in clinical decision making and communication with child welfare partners. The CBT+ model is designed to address the most common emotional and behavioral disorders in youth. Anxiety, depression, behavior problems, and traumatic stress impact about 49% of children and youth in child welfare (Bronsard et al., 2016). Further, about 43% have comorbid disorders (Lehmann et al., 2013). Though more specialty services will always be needed, the CBT+ model represents a flexible approach that addresses many needs. The training and consultation support approach appears to be feasible for larger system scale-up (Barth et al., 2020; Dorsey et al., 2016).

In our study, ongoing symptom monitoring occurred alongside a CBT-based intervention. This type of model has the potential to align well with broader child welfare efforts. For example, screening and ongoing progress monitoring is an important role for many child welfare jurisdictions (Lang et al., 2017). Coordination of these efforts with therapists trained in the full CBT+ approach could provide a common language between child welfare and mental health therapists. It is likewise important to build a culture of promotion of health and wellbeing, compared with a focus on compliance. Symptom monitoring, if shared across mental health and child welfare could help transition mental health
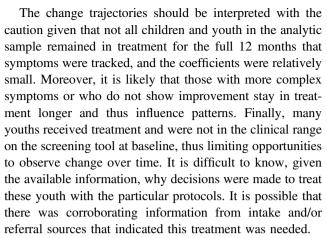
referrals from a "referral only" strategy in which referrals are made but have limited follow-up to a meaningful dialog about symptom improvements and changes, allowing for critical assessments of progress and determinations if the best fitting treatments are being delivered.

Limitations

Despite notable strengths, including the sample size and diverse racial and ethnic client representation, this study has some important limitations. Many of these limitations are compromises the research team needed to make because of the 'real world' implementation of the PfS initiative. Fidelity to the CBT+ model was monitored through consultation calls and lacked a formal fidelity tracking system. Therefore, it is not possible to assess definitively the extent to which therapists adhered to the treatment model. Although the therapists in this study received instructions about data entry to facilitate symptom monitoring and progression through treatment, substantial variability was observed, as would be expected in any implementation effort. This led to varying intervals between symptom monitoring periods. The purpose of this study was to examine change over time, rather than a formal pre- and post-treatment assessment. Interpretation of the findings should consider that the follow-up measurement period does not necessarily equate to post treatment in all cases. Further, symptom measures were not strictly standardized across sites to allow for regional variation and minimize system disruption. Children and youth with multiple treatment targets were excluded from the sample because the analytic approach necessitated a primary treatment target and CBT+ was not designed to be used to treat multiple clinical targets simultaneously. Future evaluations of CBT+ will benefit from standardizing symptom monitoring protocols across sites and exploring whether CBT+ is effective in treatment multiple treatment targets.

This study lacks a comparison group. While, overall, children and youths' symptoms improved, it is not possible to conclude if this improvement was commensurate with what would be expected within usual care settings and/or with other evidence-based interventions. Differences were observed between the full sample and the analytic sample. It is possible some differences were a result of differences in measurement availability. For example, there was only one measurement tool for behavior problems, while there were several different options for anxiety, trauma, and depression. A relatively higher proportion of children and youth who received treatment in Baltimore had symptom measurement that met the inclusion criteria. This was likely due to implementation differences across sites that likely impacted the consistency of data collection in Baltimore compared to the other sites.

The change trajectories should be interpreted with the caution given that not all children and youth in the analytic sample remained in treatment for the full 12 months that symptoms were tracked, and the coefficients were relatively small. Moreover, it is likely that those with more complex symptoms or who do not show improvement stay in treatment longer and thus influence patterns. Finally, many youths received treatment and were not in the clinical range on the screening tool at baseline, thus limiting opportunities to observe change over time. It is difficult to know, given the available information, why decisions were made to treat these youth with the particular protocols. It is possible that there was corroborating information from intake and/or referral sources that indicated this treatment was needed.

Alternatively, youth may be getting referred based on a protocol that lacks a mechanism to determine whether they have clinical needs. This creates a challenge when documenting clinical change and may indicate that the sensitivity of the screening tools should be further evaluated or that referral procedures need to be more evidence-based, or both. Finally, we were unable to evaluate the impact of cross-system collaboration on clinical outcomes given the nature of the available data. In the future, using more recently developed tools such as the "Perceptions of Overarching Cross-System Collaboration–Child Welfare and Behavioral Health Systems (POCSC-CW/BH)" (Crandal et al., 2019) may be helpful to better understand the unique contributions of system coordination.

Conclusions, Implications and Recommendations

Children and youth receiving child welfare services represent a unique treatment population given the high prevalence of mental health conditions and the dual experiences in child welfare and mental health systems. Ensuring that evidence-based interventions are flexible enough to meet their unique needs is critical. This study provides emerging evidence that CBT+ is a viable mental health treatment approach for children and youth involved in the child welfare system. It offers reach for public mental health organizations with a single training and support infrastructure covering several of the most common client clinical conditions; it uses a well-established clinical theory that underpins effective treatments for multiple disorders; and it includes measurement-based care as an ingredient of evidence-based treatment delivery.

Measurement-based care may be an important tool for supporting child wellbeing. Doing so within a training approach such as CBT+ has the advantage of being able to provide evidence-supported interventions in a manner that can address co-morbid emotional and behavioral concerns. It is frequently not feasible for public mental health organizations to have a workforce that is expert in multiple

brand name EBTs with expectations for adherence to a myriad of specialized protocols. Documenting improvement and identifying failure to improve and making clinical adjustments provides maximum benefit for clients, therapists, and organizations on the core feature of clinical interventions, client outcomes. Systematic monitoring of clinical symptoms may also assist in supporting communication and collaboration between mental health and child welfare professionals.

To further establish effectiveness, future research should evaluate CBT+ clinical outcomes in using an experimental or quasi-experimental study design. Examination of moderators and mediators of treatment may help identify patterns of clinical responsiveness that can inform future training efforts. Planned future studies of CBT+ using the current sample include examining the role of consultation and training support on client outcomes and influences on how decisions about clinical targets are made.

Compliance with Ethical Standards

Conflict of Interest The authors declare no competing interests.

References

Barth, R. P., Rozeff, L., Kerns, S. E. U., & Baldwin, M. (2020). Partnering for success: Implementing a cross-systems collaborative model between behavioral health and child welfare. *Children & Youth Services Review*, *117*, 104663 https://doi.org/10.1016/j.childyouth.2019.104663.

Barth, R. P., & Watrous, J. (2021). Partnering for success – A prototype for integrating evidence-based practices between referring professionals and mental health professionals. *Journal of Clinical Pediatrics and Neonatology*, *1*(1), 1–3.

Bell, T., Romano, E., & Flynn, R. (2015). Profiles and predictors of behavioral resilience among children in child welfare. *Child Abuse & Neglect*, *48*, 92–103. https://doi.org/10.1016/j.chiabu.2015.04.018.

Berliner, L., Fitzgerald, M. M., Dorsey, S., Chaffin, M., Ondersma, S. J., & Wilson, C. (2015). Report of the APSAC Task Force on evidence-based service planning guidelines for child welfare. *Child Maltreatment*, *20*(1), 6–16. https://doi.org/10.1177/1077559514562066.

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the screen for child anxiety related emotional disorders (SCARED): a replication study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(10), 1230–1236. https://doi.org/10.1097/00004583-199910000-00011.

Bisson, J. I., Berliner, L., Cloitre, M., Forbes, D., Jensen, T. K., Lewis, C., & Shapiro, F. (2019). The international society for traumatic stress studies new guidelines for the prevention and treatment of posttraumatic stress disorder: Methodology and development process. *Journal of Traumatic Stress*, *32*(4), 475–483. https://doi.org/10.1002/jts.22421.

Bronsard, G., Alessandrini, M., Fond, G., Loundou, A., Auquier, P., Tordjman, S., & Boyer, L. (2016). The prevalence of mental disorders among children and adolescents in the child welfare system: A systematic review and meta-analysis. *Medicine*, *95*(7), e2622 https://doi.org/10.1097/MD.0000000000002622.

Burns, B. J., Phillips, S. D., Wagner, H. R., Barth, R. P., Kolko, D. J., Campbell, Y., & Landsverk, J. (2004). Mental health need and access to mental health services by youths involved with child welfare: A national survey. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(8), 960–970. https://doi.org/10.1097/01.chi.0000127590.95585.65.

Casanueva, C., Ringeisen, H., Wilson, E., Smith, K., & Dolan, M. (2011). *NSCAW II baseline report: Child well-being*. Washington, DC: US Department of Health and Human Services.

Chamberlain, P., Price, J. M., Reid, J. B., Landsverk, J., Fisher, P. A., & Stoolmiller, M. (2006). Who disrupts from placement in foster and kinship care? *Child Abuse & Neglect*, *30*, 409–424. https://doi.org/10.1016/j.chiabu.2005.11.004.

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., & Starace, N. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, *18*(2), 154–172. https://doi.org/10.1111/j.1468-2850.2011.01247.x.

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., Chiu, A. W., Letamendi, A. M., Tsai, K. H., & Krull, J. L. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology*, *85*(1), 13–25. https://doi.org/10.1037/ccp0000133.

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, *7*(1), 5–20. https://doi.org/10.1007/s11020-005-1962-6.

Chu, B. C., Skriner, L. C., & Zandberg, L. J. (2013). Shape of change in cognitive behavioral therapy for youth anxiety: Symptom trajectory and predictors of change. *Journal of Consulting and Clinical Psychology*, *81*(4), 573 https://doi.org/10.1037/a0033390.

Cohen, J. A., Mannarino, A. P., & Deblinger, E. (2010). Trauma-focused cognitive-behavioral therapy for traumatized children. *Evidence-based Psychotherapies for Children and Adolescents*, *2*, 295–311.

Cosgrove, J. A., Lee, B. R., Greeno, E. J., & Horen, M. J. (in press). Cognitive Behavioral Therapy – Plus for youth in out-of-home care: A quasi-experimental evaluation. *Social Work Research*.

Crandal, B. R., Martin, J. K., Hazen, A. L., & Rolls Reutz, J. A. (2019). Measuring collaboration across children's behavioral health and child welfare systems. *Psychological Services*, *16*(1), 111–119. https://doi.org/10.1037/ser0000302.

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology*, *88*(3), 179–195. https://doi.org/10.1037/ccp0000482.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Dorsey, S., Berliner, L., Lyon, A. R., Pullmann, M. D., & Murray, L. K. (2016). A statewide common elements initiative for children's

mental health. *The Journal of Behavioral Health Services & Research*, 43(2), 246–261. https://doi.org/10.1007/s11414-014-9430-y.

Dorsey, S., Kerns, S. E., Trupin, E. W., Conover, K. L., & Berliner, L. (2012). Child welfare caseworkers as service brokers for youth in foster care: Findings from Project Focus. *Child Maltreatment*, 17(1), 22–31. https://doi.org/10.1177/1077559511429593.

Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice*, 4(2), 99–120. https://doi.org/10.1111/j.1468-2850.1997.tb00104.x.

Gardner, W., Lucas, A., Kolko, D. J., & Campo, J. V. (2007). Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(5), 611–618. https://doi.org/10.1097/chi.0b013e318032384b.

Gardner, W., Murphy, M., Childs, G., Kelleher, K., Pagano, M., Jellinek, M., & Chiappetta, L. (1999). The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. A report from PROS and ASPN. *Ambulatory Child Health*, 5, 225–225.

Goemans, A., van Geel, M., van Beem, M., & Vedder, P. (2016). Developmental outcomes of foster children: A meta-analytic comparison with children from the general population and children at risk who remained at home. *Child Maltreatment*, 21(3), 198–217. https://doi.org/10.1177/1077559516657637.

Hambrick, E. P., Oppenheim-Weller, S., N'zi, A. M., & Taussig, H. N. (2016). Mental health interventions for children in foster care: A systematic review. *Children and youth services review*, 70, 65–77. https://doi.org/10.1016/j.childyouth.2016.09.002.

Horwitz, S. M., Chamberlain, P., Landsverk, J., & Mullican, C. (2010). Improving the mental health of children in child welfare through the implementation of evidence-based parenting interventions. *Administration and Policy in Mental Health and Mental Health Services Research*, 37(1-2), 27–39. https://doi.org/10.1007/s10488-010-0274-3.

James, S. (2004). Why do foster care placements disrupt? An investigation of reasons for placement change in foster care. *Social Service Review*, 78, 601–627. https://doi.org/10.1086/424546.

Kerker, B. D., & Dore, M. M. (2006). Mental health needs and treatment of foster youth: Barriers and opportunities. *American Journal of Orthopsychiatry*, 76(1), 138–147. https://doi.org/10.1037/0002-9432.76.1.138.

Kerns, S. E. U., Pullmann, M. D., Putnam, B., Buher, A., Holland, S., Berliner, L., & Trupin, E. W. (2014). Child welfare and mental health: Facilitators of and barriers to connecting children and youths in out-of-home care with effective mental health treatment. *Children and Youth Services Review*, 46, 315–324. https://doi.org/10.1016/j.childyouth.2014.09.013.

Kerns, S. E. U., Pullmann, M. D., Negrete, A., Uomoto, J. A., Berliner, L., Shogren, D., & Putnam, B. (2016). Development and implementation of a child welfare workforce strategy to build a trauma-informed system of support for foster care. *Child Maltreatment*, 21(2), 135–146. https://doi.org/10.1177/1077559516633307.

Landers, A. L., McLuckie, A., Cann, R., Shapiro, V., Visintini, S., MacLaurin, B., & Carrey, N. J. (2018). A scoping review of evidence-based interventions available to parents of maltreated children ages 0-5 involved with child welfare services. *Child Abuse & Neglect*, 76, 546–560. https://doi.org/10.1016/j.chiabu.2017.09.012.

Landsverk, J. A., Burns, B. J., Stambaugh, L. F., & Reutz, J. A. (2009). Psychosocial interventions for children and adolescents in foster care: Review of research literature. *Child Welfare*, 88(1), 49–69.

Lang, J. M., Ake, G., Barto, B., Caringi, J., Little, C., Baldwin, M. J., & Stevens, K. (2017). Trauma screening in child welfare: Lessons learned from five states. *Journal of Child & Adolescent Trauma*, 10(4), 405–416. https://doi.org/10.1007/s40653-017-0155-y.

Leathers, S. J. (2006). Placement disruption and negative placement outcomes among adolescents in long-term foster care: The role of behavior problems. *Child Abuse & Neglect*, 30(3), 307–324. https://doi.org/10.1016/j.chiabu.2005.09.003.

Lehmann, S., Havik, O. E., Havik, T., & Heiervang, E. R. (2013). Mental disorders in foster children: a study of prevalence, comorbidity and risk factors. *Child and Adolescent Psychiatry and Mental Health*, 7(1), 39 https://doi.org/10.1186/1753-2000-7-39.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample size for multilevel modeling. *Methodology*, 1(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86.

McCrae, J. S. (2009). Emotional and behavioral problems reported in child welfare over 3 years. *Journal of Emotional and Behavioral Disorders*, 17(1), 17–28. https://doi.org/10.1177/1063426608319141.

McCrae, J. S., Barth, R. P., & Guo, S. (2010). Changes in maltreated children's emotional–behavioral problems following typically provided mental health services. *American Journal of Orthopsychiatry*, 80(3), 350 https://doi.org/10.1111/j.1939-0025.2010.01039.x.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. https://doi.org/10.1037/met0000144.

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28(2), 295–314. https://doi.org/10.1007/s10648-014-9287-x.

Parker, E. M., Jacobson, J., Pullmann, M. D., & Kerns, S. E. U. (2019). Identifying psychosocial problems among children and youth in the child welfare system using the PSC-17: Exploring convergent and discriminant validity with multiple informants. *Child Psychiatry & Human Development*, 50(1), 108–120. https://doi.org/10.1007/s10578-018-0824-6.

Peris, T. S., Compton, S. N., Kendall, P. C., Birmaher, B., Sherrill, J., March, J., & Piacentini, J. (2015). Trajectories of change in youth anxiety during cognitive—behavior therapy. *Journal of Consulting and Clinical Psychology*, 83(2), 239 https://doi.org/10.1037/a0038402.

Pritchard, D., Hoerger, M., & Mace, F. C. (2014). Treatment relapse and behavioral momentum theory. *Journal of Applied Behavior Analysis*, 47(4), 814–833. https://doi.org/10.1002/jaba.163.

Romano, E., Weegar, K., Babchishin, L., & Saini, M. (2018). Cross-informant agreement on mental health outcomes in children with maltreatment histories: A systematic review. *Psychology of Violence*, 8(1), 19–30. https://doi.org/10.1037/vio0000086.

Sachser, C., Berliner, L., Holt, T., Jensen, T. K., Jungbluth, N., Risch, E., Rosner, R., & Goldbeck, L. (2017). International development and psychometric properties of the Child and Adolescent Trauma Screen (CATS). *Journal of Affective Disorders,*, 210, 189–195. https://doi.org/10.1016/j.jad.2016.12.040.

Snijders, T., & Bosker, R. J. (2012). *Multilevel models*. Thousand Oaks, CA: Sage.

Sunderland, M., Wong, N., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour Research and Therapy*, 50(6), 374–380. https://doi.org/10.1016/j.brat.2012.03.005.

Thomas, K. C., Ellis, A. R., Konrad, T. R., Holzer, C. E., & Morrissey, J. P. (2009). County-level estimates of mental health professional shortage in the United States. *Psychiatric Services*, 60(10), 1323–1328. https://doi.org/10.1176/ps.2009.60.10.1323.

U.S. Department of Health and Human Services. (2013, July 11). [Letter to State Medicaid Directors]. http://www.medicaid.gov/Federal-PolicyGuidance/Downloads/SMD-13-07-11.pdf

Van der Meer, M., Dixon, A., & Rose, D. (2008). Parent and child agreement on reports of problem behaviour obtained from a screening questionnaire, the SDQ. *European Child & Adolescent Psychiatry, 17* (8), 491–497. https://doi.org/10.1007/s00787-008-0691-y.

Walker, S. C., Hurvitz, P. M., Leith, J., Rodriguez, F. I., & Endler, G. C. (2016). Evidence-based program service deserts: A geographic information systems (GIS) approach to identifying service gaps for state-level implementation planning. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(6), 850–860. https://doi.org/10.1007/s10488-016-0743-4.

Weems, C. F., Feaster, D. J., Horigian, V. E., & Robbins, M. S. (2011). Parent and child agreement on anxiety disorder symptoms using the DISC predictive scales. *Assessment, 18*(2), 213–216. https://doi.org/10.1177/1073191110389278.

Weisz, J. R., Bearman, S. K., Ugueto, A. M., Herren, J. A., Evans, S. C., Cheron, D. M., & Jensen-Doss, A. (2020). Testing robustness of Child STEPs effects with children and adolescents: A randomized controlled effectiveness trial. *Journal of Clinical Child & Adolescent Psychology, 49*(6), 883–896. https://doi.org/10.1080/15374416.2019.1655757.

Wood, A., Kroll, L., Moore, A., & Harrington, R. (1995). Properties of the mood and feelings questionnaire in adolescent psychiatric outpatients: a research note. *Journal of Child Psychology and Psychiatry, 36*(2), 327–334. https://doi.org/10.1111/j.1469-7610.1995.tb01828.x.

Zahn-Waxler, C. (1993). Warriors and worriers: Gender and psychopathology. *Development and Psychopathology, 5*(1-2), 79–89. https://doi.org/10.1017/S0954579400004272.