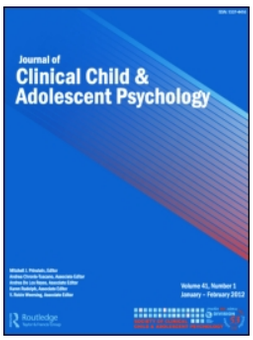

Routledge
Taylor & Francis Group

Journal of Clinical Child & Adolescent Psychology

ISSN: 1537-4416 (Print) 1537-4424 (Online) Journal homepage: http://www.tandfonline.com/loi/hcap20

Initial Test of a Principle-Guided Approach to Transdiagnostic Psychotherapy With Children and Adolescents

John Weisz, Sarah Kate Bearman, Lauren C. Santucci & Amanda Jensen-Doss

To cite this article: John Weisz, Sarah Kate Bearman, Lauren C. Santucci & Amanda Jensen-Doss (2017) Initial Test of a Principle-Guided Approach to Transdiagnostic Psychotherapy With Children and Adolescents, Journal of Clinical Child & Adolescent Psychology, 46:1, 44-58, DOI: 10.1080/15374416.2016.1163708

To link to this article: https://doi.org/10.1080/15374416.2016.1163708

Journal of Clinical Child & Adolescent Psychology, 46(1), 44–58, 2017

ISSN: 1537-4416 print/1537-4424 online
DOI: 10.1080/15374416.2016.1163708

Routledge
Taylor & Francis Group

CROSS-CULTURAL RESEARCH AND IMPLEMENTATION SCIENCE

Initial Test of a Principle-Guided Approach to Transdiagnostic Psychotherapy With Children and Adolescents

John Weisz
Department of Psychology, Harvard University

Sarah Kate Bearman
Department of Educational Psychology, University of Texas at Austin

Lauren C. Santucci
Department of Psychology, Harvard University

Amanda Jensen-Doss
Department of Psychology, University of Miami

To address implementation challenges faced by some evidence-based youth psychotherapies, we developed an efficient transdiagnostic approach—a potential "first course" in evidence-based treatment (EBP)—guided by five empirically supported principles of therapeutic change. An open trial of the resulting FIRST protocol was conducted in community clinics. Following a 2-day training, staff practitioners treated 24 clinically referred youths ages 7–15, 50% male, 87% White and 13% Latino, all with the Schedule for Affective Disorders and Schizophrenia for School-Age Children *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) anxiety, depressive, or conduct-related disorders, and averaging 2.21 disorders. We evaluated the protocol's (a) feasibility for use in everyday clinical practice (examining therapy process, client engagement, and therapist adherence and competence in using the protocol), (b) acceptability (examining therapeutic alliance and treatment satisfaction by youths, caregivers, and therapists), and (c) potential for clinical benefit (examining treatment outcomes across multiple measures and time points). FIRST scored well on measures of feasibility, acceptability to clients and clinicians, and clinical outcomes, matching or exceeding the corresponding scores in most benchmarking comparisons. Observational coding of sessions showed high levels of protocol adherence (86.6%) and good therapist competence in the evidence-based skills. Weekly assessments throughout treatment showed effect sizes for clinical improvement ranging from .41 to 2.66 on weekly total problems and problems deemed "most important" by caregivers and youths. The FIRST protocol showed evidence of feasibility, acceptability, and clinical benefit when used by practitioners with referred youths treated in community clinics. The findings suggest sufficient potential to justify a full randomized controlled trial of FIRST.

A notable success story in clinical science has been the emergence of evidence-based psychotherapies (EBPs) for children and adolescents (herein "youths"). Over the past 50 years, scores of manual-guided youth protocols have shown beneficial effects in research trials, and dozens of these now meet criteria for the status of "evidence-based

Correspondence should be addressed to John Weisz, Harvard University, Department of Psychology, William James Hall, 33 Kirkland Street, Cambridge, MA 02138. E-mail: john_weisz@harvard.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hcap.

psychotherapies" (National Registry of Evidence-Based Programs and Practices [http://www.nrepp.samhsa.gov/]; Silverman & Hinshaw, 2008; Southam-Gerow & Prinstein, 2014). Meta-analyses and systematic reviews (e.g., Piquero, Farrington, Welsh, Tremblay, & Jenkins, 2008; Verdeli, Mufson, Lee, & Keith, 2006; Weisz, Weiss, Han, Granger, & Morton, 1995) have shown significant mean effect sizes for tested youth therapies, with most means in the 0.5 to 0.8 range—*medium* to *large* effects by Cohen's (1988) standards.

Most of the youth EBPs are *focal treatments*, that is, interventions designed for use with one specific disorder or problem domain, or a homogeneous cluster (e.g., one protocol for obsessive-compulsive disorder, another for conduct problems). A case can be made that these focal EBPs are the treatments of choice for their specific target conditions and that an important goal for our field is to increase practitioner expertise in, and thus community access to, these treatments. There is broad support for this goal, within the United States (e.g., Comer & Barlow, 2014; U.S. Surgeon General, 2004) and internationally (e.g., Gardner, Montgomery, & Knerr, 2015).

Although this goal is widely endorsed, efforts at broad implementation face certain challenges:

1. *Implementation burden*. Effective implementation on a broad scale requires significant time and funding, for the clinician training, case consultation, and skill building that are needed, particularly given the number of focal EBPs that would be required to address the heterogeneous caseloads found in everyday practice (Bearman & Weisz, 2015). Funding for training is quite limited in many practice settings, and productivity requirements often limit the time available for any unbillable activity. A recent report on practitioners' perspectives identified the need to learn multiple EBPs to address caseload demands as a major barrier to implementation (Powell, Hausmann-Stabile, & McMillen, 2013).
2. *Co-occurring problems and comorbidity*. Focal EBPs may not be a perfect match to the structure of psychopathology in most clinically referred youths, for whom problem co-occurrence and diagnostic comorbidity are common (Angold, Costello, & Erkanli, 1999; Hogue & Dauber, 2013; Merikangas et al., 2010). Experts in focal EBPs *do* often treat comorbid youths, but the treatments are typically not designed to address the comorbid conditions, and clients with potentially interfering comorbidities have often been excluded from trials of focal EBPs (Bearman & Weisz, 2015). A challenge for everyday implementation is that co-occurring conditions often warrant treatment in their own right.
3. *Flux during treatment*. Youth treatment needs in everyday practice may shift during episodes of care. Changes in problem severity, functional impairment, or environmental conditions during treatment may lead to changes in what is most pressing, or "primary" (Weisz, Krumholz, Santucci, Thomassin, & Ng, 2015). Focal EBPs often entail a series of sessions delivered in a prescribed order, typically built on sound logic in which early level skills build a foundation for more complex skills. This logic, although compelling, may not work so well when treatment needs shift markedly—as when a youth being treated for depression develops serious conduct problems that must be addressed during the treatment episode.
4. *Insufficient dose*. Youths in clinical care often fail to complete the course of treatment their therapist had planned. When focal EBP protocols require completion of multiple logically sequenced sessions, many youths fail to complete the protocol (see Zhou et al., 2015), leaving application of the EBP incomplete and potentially violating the logic of the treatment models. As an example, in a recent RCT with referred youths in community practice settings (Weisz et al., 2012), clinicians in one of the treatment conditions used a focal CBT protocol (16 sessions) to treat anxiety, another focal CBT protocol (20 sessions) to treat depression, and a focal behavioral parent training protocol (15 sessions) to treat conduct-related problems. In this study, 13% of anxiety cases, 33% of depression cases, and 52% of conduct cases completed less than half the number of sessions prescribed in the protocol; only 65% of the anxiety cases, 60% of depression cases, and 20% of conduct cases completed 90% or more of the prescribed sessions. In such situations, many youths may miss the dose prescribed in the EBP protocol, and perhaps the dose required to generate beneficial effects.
5. *Practitioner choice*. Even when practitioners learn to deliver focal EBPs with sessions in a prescribed order—as in treatment trials or training initiatives—the use of the full EBP as prescribed tends to break down afterward, when the clinicians are free to choose (Adelman & Taylor, 2003; Chu et al., 2015; Rye & Kimberly, 2007; Simpson & Flynn, 2007). Practitioners often gravitate toward using some of the treatment components, flexibly on an as-applicable basis, without regard to the order prescribed in the EBP protocol (see Chu et al., 2015).

To summarize, although focal EBPs are excellent in many ways, they may face implementation challenges when applied in everyday clinical practice, particularly by clinicians who have not had previous EBP experience. It may thus be useful to complement focal EBPs with approaches designed to address barriers in everyday clinical care. Based on the preceding review, such a complementary approach might include the following characteristics:

1. Brevity and efficient learnability. To address the challenge faced by clinicians who have limited funding and limited time for new learning, the approach might consist of a manageable amount of content related to evidence-based intervention, to permit busy clinicians to learn it efficiently, without high cost or serious loss of productivity.
2. Caseload and comorbidity coverage. The evidence-based content should be appropriate for multiple disorders in a typical clinician's caseload, and for common forms of comorbidity. For youth treatment this should encompass both internalizing and externalizing problems for which psychotherapies have been shown to be effective.
3. Flexibility for flux. The approach should permit informed and appropriate shifts in focus in response to changes in treatment needs during episodes of care.
4. Stand-alone elements and skills. To permit selective use by clinicians, and to benefit youths who do not stay in treatment very long, the approach should ideally include components that have the potential for beneficial effects as solo interventions.
5. Informed by practitioner and researcher feedback. To enhance prospects for sustained use by practitioners, and effectiveness with referred youths, the approach should reflect guidance from practitioners on what features will work in everyday practice and from treatment researchers on what will be effective.

We designed a treatment protocol with these five characteristics in mind. To do so, we shifted from a focus on detailed descriptions of many specific treatment procedures to a focus on broad principles of therapeutic change. Using research literature reviews (e.g., Silverman & Hinshaw, 2008; Weisz, Hawley, & Jensen Doss, 2004), we identified five core principles that (a) are often included in youth EBPs for common internalizing and externalizing disorders, (b) have been shown to be efficacious as stand-alone interventions, (c) can be applied to multiple youth problems that often co-occur, and (d) can be learned efficiently. The nature of the evidence base led us to draw most heavily from behavioral and cognitive-behavioral models of intervention.

Our principle-guided protocol joins other approaches to transdiagnostic treatment. Examples include the "unified protocol" for treatment of emotional disorders across the affective-anxiety spectrum in adults (Barlow et al., 2011), and modified for children (Ehrenreich-May & Bilek, 2012); other youth protocols that address internalizing disorders (Chu, Merson, Zandberg, & Areizaga, 2012; Weersing, Rozenman, Maher-Bridge, & Campo, 2012); and Modular Approach to Therapy for Children (MATCH; Chorpita & Weisz, 2009), which integrates 33 treatment procedures derived from EBPs for youth disorders and problems in the anxiety, depression, and misconduct domains. Each of these protocols meets some, but not all, of the goals described in the introduction. Except for MATCH, the transdiagnostic protocols address only internalizing problems, and MATCH is lengthier and more complex, with 33 treatment modules, requiring more training (6 days) than may be feasible for some clinicians.

We viewed our new protocol as a potential "first course" in evidence-based practice, one that might complement other transdiagnostic treatments by offering a simpler design and fewer detailed instructions, paired with a conceptual approach to learning that might enhance training efficiency. The protocol, called FIRST, was organized around five well-tested principles of youth psychotherapy. It was reviewed by practitioners and treatment researchers and refined based on their feedback. We thought the brevity and efficiency of FIRST might make it a useful addition but might also carry a risk: Shifting from a long list of procedures with detailed instructions to a short list of core principles more briefly described might be insufficient to meet clinicians' needs. This could undermine the feasibility, acceptability, or clinical effectiveness of FIRST. As an initial step toward finding out, we tested FIRST in an open trial.

To provide evidence on FIRST, we evaluated (a) its feasibility for use in everyday clinical practice (examining therapy process, client engagement, and therapist adherence and competence in using the protocol), (b) its acceptability (examining therapeutic alliance and treatment satisfaction by youths, caregivers, and therapists), and (c) its potential for clinical benefit (examining treatment outcomes across multiple measures and time points). Because MATCH (Chorpita & Weisz, 2009, as just noted) is similar to FIRST in the youth problem areas addressed, but much more detailed and comprehensive, we compared our findings in the present study to findings from MATCH research; we made benchmarking comparisons to other research, as well, where measures and samples were comparable to those used in this study.

METHOD

Protocol Development and Refinement

Our review of the youth psychotherapy research literature and of systematic research reviews (e.g., Silverman & Hinshaw, 2008; Weisz et al., 2004) led to identification of five core principles of evidence-based practice with youths. These principles appear in multiple tested interventions for different problems and disorders, and each has shown significant effects when used alone (Weisz et al., 2004). In addition, each of these principles can be readily applied to treatment of anxiety, depression, and conduct problems. The five principles are as follows:

- **Feeling calm**. This is self-calming and relaxation, including both progressive muscle relaxation and

quick calming techniques for reducing short-term situational tension and the accompanying emotional arousal that can be associated with anxiety, depression, or conduct problems.
- **Increasing motivation**. Using environmental contingencies to make adaptive behavior more rewarding than maladaptive behavior, in the context of anxiety, depression, or conduct-related dysfunction. Motivation may be enhanced via such standard principles as differential attention and judicious use of praise and tangible rewards.
- **Repairing thoughts**. Identifying and changing biased or distorted cognitions. The cognitions may be overestimates of threat in anxiety, unrealistically pessimistic or self-blaming thoughts in depression, or bias toward perceiving hostile intent in others in the case of conduct-related disorders and problems.
- **Solving problems**. Building skill in sequential problem solving (e.g., identify the problem, set a goal, think of solutions, weigh the pros and cons, etc.). The skills may be applied to problems arising from anxiety, depression, or conduct-related dysfunction.
- **Trying the opposite**. Engaging in activities that directly counter the behavioral problem. Examples include graduated exposure for anxiety, behavioral activation for depression, and practicing adaptive responses to interpersonal conflict in the context of conduct-related problems.

We integrated these principles to create an initial manual prototype, then refined it based on feedback from a team of seven community practitioners (80% female; 57% doctoral and 43% master's level) who were experienced in youth psychotherapy; this involved six 3-hr collaborative meetings spanning 6 months. The practitioners read draft materials prior to each meeting, discussed and critiqued them during the meetings, then reviewed revised drafts prior to the next meeting. After the final clinician feedback, we submitted the draft to five prominent treatment researchers for their feedback; all offered detailed comments, and we modified the draft based on their input.

Structure of the Protocol

The resulting FIRST protocol was designed for individual administration to youths, caregivers, and youth–caregiver combinations as indicated. The protocol included an initial overview of the five principles and their potential role in treatment of youth anxiety, depression, and conduct problems. This was followed by guidelines to assist clinicians in beginning treatment; a collection of relevant information from both youth and caregiver to determine initial treatment target; an understanding of the nature of youth anxiety, depression, and conduct problems; and materials to help clients and caregivers understand the nature of these three problem areas and the need for intervention. Subsequent sections focused, in order, on each of the five psychotherapy principles, with discussion and examples of how each could be applied to anxiety, depression, and conduct problems. An appendix focused on special cases (e.g., how to address panic disorder or obsessive-compulsive disorder [OCD] within FIRST), frequently asked questions, and additional treatment resources for the clinician (e.g., sample fear hierarchies, sample reward system plans), and a research bibliography. A decision tree was used together with weekly problem reports from youths and caregivers (next) to guide such decisions as how to sequence treatment steps and whether or when to shift from one principle or treatment focus to another.

Youth and Caregiver Participants

Participants were 24 youths who had been referred for treatment through normal community pathways to one of two urban community mental health clinics in the northeastern United States. The study was approved by the Institutional Review Board, with informed consent and assent obtained by research staff, and compensation was provided to participants for their time completing study measures, as per Institutional Review Board review. Figure 1 shows the steps leading to enrollment.

Youth age ranged from 7.11 to 15.40 (M = 11.03, SD = 2.69), reflecting our effort to carry out FIRST treatment across a relatively broad developmental range. The lower end of the range was based on the youngest age level at which study self-report measures were deemed appropriate; the upper end was based on the approximate oldest age at which significant parental involvement is a common part of outpatient youth psychotherapy. Half the sample members were boys; the ethnic composition was 87% White and 13% Latino. One youth lived with adoptive parents and 23 with biological parents. Annual family income was below $40,000 for 25% of the sample, $40,000–$79,000 for 29%, $80,000–$119,000 for 8%, and $120,000 or more for 25% (three families did not report their income). Some 54% of caregivers were married, 42% divorced/separated, and 4% never married. Mean total problems T score was 60.46 (SD = 9.47) on the Child Behavior Checklist (CBCL) and 55.96 (SD = 11.10) on the Youth Self-Report (YSR; see Table 1). All youths met criteria for at least one *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM-IV*]; American Psychiatric Association, 1994) disorder based on a standardized diagnostic interview (see next), and the mean number of disorders was 2.21 (SD = 1.18; see Tables 1 and 2).

Participating Therapists and Clinics

Study therapists were 14 licensed practitioners from two community mental health clinics. Both clinics served youths

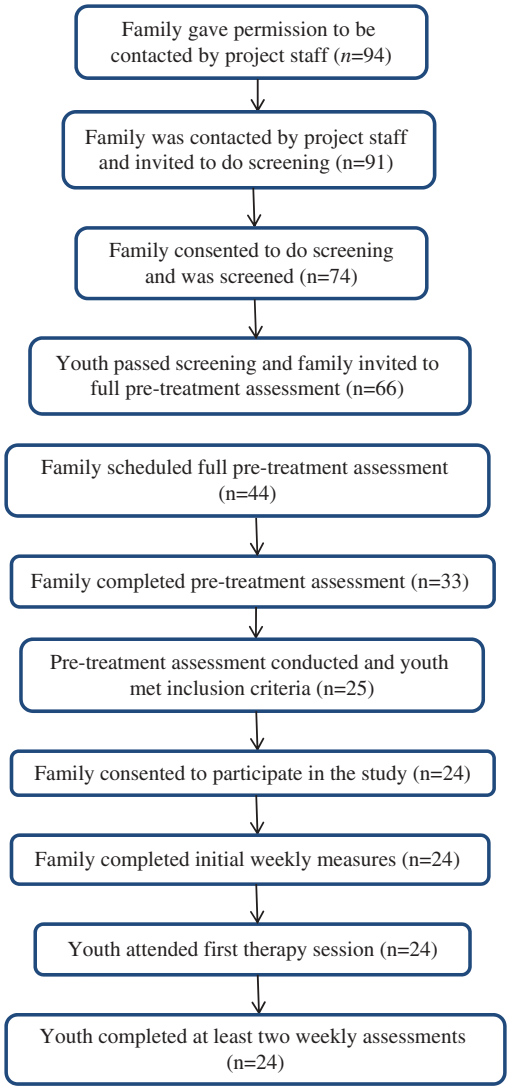

FIGURE 1 Participant enrollment flowchart.

with a broad range of referral problems. Therapists were 79% female, 93% Caucasian, and 7% Latino, with mean age of 35.57 (*SD* = 8.69) and mean years of clinical experience at 7.96 (*SD* = 9.16). Some 43% were psychologists, 50% social workers, and 7% Licensed Mental Health Counselors. The therapists reported mainly eclectic orientations, with all indicating they used multiple theoretical orientations for both case formulation and therapy.

Experimental Design and Study Procedure

The open trial (pre/posttreatment design) was structured to assess how FIRST would function within an outpatient community clinic context, including its feasibility and acceptability, and whether changes in clinical measures from pre- to posttreatment might suggest potential for clinical benefit and thus readiness for a full randomized controlled trial (RCT). Because FIRST was comparable in its problem coverage (anxiety, depression, and conduct) to the more detailed and extensive MATCH program (Chorpita & Weisz, 2009), we used data from previous MATCH studies (Chorpita et al., 2015; Weisz et al., 2012, 2016) as benchmarks to help interpret the present findings; missing comparisons, in the Results section, indicate data for which no benchmarks were available from MATCH research.

Youths were included only if they met full diagnostic criteria for one or more disorders within the anxiety, depression, or conduct-related clusters (based on the K-SADS; Kaufman et al., 1997; see next). Because evidence (e.g., MTA Cooperative Group, 1999, 2004) suggests that stimulant medication produces larger acute effects than psychological interventions for the core symptoms of attention deficit/hyperactivity disorder (ADHD), and because the principles included in FIRST were not expected to ameliorate attention deficits, impulsivity, or overactivity, we excluded youths for whom ADHD was identified as the primary disorder (but we included youths with ADHD if they met criteria for an anxiety-, depression-, or conduct-related disorder that was identified as primary).

Clinic reception and intake staff were asked to invite families who inquired about treatment for their 7- to 15-year-old child to be contacted about the study by our staff. Families who gave permission were contacted by project staff and screened by phone. Those who met initial inclusion criteria were invited to a project interview. If the K-SADS in that interview identified only one disorder, and it was an anxiety, depressive, or conduct-related disorder, the family was invited to enroll in the trial. If the K-SADS identified additional disorders, treatment priority was assessed, using K-SADS symptom counts and interference-impairment ratings, plus other clinical measures in the measurement model (see next). Project staff and family members discussed these data to reach a consensus on which disorder should be the initial focus of treatment. This focus could change later if the most pressing treatment needs shifted during treatment. Such decisions were informed by weekly data on youth treatment response (see Measurement Model section).

Measurement Model

Project measures were selected to provide evidence bearing on (a) the feasibility of FIRST for use in everyday clinical practice (measures of therapy process, client engagement, and therapist adherence to the protocol), (b) acceptability (measures of therapeutic alliance, and treatment satisfaction by youths, caregivers, and therapists), and (c) potential for mental health benefit (measures of outcome across multiple clinical measures and time points).

TABLE 1
Descriptive Statistics and Outcome Analyses for Measures of Clinical Benefit

	Pre	*Post*	*Slope Estimates*[a]			
Measure	*Treatment M (SD)*	*Treatment M (SD)*	*Slope*	*95% CI*	*p*	*ES*
Primary Clinical Measures[b]						
BPC Total – Parent	9.04 (4.61)	6.25 (5.74)	−.80	−1.18, −.41	< .001	.66
BPC Total – Youth	6.12 (4.57)	4.70 (5.01)	−.35	−.61, −.089	.013	.41
BPC Internalizing - Parent	4.62 (3.44)	3.58 (3.59)	−.42	−.68, −.16	.004	.29
BPC Internalizing - Youth	2.92 (2.77)	2.74 (3.53)	−.12	−.30, .060	.124	.11
BPC Externalizing - Parent	4.42 (2.98	2.67 (2.93)	−.38	−.57, −.19	.001	.67
BPC Externalizing - Youth	3.21 (2.48)	1.96 (1.89)	−.22	−.40, −.040	.023	.62
TPA – Parent	7.85 (1.78)	6.25 (5.74)	−.78	−1.04, −.32	< .001	1.91
TPA – Youth	7.54 (1.29)	4.70 (5.01)	−.72	−1.07, −.37	.001	2.66
CGI	3.80 (.41)	2.09 (1.23)	−.68	−.85, −.51	< .001	3.43
Secondary Clinical Measures						
K-SADS – No. of Diagnoses	2.21 (1.38)	.88 (1.32)	−1.29	−2.01, −.53	.009	.96
CBCL - Total Problems	60.46 (9.47)	50.94 (11.00)	−8.68	−14.24, −3.12	.012	.85
YSR - Total Problems	55.96 (10.58)	43.16 (10.86)	−12.62	−17.60, −7.64	< .001	1.15
CBCL - Internalizing	62.17 (11.98)	56.18 (11.95)	−6.08	−11.63, −.53	.038	.51
YSR – Internalizing	57.26 (11.96)	46.00 (11.78)	−11.18	−17.61, −4.75	.005	.94
CBCL – Externalizing	56.58 (11.40)	47.41 (7.42)	−8.73	−15.29, −2.17	.021	.93
YSR – Externalizing	50.96 (10.58)	43.16 (10.86)	−8.62	−13.98, −3.27	.008	.94

Note: Effect sizes are computed as estimated pre–post Cohen's *d* values. BPC = Brief Problem Checklist (Chorpita et al., 2010); Top Problems = Top Problems Assessment (TPA; Weisz et al., 2011); K-SADS = Schedule for Affective Disorders and Schizophrenia for School-Age Children (Kaufman et al., 1997); CGI = Clinical Global Impression-Improvement (Guy, 1976); CBCL = Child Behavior Checklist (Achenbach & Rescorla, 2001); YSR = Youth Self Report (Achenbach & Rescorla, 2001).

[a]Results of multilevel modeling analyses estimating the change per log day in the primary clinical measures and pre–post change for the secondary clinical measures.

[b]Primary clinical measures were administered weekly throughout treatment. Means and standard deviations are presented for the first and last assessment points.

Feasibility Measures

Family engagement. Engagement measures included (a) percentage of scheduled sessions attended, attended on time, cancelled, and missed due to no-show and (b) whether treatment was terminated as planned with therapist agreement versus against therapist advice.

Therapist treatment integrity. Treatment integrity—including therapist adherence and therapist competence in delivering the FIRST protocol—was assessed using the Therapist Integrity in Evidence Based Interventions coding system for FIRST (TIEBI; Bearman, Herren, & Weisz, 2012; adapted from a coding system used in Weisz et al., 2012). In this system, treatment session recordings are coded, in 5-min segments, for the presence/absence of 19 items reflecting FIRST content. Coding includes therapist adherence (based on the percentage of 5-min segments in which prescribed content from FIRST was present) and therapist competence (skillfulness of delivery, rated from 0 [*not at all*] to 4 [*expert*]). TIEBI coders (N = 7) were bachelor's- and master's-level research assistants in the primary coding system developer's laboratory. Coding training consisted of four steps. Coders first observed a full clinician training for the treatment of anxiety, depression, and conduct, which covered all therapeutic content included in the FIRST protocol. Next, coders attended a didactic training in which both the FIRST manual and the TIEBI coding manual were reviewed in detail, and in which each item in the coding manual was paired with exemplar sections of session recordings. Following the didactic training, coders independently coded three sessions and met together with the coding system developer to review each session. Finally, all coders independently coded the same six sessions, and reliability was assessed against master codes produced by the first two authors of the coding system. To be permitted to code independently for the study, coders had to demonstrate mean item agreement for both adherence and competence that was above the threshold for "good" reliability (intraclass correlation coefficient [ICC] > .59), according to the standards recommended by Cicchetti (1994). All coders scored in the good-to-excellent range for both adherence, M ICC [2, 2] = .83 (range = .65–.90) and competence, ICC [2, 2] = .78 (range = .66–.90) on this sample, and thus qualified to code the remainder of the sessions randomly assigned to them.

All client treatment sessions were video-recorded throughout the study. Sessions in each treatment episode were randomly selected for coding using the following procedures: (a) First sessions were omitted (these often included clinic administrative content), (b) all remaining

TABLE 2
Diagnostic and Statistical Manual of Mental Disorders (4th ed.)
Diagnostic Composition of Sample

Diagnosis	*Primary, No. (%)*	*Anywhere, No. (%)*
ADHD (Any Type)	0	16 (66.6)
ADHD, Combined type	0	6 (25.0)
ADHD, Predominantly Inattentive Type	0	3 (12.5)
ADHD NOS	0	7 (29.2)
ADHD, Predominantly Hyperactive-Impulsive Type	0	0
Adjustment Disorder (Any Type)	0	1 (4.00)
Anxiety Disorder (Any Type)	10 (41.6)	14 (58.3)
Specific Phobia	0	4 (16.7)
Separation Anxiety Disorder	2 (8.3)	3 (12.5)
Generalized Anxiety Disorder	3 (12.5)	9 (37.5)
Social Phobia	2 (8.3)	4 (16.7)
Obsessive-Compulsive Disorder	2 (8.3)	2 (8.3)
Posttraumatic Stress Disorder	0	0
Panic Disorder Without Agoraphobia	0	1 (4.16)
Anxiety Disorder NOS	1 (4.2)	5 (20.8)
Conduct-Related Disorder (Any Type)	11 (45.8)	15 (62.5)
Oppositional Defiant Disorder	5 (20.8)	5 (20.8)
Conduct Disorder	0	0
Disruptive Behavior Disorder NOS	6 (25.0)	10 (41.7)
Elimination Disorder	0	1 (4.2)
Mood Disorder (Any Type)	3 (12.5)	5 (20.8)
Major Depressive Disorder	1 (4.2)	3 (12.5)
Dysthymic Disorder	0	0 (0.00)
Depressive Disorder NOS	2 (8.3)	2 (8.3)
Bipolar Disorder	0	0 (0.00)
Psychosis	0	1 (4.2)
Autism Spectrum Disorder	0	1 (4.2)
Tic Disorder	0	1 (4.2)

Note: ADHD = attention deficit/hyperactivity disorder; NOS = not otherwise specified.

sessions were divided into thirds (early, middle, late phase of treatment), and (c) 30% of audible sessions within each phase were randomly selected to be coded. If three or fewer sessions were available, all were coded. Sessions shorter than 15 min or longer than 75 min were omitted (these were typically unrepresentative, e.g., clinic paperwork). Of the 360 recordings, 107 were randomly assigned by session phase to the seven coders, who were blind to study purposes. To generate a mean adherence score for each coded treatment session for each FIRST content item, the number of 5-min segments in which that item was coded "present" was summed, multiplied by 5 (number of minutes in each segment), and divided by the total time of the session in minutes. The resulting means were averaged across all the coded sessions for each youth's treatment episode. To generate a mean competence score for each youth's treatment episode for each FIRST content item, each of the global compentency codes for each "present" item were averaged across all global competency codes in a given session, and these means were, in turn, averaged across all coded sessions in the treatment episode.

Over the course of the coding, 31 sessions (29%) were randomly selected for double coding to assess agreement between independent coders. Each session was coded by a random subset of two raters, intended to represent the larger population of similar raters, and reliability estimates corresponded to a single measurement. Average reliability was in the excellent range for both adherence, *M* ICC [1, 1] = .87 (range = .53–.99), and competence, *M* ICC [1, 1] = .88 (range = .63–.97). We compared FIRST findings on both measures to findings with parallel TIEBI measures employed to code the modular MATCH program used by community practitioners with their young clients in two separate RCTs (both including the standard 6-day clinician training, and weekly clinician case consultation with MATCH experts). One of these RCTs (Weisz et al., 2012) had used the TIEBI to assess MATCH adherence (309 sessions coded), the other (Weisz et al., 2016) had used the TIEBI to assess MATCH competence ($N = 391$ sessions coded). The coding team and training procedures were the same for the FIRST findings and for the second MATCH RCT; for the first MATCH RCT, a different coding team was used, but training procedures were identical. In both studies, coders were unaware of study hypotheses and study condition.

Measures of Treatment Acceptability

Youth and parent-reported therapeutic alliance.

Alliance was assessed via the Therapeutic Alliance Scale for Youths (TASC-C) and Parents (TASC-P). The TASC-C has shown good internal consistency (α = .84), and 7- to 14-day test–retest reliability (r = .65) in samples of clinic-referred youths (Shirk & Saiz, 1992). The TASC-P has also shown good internal consistency (α = .92) and 7- to 14-day test–retest reliability (r = .82) in parents of clinic-referred youth (Shirk & Saiz, 1992). In the current study, the TASC-C and TASC-P showed alphas of .78 and .83, respectively.

Youth and caregiver satisfaction with treatment.

The Client Satisfaction Questionnaire (CSQ-8; Larsen, Attkisson, Hargreaves, & Nguyen, 1979) and Youth Satisfaction Questionnaire (YSQ) were completed by caregivers and youths, respectively, to assess satisfaction with treatment received. The eight-item caregiver measure (CSQ-8; sample item: "In an overall, general sense, how satisfied are you with the service you have received?") has been found to be substantially correlated with treatment dropout, number of therapy sessions attended, and change in client-reported symptoms (Attkisson & Zwick, 1982). The eight-item youth measure (YSQ) was developed using existing satisfaction measures as a guide (e.g., CSQ-8; Larsen et al., 1979) but using youth-appropriate language (sample item: "Do you think the help you got here will make things better for you later on?"). CSQ-8 and YSQ

items use a 4-point response scale with total scores ranging from 8 to 32 (higher scores indicate greater satisfaction). In the current sample, the parent and youth measures showed Cronbach's alphas of .91 and .84, respectively.

Therapist satisfaction with treatment. Each time treatment ended for a project case, the therapist completed the Therapist Satisfaction Inventory (TSI; Chorpita et al., 2015) for that case. The TSI is a 16-item therapist-report measure derived partly from Addis and Krasnow's (2000) report on therapist attitudes toward manualized treatments. A study with 145 youths treated by 77 therapists (Chorpita et al., 2015) revealed two factors reflecting perceived responsiveness and perceived effectiveness of treatment (Cronbach's α = .83 for total score, .82 for responsiveness, .81 for effectiveness). Total TSI satisfaction scores were found to be higher for cases in which therapists used a flexible modular treatment program based on evidence-based elements than for cases in which usual care was employed; other comparisons showed significant differences on the Responsiveness and Effectiveness subscales. Cronbach's alpha for the current sample was .79 for total score, .72 for responsiveness, and .77 for effectiveness.

Measures of Clinical Benefit

Primary clinical measures: Weekly symptom, target problem, and clinical ratings. The Brief Problem Checklist (BPC; Chorpita et al., 2010) is a 12-item measure of internalizing (six items; scores range 0–12), externalizing conduct (six items; range = 0–12), and total problems, developed by applying item response theory and factor analysis to large samples of data from the CBCL and YSR. In a psychometric study with 184 clinic-referred 8- to 13-year-olds (Chorpita et al., 2010), the 12-item total score showed alphas of .82 for parents and .76 for youths, and correlations between scores on corresponding BPC and CBCL/YSR scales were substantial (all > .57). The BPC was used for weekly tracking of problem trajectories. (Note: The BPC has now been replaced by the Brief Problem Monitor, which includes ADHD symptoms [Achenbach, McConaughy, Ivanova, & Rescorla, 2011], but the BPC was included here to permit benchmarking with Weisz et al. [2012], in which the BPC was used.)

The Top Problems Assessment (TPA; Weisz et al., 2011) entails youth and parent severity ratings (on a scale of 0–10) of the top three problems the youth and parent independently identified as most important to them in separate structured pretreatment interviews. Psychometric analyses in a sample of 178 clinic-referred youths showed substantial evidence of test–retest reliability (.69–.91, across 5- to 21-day intervals), convergent and discriminant validity (in relation to standardized problem-report measures), and sensitivity to clinical change during treatment (Weisz et al., 2011).

The 7-point Clinical Global Impression–Improvement (CGI) Scale (Guy, 1976) was completed by practitioners each week to provide their rating of client change relative to baseline severity: 1 [*very much improved*] to 7 [*very much worse*]. The CGI has been used to assess treatment response in studies testing treatments for youth anxiety disorders (e.g., Walkup et al., 2008), youth depression (e.g., TADS Team, 2004), and youth ADHD (Swanson et al., 2001). The measure has shown convergent validity between clinicians and independent assessors ratings of youth response to treatment (κ = .58; Kratochvil et al., 2006).

Secondary clinical measures: Pre- and posttreatment diagnoses and symptoms. The K-SADS (Kaufman et al., 1997) was used to diagnose *DSM-IV* disorders. Research on the K-SADS has shown acceptable test–retest reliability (κs = .60–1.00) and interrater reliability (κs = .60–1.00), and discrimination between disordered and nondisordered youths (Ambrosini, 2000; Kaufman et al., 1997; Lewinsohn et al., 1994). Interviewers (minimum postbachelor's) received 24 hr of training, with interview simulation, observation, and feedback. Next, they reviewed and scored six prerecorded interviews and were required to show minimum kappa of .80 with expert raters before beginning data collection. A refresher training workshop was conducted 6 months after the initial training. Assessors audio-recorded all their interviews, and a random sample was monitored by supervisors throughout the study to prevent drift.

The CBCL (Achenbach & Rescorla, 2001) obtains parent reports on 118 youth/adolescent problems, each rated on a 0–1–2 scale (2 = *very true* or *very often true*). Age- and gender-adjusted *T* scores for the broadband Internalizing and Externalizing scales, and Total Problems, were used as outcome measures. CBCL validity and reliability evidence is strong and extensive (Achenbach & Rescorla, 2001).

The YSR (Achenbach & Rescorla, 2001) is a 112-item youth-report counterpart to the CBCL. As with the CBCL, we used *T* scores for Internalizing, Externalizing, and Total Problems. YSR validity and reliability evidence is strong and extensive (Achenbach & Rescorla, 2001).

Pretreatment and weekly medication monitoring. Medication questions from the Services for Youths & Adolescents–Parent Interview (SCAPI; Hoagwood et al., 2004; Jensen et al., 2004) were used to assess prescription medications for behavioral and emotional problems at pre- and posttreatment and weekly during treatment. The SCAPI was administered to parents at each time point, except when they reported no change in medications. The SCAPI has

been found to have acceptable test-–retest reliability (κ = .98; Hoagwood et al., 2004) and to discriminate participants in a randomized trial (MTA) from a local control group with regard to total daily medication dose (Jensen et al., 2004). Interviewers (minimum postbachelor's) received 2 hr of training, followed by interview simulation, observation, and feedback.

Therapist Training, Treatment, and Study Procedures

FIRST therapists completed all routine clinic procedures (including paperwork) and conducted treatment on clinic premises following usual clinic routines, procedures, and requirements. At the beginning of the project, the therapists received one 2-day (14-hr) training. The training combined didactics, modeling, video illustrations, and practitioner role-plays with trainer feedback. After training and throughout the study, therapists joined group consultation meetings 1 hr per week, focused on implementation of FIRST with study cases, and therapists and consultants had web-based access to the weekly youth and caregiver BPC and TPA reports. All treatment involved individual sessions, typically 45- to 60-min long, with youths most often, caregivers less frequently, and sometimes joint youth–caregiver sessions. All sessions were video-recorded, with subsequent coding (see Measurement Model section).

Data Analysis Plan

All study analyses were conducted using SPSS Statistics version 22.0 (IBM Corp, 2013) or HLM 7.01 (Raudenbush, Bryk, Cheong, Congdon, & Du Toit, 2011). Where our data permitted, we used an intent-to-treat data analysis strategy, with all youths included who had been enrolled initially; ITT was not used with measures for which we had only baseline assessments. In reporting findings, we include benchmarking against comparable previous studies where the same measures were reported (summary provided in Table 3).

Feasibility analyses were primarily descriptive in nature. Analyses of trajectories of change across weekly measures during treatment (shown in Figure 2) were structured as random-effects three-level multilevel models, estimating the rate of change over time using all available measurement occasions for each participant accounting for clustering within clinicians (see Hedeker & Gibbons, 2006). These analyses focused on the BPC Total Problems, Internalizing, and Externalizing measures, the TPA top problem severity ratings, and the therapists' GGI-I ratings. Prior to analysis, following Snijders and Bosker (1999), deviance statistics for both linear and loglinear models were compared to intercept-only models for each dependent variable; these analyses suggested either model would be appropriate, so the loglinear models were selected to facilitate comparison to previous trials using these same measures (see, e.g., Table 3 in Weisz et al., 2012). To assess the appropriateness of a three-level modeling approach, the proportion of variance in model intercepts and slopes at the youth and therapist levels were computed. Across outcome variables, between 2.1% and 70.1% of variance in intercept and between 14.4% and 74.2% of the variance in slope were at the clinician level. This information was used to compute the design effect, which takes cluster size into account. Design effect values ranged from 1.02 to 1.74. In studies such as this one, which had an average therapist cluster size of 2, design effects greater than 1.1 are considered nonignorable (Lai & Kwok, 2015), so three level models were used to analyze the data.

No accepted standard for computing effect sizes for models like this exists; effect sizes were computed by using the model parameters to estimate the group mean at 168.28 days (i.e., the average length of treatment), subtracting that estimated mean from the baseline mean, and dividing the difference by the pretreatment standard deviation to estimate a within-subjects effect size (Hedges's g_{av}; Lakens, 2013; a version of Cohen's d for within-subject comparisons), where .20 is considered a small effect, .50 a medium effect, and .80 a large effect (Cohen, 1988). Negative signs were changed to positive, such that more positive effect sizes indicated greater treatment benefit.

Analysis of measures collected only at pre and post (i.e., the KSADS, CBCL, and YSR) were also analyzed via three-level multilevel models. Estimates of changes from pre to post were generated via these models, and effect sizes were computed by dividing the pre–post slope by the pre–standard deviation (with negative signs converted to positive). For both weekly and pre–post measures, all analyses were also conducted a second time, controlling for medication use.

RESULTS

Preliminary Analyses

To assess the impact of attrition on study findings, participants without posttreatment data (n = 4) were compared to those with available data on baseline demographic and clinical characteristics. No significant differences were found between postassessment completers and noncompleters on age, gender, ethnicity, baseline CBCL Internalizing, Externalizing or Total Problem scores, baseline YSR Internalizing, Externalizing or Total Problem scores, or number of parent- or youth-reported K-SADS diagnoses (all ps > .20). Postassessment completers and noncompleters did not differ significantly in their trajectories of change on the weekly outcome measures (see next for a description of the trajectory analyses), except on the CGI, where measure completers showed larger gains during treatment than noncompleters (p = .047). The two

TABLE 3
Benchmarking Comparisons of FIRST Findings With Previous Research Findings

Measures of Treatment Process, Feasibility, Acceptability, and Diagnostic Change				
Measure	*FIRST*	*Usual Care*	*Standard EBPs*	*MATCH*
Training Time Required	2 days	0 days[a]	6 days[a]	6 days[a]
M Treatment Duration	24.06 wk.	39.35 wk.[a]	28.03 wk.[a]	30.02 wk.[a]
Protocol adherence	86.60%	8.47%[a]	92.75%[a]	82.95%[a]
Therapist Competence	2.27	0.93b	NA	2.15[b]
TSI Therapist Satisfaction Total	25.59	23.31c	23.88[c]	25.74[c]
TSI Ther. Sat. Responsiveness	12.31	12.00[c]	10.90[c]	12.28[c]
TSI Ther. Sat. Effectiveness	13.27	11.31[c]	13.00[c]	13.47[c]
% Reduction in K-SADS Dx	58.64%	24.70%[a]	52.55%[a]	59.92%[a]
Slopes of Clinical Change: Brief Problem Checklist and Top Problems Severity				
Measure	FIRST	Usual Care	Standard EBPs	MATCH
BPC Total – Parent	−.80	−.50[a]	−.59[a]	−.94[a]
BPC Total – Youth	−.35	−.44[a]	−.23[a]	−.67[a]
Top Problems – Parent	−.78	−.32[a]	−.54[a]	−.65[a]
Top Problems – Youth	−.72	−.47[a]	−.34[a]	−.61[a]
One-Year Change Estimates: Brief Problem Checklist and Top Problems Severity				
Measure	FIRST	Usual Care	Standard EBPs	MATCH
BPC Total – Parent	−4.72	−2.94[a]	−3.48[a]	−5.55[a]
BPC Total – Youth	−2.06	−2.61[a]	−1.33[a]	−4.04[a]
Top Problems – Parent	−4.60	−1.87[a]	−3.17[a]	−3.84[a]
Top Problems – Youth	−4.01	−2.75[a]	−2.02[a]	−3.57[a]

Note: The table includes only study measures for which benchmarking measure comparisons could be identified in previous research using Modular Approach to Therapy for Children (MATCH), usual care, and standard evidence-based treatment protocols. Protocol adherence and therapist competence were coded from treatment sessions using the Therapist Integrity in Evidence-Based Interventions observational coding system (Bearman et al., 2012). EBPs = evidence-based psychotherapies; wk. = week; TSI = Therapist Satisfaction Inventory (Chorpita et al., 2015); K-SADS = Schedule for Affective Disorders and Schizophrenia for School-Age Children (Kaufman et al., 1997); BPC = Brief Problem Checklist (Chorpita et al., 2010); Top Problems = Top Problems Assessment (Weisz et al., 2011).

[a]Weisz et al. (2012).

[b]Weisz et al. (2016).

[c]Chorpita et al. (2015).

groups also did not differ significantly on measures of treatment feasibility, including process (number of sessions and weeks in treatment), engagement (percentage of sessions attended and attended on time, agreement with termination), or integrity (adherence and competence with FIRST). In terms of acceptability, therapist report of satisfaction with treatment did differ by group, with satisfaction higher for clients who completed the assessments (p = .035); families who did not complete posttreatment assessments provided no data on the youth or parent satisfaction measures, so we were not able to compare groups on these measures.

Feasibility of FIRST for Use by Practitioners in Everyday Clinical Practice

Treatment process. Across the full sample, the mean number of therapy sessions was 18.63 (SD = 13.05, range = 2–59), and mean treatment duration was 24.06 weeks (SD = 16.10), somewhat shorter than the mean of 30.02 for the modular MATCH program in Weisz et al. (2012). On average, sessions lasted 48.6 min (SD = 7.07). Across the 24 treatment episodes, 23 included some involvement of caregivers in sessions; in those cases, 76.2% of sessions included some caregiver involvement.

Engagement. Session attendance was high, with 92.7% of scheduled sessions attended, and 98.9% of these on time. For 68.2% of cases, termination was planned in advance and agreed upon between therapist and client/caregiver.

Treatment integrity. TIEBI adherence coding showed that therapists adhered closely to the evidence-based content of the FIRST protocol. Some 86.6% of the 5-min segments of each session contained that content (M = 86.60% of the 5-min segments, SD = 26.20), somewhat higher than the 82.95% and the 80.14% found for therapists trained in the modular MATCH protocol in two previous studies (Weisz et al., 2012, 2016). TIEBI coding showed mean therapist competence at 2.27 (SD = 0.80), somewhat higher than the means of 2.15 obtained for MATCH in the second RCT noted previously (Weisz et al., 2016).

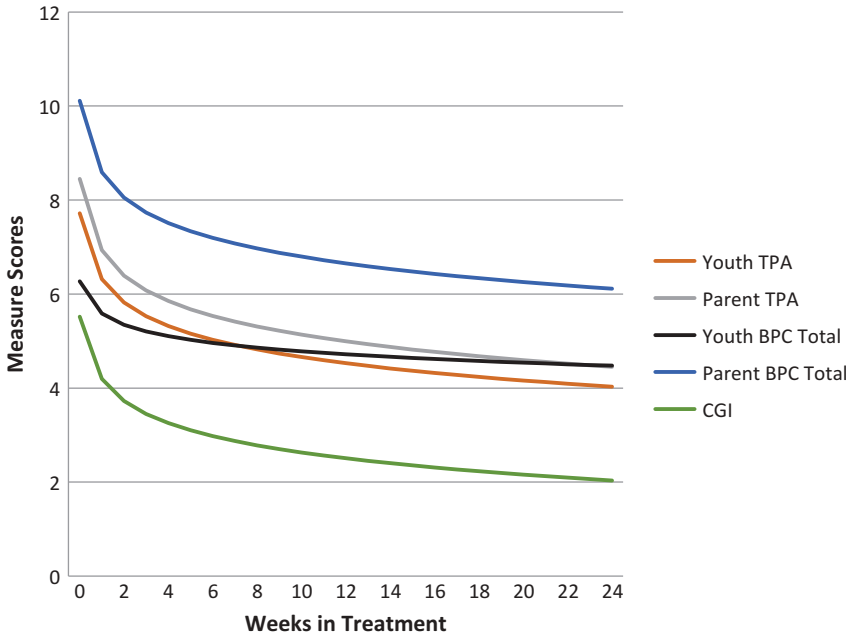

FIGURE 2 Mean change over time in weekly youth, caregiver, and therapist measures. *Note*: TPA = Top Problems Assessments (scores range 0–10, with 10 being most severe); BPC = Brief Problem Checklist (scores range 0–12, with 12 being most severe); CGI = Clinical Global Impression: Improvement (scores range 1 [*very much improved relative to baseline*] to 7 [*very much worse relative to baseline*]).

Acceptability of FIRST

Therapeutic alliance. Caregiver report of therapeutic alliance on the TASC-P was high, with a mean score of 34.00 (*SD* = 2.45) out of a possible 36. Youth alliance on the TASC-C was also relatively high, with a mean of 29.60 (*SD* = 6.00) out of a possible 36.

Treatment satisfaction. Satisfaction with treatment was rated high by caregivers using the CSQ (*M* = 29.11, *SD* = 3.51) and youths using the YSQ (*M* = 28.12, *SD* = 3.79), both out of maximum total scores of 32. The caregiver mean compares favorably to the mean of 26.75 found in a treatment trial using CBT with depressed adolescents exposed to interpersonal trauma (Shirk, DePrince, Crisostomo, & Labus, 2014) and 26.80 found in another treatment trial using CBT to treat children with anxiety disorders (Khanna & Kendall, 2010). Therapist satisfaction with treatment was also high. The mean TSI Total score was 25.59 (*SD* = 3.54) out of a possible 30, very similar to the mean of 25.74 for the MATCH modular program in Chorpita et al. (2015; and that mean was significantly higher than the mean for standard EBPs and for usual care in the same study). The mean TSI Responsiveness scale score was 12.31 (*SD* = 1.94) out of a possible 15, very similar to the mean of 12.28 for MATCH in Chorpita et al. (2015; and that mean was significantly higher than the mean for standard EBPs). The mean TSI Effectiveness scale score was 13.27 (*SD* = 2.21) out of a possible 15, very similar to the mean of 13.47 for MATCH in Chorpita et al. (2015; and that mean was significantly higher than the mean for usual care).

Primary Clinical Measures: Trajectories of Change Across Weekly Assessments

Descriptive statistics and outcome analyses for the primary clinical measures are presented in Table 1. On the BPC Total Problems scale, both parents (slope = −0.80, $p < .001$, ES = .66), 95% CI [–1.18, .41], and youths (slope = –0.35, $p = .013$, ES = .41), 95% CI [–.61, .089], reported significant improvements over the course of treatment. Similarly, both reported significant improvements in the BPC Externalizing scores: parents (slope = –0.38, $p = .001$, ES = .67), 95% CI [–.57, –.19]; youths (slope = –0.22, $p = .023$, ES = .62), 95% CI [–.40, –.040]; however, only parent-reported improvements were significant for the BPC Internalizing scale: parents (slope = –0.42, $p = .004$, ES = .29), 95% CI [–.68, –.16]; youths (slope = –0.12, $p = .124$, ES = .11), 95% CI [–.30, –.060]. Both parents (slope = –0.78, $p < .001$, ES = 1.91), 95% CI [–1.04, –.32], and youths (slope = –0.72, $p = .001$, ES = 2.66), 95% CI [–1.07, –.37], reported significant improvements on the Top Problems measure, and clinicians reported significant improvements on the CGI (slope = –0.68, $p < .001$, ES = 3.43), 95% CI [–.85, –.51].

Controlling for medication use. Weekly measures analyses were repeated controlling for medication use. Nearly all slopes remained significant, except for youth-reported BPC Total (slope = –0.35, p = .062, ES = .38), 95% CI [–.71, .023], and Externalizing (slope = –0.20, p = .11, ES = .56), 95% CI [–.45, .051], scores. Although the p values in these analyses increased slightly to over .05, the slopes and effect sizes remained essentially the same, suggesting that this p value change was largely a function of adding a predictor to an analysis with a relatively small sample size, rather than a meaningful change in the magnitude of the slopes.

Secondary Clinical Measures: Pre- Versus Posttreatment Assessments

Descriptive statistics and outcome analyses for the secondary clinical measures are presented in Table 1. On the K-SADS, we found significant reductions from pre to post in the total number of diagnoses. Slope estimates indicated that, on average, the number of diagnoses was reduced by 1.29 (58.64%) over the course of treatment (p = .009, ES = .96), 95% CI [–2.01, –.53], from a pretreatment mean of 2.20 to a posttreatment mean of .91[1]; for 81.30% of the sample, the index disorder identified as the primary intervention target at pretreatment had been eliminated by posttreatment. The 58.64% reduction in diagnoses achieved by FIRST can be compared to the 59.92% reduction achieved by MATCH, 52.55% by standard EBPs, and 24.70% by usual care, in Weisz et al. (2012).

Pre- versus post comparisons on the CBCL Total Problems, Internalizing, and Externalizing T scores showed significant decreases on all three scales. Slope estimates indicated that, on average, parents reported 8.68-point decreases (p = .012, ES = .85), 95% CI [–14.24, –3.12], on the Total Problems T score; 6.08-point decreases (p = .038, ES = .51), 95% CI [–11.63, –.53], on the Internalizing T score; and 8.73-point decreases (p = .021, ES = .93), 95% CI [–15.29, –2.17], on the Externalizing T score. In terms of clinical significance, at post- compared to pretreatment, fewer youths scored above the clinical cutoff on the Total Problems score (pre = 33.3%; post = 11.8%), Internalizing score (pre = 54.2%, post = 29.4%), and Externalizing score (pre = 33.3%; post = 8.3%).

Pre versus post comparisons on the YSR Total Problems, Internalizing, and Externalizing T scores also showed significant decreases on all three scales. Slope estimates indicated that, on average, youths reported 12.62-point decreases (p < .001, ES = 1.15), 95% CI [–17.60, –7.64], on the Total Problems T score; 11.18-point decreases (p = .005, ES = .94), 95% CI [–17.61, –4.75], on the Internalizing T score; and 8.62-point decreases (p = .008, ES = .94), 95% CI [–13.98, –3.27], on the Externalizing T score. At post- compared to pretreatment, fewer youth scored above the clinical cutoff on the Total Problems score (pre = 17.4%, post = 5.3%), Internalizing score (pre = 13.0%, post = 10.5%), and Externalizing score (pre = 17.4%, post = 5.3%).

Controlling for medication use. In an additional wave of analyses, to determine whether findings differed with medication controlled, all the analyses of pre–post data were rerun examining use of medication (n = 10 yes vs. n = 14 no) as a time-invariant covariate. The findings showed that decreases in CBCL Total Problems (slope = −7.91, p = .055, ES = .77), 95% CI [–16.14, .32]; CBCL Internalizing (slope = –4.83, p = .15, ES = .40), 95% CI [–12.87, 3.21]; CBCL Externalizing (slope = –7.89, p = .082, ES = .83), 95% CI [–17.64, –1.86]; and YSR Internalizing (slope = –8.05, p = .062, ES = .68), 95% CI [–12.87, 3.21], scores were no longer statistically significant. However, the magnitudes of the slopes were again very similar, and effect sizes remained in the medium-to-large range, suggesting that the small increases in p values were likely artifacts of lowered power from adding a predictor to the models.

DISCUSSION

FIRST was developed as a way to make transdiagnostic treatment, informed by evidence, accessible to clinicians who are motivated but have limited time and resources for new learning. We used as a simplifying framework the fact that diverse EBP approaches share certain core principles of therapeutic change. Five such core principles were integrated to form FIRST. We examined the performance of FIRST across multiple dimensions, with community clinic practitioners treating clinic-referred youths.

The findings suggest that this efficient principle-guided approach may have potential as a complement to the existing armamentarium of excellent youth treatment programs. Our data analyses focused in part on measures of the feasibility of FIRST for use in clinical practice and on acceptability to the participants in therapy. FIRST appeared to generate strong therapeutic engagement and high levels of satisfaction. Session attendance was strong, with low levels of missed appointments. There were also relatively high rates of therapy completion, defined in terms of therapist concurrence with the decision to terminate. Findings also showed high levels of therapeutic alliance, as rated by young clients and their caregivers; scores were high on questionnaire measures of youth, caregiver, and therapist satisfaction with treatment. It is important that coding of FIRST session recordings showed levels of both protocol

[1] Data presented here are model estimates of means and changes over time that take therapist-level nesting into account. Data in Table 1 are raw data.

adherence and competence in delivery by therapists to be slightly higher than adherence and competence scores achieved by therapists using MATCH in earlier research. This suggests that the relatively brief training required for FIRST, combined with modest levels of subsequent group consultation, was sufficient to produce good fidelity to the FIRST treatment model.

Other analyses focused on whether FIRST showed evidence of potential clinical benefit. In support of this notion, we found significant improvement on multiple clinical measures. These findings were derived from pre/post comparisons on standardized diagnoses and on standardized parent and youth symptom checklists, as well as trajectory of change analyses of weekly standardized problem reports by youths and caregivers, weekly severity ratings on the "top problems" identified by youths and caregivers at pretreatment, and weekly clinical global improvement ratings by the therapists.

To put these patterns into context, we provided benchmarking comparisons, where available, with findings from previous research with the modular MATCH treatment program (Chorpita & Weisz, 2009). The comparison was useful because MATCH and FIRST were designed to address youths within a similar age range and with a similar range of mental health problems, and because studies with MATCH and FIRST have both been done in similar clinical service contexts, with practitioners treating clinically referred youths. However, the two treatment programs are designed differently, with MATCH focused on 33 specific treatment procedures embodied within modules and FIRST focused on five broad therapeutic principles. The MATCH manual is 267 pages, and training requires 6 days. The FIRST manual is 98 pages with a 14-page appendix, and FIRST training requires 2 days. Given these substantial differences, we would certainly not expect FIRST to outperform MATCH, but benchmarking seemed useful as a way to determine how much might be lost—in feasibility, acceptability, and clinical outcomes—when a simpler, principle-guided protocol was used. It is helpful that the MATCH research also provided some comparative evidence on feasibility, acceptability, and clinical outcomes associated with usual clinical care and with standard, focal EBPs (i.e., single-disorder treatment protocols for anxiety, depression, and conduct problems; in Chorpita et al., 2015, and Weisz et al., 2012).

In general, the benchmarking comparisons (shown in Table 3) showed more favorable scores for FIRST than for usual care and standard EBPs, with FIRST and MATCH rather similar on most measures. On our trajectory of change analyses of the weekly clinical measures, MATCH showed somewhat steeper slopes of improvement on the standardized problem reports, whereas FIRST showed steeper slopes than MATCH on the top problems identified as most important by youths and caregivers. On diagnostic change from pre- to posttreatment, and on several other measures, FIRST and MATCH were strikingly similar. Without tests of statistical significance, these benchmarking comparisons are not definitive, but they do suggest that the principle-guided approach represented by FIRST may be sufficiently feasible and acceptable and may have sufficient potential for clinical benefit, to merit further testing.

Certain limitations of the study should be noted. The purpose was to examine feasibility and determine whether there was sufficient evidence of treatment benefit to justify further research. Thus, we used a small sample and an open trial design, features that impose limitations on statistical power and the interpretation of findings. Diversity was also limited, with an 87% White sample, and income distribution was rather wide. Our sample also appeared somewhat less severe on CBCL and YSR baseline scores than that found in some community clinics, and the baseline similarity to benchmarking samples varied across measures; the FIRST sample, compared to benchmarks at baseline, was similar in number of *DSM* disorders, somewhat higher in top problem severity, and somewhat lower in BPC ratings. A useful next step will be a full randomized trial with sample size and diversity sufficient to provide a rigorous test of the treatment and shed light on potential moderators of treatment impact. Another limitation, from the perspective of several transdiagnostic treatments noted in the introduction, may be our limited coverage, in FIRST, of the broad array of specific treatment procedures available in the EBPs. This reflects our effort to complement the rich procedurally based approaches already available with a protocol that emphasizes broad principles of therapeutic change more than detailed descriptions of treatment techniques. Whether this principle-guided model has value as a complement to the excellent transdiagnostic approaches currently available is ultimately an empirical question.

If the principle-guided FIRST approach does prove to be useful and effective, the clinical implications could be significant, particularly for efforts to implement EBPs in clinical practice contexts. A treatment protocol guided by a small number of empirically supported therapeutic principles, and designed for efficient training and ease of implementation, could help extend empirically supported treatment to youths across a range of clinical care contexts in which more extensive and expensive programs might be out of reach. Further testing, within a more fully elaborated randomized controlled trial, will help evaluate this possibility.

FUNDING

This work was funded by the National Institute of Mental Health [Grant number MH085963] and the Norlien Foundation.

REFERENCES

Achenbach, T. M., McConaughy, S. H., Ivanova, M. Y., & Rescorla, L. A. (2011). *Manual for the ASEBA Brief Problem Monitor (BPM)*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington: University of Vermont, Research Center for Youths, Youth and Families.

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology*, *68*, 331–339. doi:10.1037/0022-006X.68.2.331

Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation*, *14*, 1–25. doi:10.1207/S1532768XJEPC1401_01

Ambrosini, P. J. (2000). Historical Development and Present Status of the Schedule for Affective Disorders and Schizophrenia for School-Age Youths (K-SADS). *Journal of the American Academy of Youth Psychiatry*, *39*, 49–58. doi:10.1097/00004583-200001000-00016

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, *40*, 57–87. doi:10.1111/1469-7610.00424

Attkisson, C. C., & Zwick, R. (1982). The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning*, *5*(3), 233–237. doi:10.1016/0149-7189(82)90074-X

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. (2011). *The unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. New York, NY: Oxford University Press.

Bearman, S. K., Herren, J., & Weisz, J. R. (2012). *Therapy Integrity in Evidence Based Interventions: Observational Coding System, Coding manual*. Austin: University of Texas at Austin.

Bearman, S. K., & Weisz, J. R. (2015). Review: Comprehensive treatments for youth comorbidity - evidence-guided approaches to a complicated problem. *Child and Adolescent Mental Health*, *20*, 131–141. doi:10.1111/camh.12092

Chorpita, B. F., Park, A., Tsai, K., Korathu-Larson, P., Higa-McMillan, C., Nakamura, B., … Krull, J. (2015). Balancing effectiveness with responsiveness: Therapist satisfaction across different treatment designs in the Child STEPs randomized effectiveness trial. *Journal of Consulting and Clinical Psychology*, *83*, 709–718. doi:10.1037/a0039301

Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., Krull, J. L., & the Research Network on Youth Mental Health. (2010). Evaluation of the Brief Problem Checklist: Child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology*, *78*, 526–536. doi:10.1037/a0019602

Chorpita, B. F., & Weisz, J. R. (2009). *Modular approach to therapy for children with anxiety, depression, or conduct problems (MATCH-ADC)*. Satellite Beach, FL: PracticeWise, LLC.

Chu, B. C., Crocco, S. T., Arnold, C. C., Brown, R., Southam-Gerow, M. A., & Weisz, J. R. (2015). Sustained implementation of cognitive-behavioral therapy for youth anxiety and depression: Long-term effects of structured training and consultation on therapist practice in the field. *Professional Psychology: Research and Practice*, *46*, 70–79. doi:10.1037/a0038000

Chu, B. C., Merson, R. A., Zandberg, L. J., & Areizaga, M. (2012). Calibrating for comorbidity: Clinical decision-making in youth depression and anxiety. *Cognitive and Behavioral Practice*, *19*, 5–16. doi:10.1016/j.cbpra.2010.10.005

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. doi:10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Comer, J. S., & Barlow, D. H. (2014). The occasional case against broad dissemination and implementation: Retaining a role for specialty care in the delivery of psychological treatments. *American Psychologist*, *69*, 1–18. doi:10.1037/a0033582

Ehrenreich-May, J., & Bilek, E. L. (2012). The development of a transdiagnostic cognitive behavioral group intervention for childhood anxiety disorders and co-occurring depression symptoms. *Cognitive and Behavioral Practice*, *19*, 41–55. doi:10.1016/j.cbpra.2011.02.003

Gardner, F., Montgomery, P., & Knerr, W. (2015). Transporting evidence-based parenting programs for child problem behavior (age 3–10) between countries: Systematic review and meta-analysis. *Journal of Clinical Child & Adolescent Psychology*, Advance online publication. 1–14. doi:10.1080/15374416.2015.1015134

Guy, W. (1976). *ECDEU assessment manual for psychopharmacology* (2nd ed.). Washington, DC: U.S. Government Printing Office.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. New York, NY: Wiley.

Hoagwood, K. E., Jensen, P. S., Arnold, L. E., Roper, M., Severe, J., Odbert, C., … MTA Cooperative Group. (2004). Reliability of the Services for Youths and Adolescents-Parent Interview. *Journal of the American Academy of Youth and Adolescent Psychiatry*, *43*, 1345–1354. doi:10.1097/01.chi.0000139558.54948.1f

Hogue, A., & Dauber, S. (2013). Diagnostic profiles among urban adolescents with unmet treatment needs: Comorbidity and perceived need for treatment. *Journal of Emotional and Behavioral Disorders*, *21*, 18–32. doi:10.1177/1063426611407500

IBM Corp. (2013). *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: Author.

Jensen, P. S., Hoagwood, K. E., Roper, M., Arnold, L. E., Odbert, C., Crowe, M., & Wells, K. (2004). The Services for Youths and Adolescents-Parent Interview: Development and performance characteristics. *Journal of the American Academy of Youth and Adolescent Psychiatry*, *43*, 1334–1344. doi:10.1097/01.chi.0000139557.16830.4e

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age youths–present and lifetime version (KSADS-PL): Initial Reliability and Validity Data. *Journal of the American Academy of Youth Psychiatry*, *36*, 980–988. doi:10.1097/00004583-199707000-00021

Khanna, M. S., & Kendall, P. C. (2010). Computer-assisted cognitive behavioral therapy for child anxiety: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *78*, 737–745. doi:10.1037/a0019739

Kratochvil, C., Emslie, G., Silva, S., McNulty, S., Walkup, J., Curry, J., . . . TADS Team. (2006). Acute time to response in the treatment for adolescents with depression study (TADS). *Journal of the American Academy of Child and Adoldescent Psychiatry*, *45*, 1412–1418.

Lai, M. H. C., & Kwok, O. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *The Journal of Experimental Education*, *83*, 423–438. doi:10.1080/00220973.2014.907229.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. doi:10.3389/fpsyg.2013.00863

Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, *2*, 197–207. doi:10.1016/0149-7189(79)90094-6

Lewinsohn, P., Roberts, R., Seeley, J., Rohde, P., Gotlib, I., & Hops, H. (1994). Adolescent psychopathology: II. Psychosocial risk factors for depression. *Journal of Abnormal Psychology*, *103*, 302–315. doi:10.1037/0021-843X.103.2.302

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Study—Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*, 980–989. doi:10.1016/j.jaac.2010.05.017

MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, *56*, 1073–1086. doi:10.1001/archpsyc.56.12.1073

MTA Cooperative Group. (2004). National Institute of Mental Health Multimodal Treatment Study of ADHD follow-up: 24-month outcomes of treatment strategies for attention-deficit/hyperactivity disorder. *Pediatrics*, *113*, 754–761. doi:10.1001/archpsyc.56.12.1073

Piquero, A. R., Farrington, D. P., Welsh, B. C., Tremblay, R., & Jenkins, W. (2008). *Effects of early family/parent training on antisocial behavior and delinquency.* Stockholm, Sweden: National Council on Crime Prevention. Retrieved from http://www.bra.se

Powell, B. J., Hausmann-Stabile, C., & McMillen, J. C. (2013). Mental health clinicians' experiences of implementing evidence-based treatments. *Journal of Evidence-Based Social Work*, *10*, 396–409. doi:10.1080/15433714.2012.664062

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & Du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Rye, C. B., & Kimberly, J. R. (2007). The adoption of innovations by provider organizations in health care. *Medical Care Research and Review*, *64*, 235–278. doi:10.1177/1077558707299865

Shirk, S. R., DePrince, A. P., Crisostomo, P. S., & Labus, J. (2014). Cognitive behavioral therapy for depressed adolescents exposed to interpersonal trauma: An initial effectiveness trial. *Psychotherapy*, *51*, 167–179. doi:10.1037/a0034845

Shirk, S. R., & Saiz, C. (1992). The therapeutic alliance in youth therapy: Clinical, empirical, and developmental perspectives. *Development and Psychopathology*, *4*, 713–728. doi:10.1017/S0954579400004946

Silverman, W. K., & Hinshaw, S. P. (2008). The second special issue on evidence-based psychosocial treatments for youths and adolescents: A ten-year update. *Journal of Clinical Youth and Adolescent Psychology*, *37*, 1–7. doi:10.1080/15374410701817725

Simpson, D. D., & Flynn, P. M. (2007). Moving innovations into treatment: A stage-based approach to program change. *Journal of Substance Abuse Treatment*, *33*, 111–120. doi:10.1016/j.jsat.2006.12.023

Snijders, T. A. B., & Bosker, R. J. (1999). *Mulitlevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Southam-Gerow, M. A., & Prinstein, M. J. (2014). Evidence base updates: The evolution of the evaluation of psychological treatments for children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *43*, 1–6. doi:10.1080/15374416.2013.855128

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., & Wu, M. (2001). Clinical relevance of the primary findings of the MTA success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Youth and Adolescent Psychiatry*, *40*, 168–179. doi:10.1097/00004583-200102000-00011

TADS Team. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for adolescents with depression study (TADS) randomized controlled trial. *Journal of the American Medical Association*, *292*, 807–820. doi:10.1001/jama.292.7.807

U.S. Surgeon General. (2004). *Report of the Surgeon General's conference on children's mental health: A national action agenda*. Rockville, MD: U.S. Department of Health and Human Services.

Verdeli, H., Mufson, L., Lee, L., & Keith, J. A. (2006). Review of evidence-based psychotherapies for pediatric mood and anxiety disorders. *Current Psychiatry Reviews*, *2*, 1–27. doi:10.2174/157340006778018102

Walkup, J. T., Albano, A. M., Piacentini, J., Birmaher, B., Compton, S., Sherrill, J. T., & Kendall, P. C. (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *The New England Journal of Medicine*, *359,* 2753–2766. doi:10.1056/NEJMoa0804633

Weersing, V. R., Rozenman, M. S., Maher-Bridge, M., & Campo, J. V. (2012). Anxiety, depression, and somatic distress: Developing a transdiagnostic internalizing toolbox for pediatric practice. *Cognitive and Behavioral Practice*, *19*, 68–82. doi:10.1016/j.cbpra.2011.06.002

Weisz, J. R., Bearman, S. K., Ugueto, A. M., Herren, J., Allene, A., Cheron, D. M., …. Tweed, L. (2016). *Pragmatic trial of transdiagnostic intervention for youth mental health problems*. Cambridge, MA: Harvard University.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., & Hoagwood, K. (Research Network on Youth Mental Health). (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and track change during psychotherapy. *Journal of Consulting and Clinical Psychology*, *79*, 369–380. doi:10.1037/a0023307

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., & Gibbons, R. D. (Research Network on Youth Mental Health). (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), 274–282.

Weisz, J. R., Hawley, K. M., & Jensen Doss, A. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Youth and Adolescent Psychiatric Clinics of North America*, *13*, 729–816. doi:10.1016/j.chc.2004.05.006

Weisz, J. R., Krumholz, L. S., Santucci, L., Thomassin, K., & Ng, M. (2015). Shrinking the gap between research and practice: Tailoring and testing youth psychotherapies in clinical care contexts. *Annual Review of Clinical Psychology*, *11*, 139–163. doi:10.1146/annurev-clinpsy-032814-112820

Weisz, J. R., Weiss, B., Han, S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with youths and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117*, 450–468. doi:10.1037//0033-2909.117.3.450

Zhou, X., Hetrick, S. E., Barth, J., Qin, B., Whittington, C. J., Cohen, D., . . . Xie, P. (2015). Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: Systematic review and network meta-analysis. *World Psychiatry*, *14*, 207–222. doi:10.1002/wps.20217