A Second and Third Look at FIRST: Testing Adaptations of A Principle-Guided Youth Psychotherapy

Evelyn Cho , Sarah Kate Bearman , Rebecca Woo , John R. Weisz & Kristin M. Hawley

Published online: 07 Aug 2020.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

A Second and Third Look at FIRST: Testing Adaptations of A Principle-Guided Youth Psychotherapy

Evelyn Cho[a], Sarah Kate Bearman [b], Rebecca Woo[b], John R. Weisz [c], and Kristin M. Hawley [a]

[a]Department of Psychological Sciences, University of Missouri; [b]Department of Educational Psychology, University of Texas at Austin; [c]Department of Psychology, Harvard University

ABSTRACT
Objective: We examined the acceptability, integrity, and symptom trajectories associated with FIRST, a principle-guided treatment for youth internalizing and externalizing problems designed to support efficient uptake and implementation.
Method: We conducted two open trials of an adapted FIRST, focusing on uptake and implementation by novice trainees in a university-affiliated clinic, limiting treatment duration to six sessions, and benchmarking findings against a 2017 FIRST trial with community therapists. In Study 1, trainees received a two-day training and weekly two-hour supervision (N = 22 youths, ages 7–17, 50% female, 54.54% Caucasian, 4.55% Latinx). In Study 2, trainees received a one-day training and weekly one-hour supervision, delivering the six-session FIRST in a predetermined sequence (N = 26 youths, ages 11–17, 42.31% female, 65.38% Caucasian, 7.69% Latinx). In Study 3, the original study therapists – now practitioners – evaluated FIRST's effectiveness and implementation difficulty, and reported their own post-study FIRST use.
Results: Acceptability (treatment completion, session attendance, caregiver participation) and integrity (adherence, competence) were comparable across Study 1, Study 2 and the 2017 trial. Improvement effect sizes across ten outcome measures were in the large range in all three trials: M ES = 1.10 in the 2017 trial, 0.83 in Study 1, and 0.81 in Study 2. Study 3 showed high effectiveness ratings, low difficulty ratings, and continued use of FIRST by a majority of clinicians.
Conclusions: Across two open trials and a follow-up survey, FIRST showed evidence of acceptability and integrity, with youth symptom reduction comparable to that in prior research.

The number of psychosocial treatments for youth mental health disorders developed and tested in randomized controlled trials (RCTs) has multiplied in recent decades (Southam-Gerow et al., 2014), far outpacing their implementation in community settings (e.g., Jensen-Doss et al., 2009). This "research-to-practice-gap" persists despite considerable evidence for the efficacy of evidence-based treatments (EBTs) from RCTs (Weisz, Kuppens et al., 2017), the smaller but still significantly superior effects of EBTs relative to treatment-as-usual (Weisz et al., 2013), and naturalistic examples of improved outcomes associated with therapists' use of EBT strategies (e.g., Garland et al., 2014). This relatively infrequent EBT use in community settings has prompted thoughtful consideration of innovative approaches to treatment design, to expand opportunities for EBT implementation in real world practice contexts (Herschell et al., 2004).

Some alternatives have been developed to address the fact that EBTs targeting a single disorder, or a small family of related problems, may not be an ideal fit to all youths who need care (Powell et al., 2013). Some clinically referred youths may be more complex and comorbid than those often represented in RCTs (Ehrenreich et al., 2011), and therapist caseloads routinely span a broad range of problem areas (Cho et al., 2019). Additionally, some standard, single-disorder EBTs prescribe 16–24 sessions, with an adequate treatment dose requiring completion of all or most of the protocol. Many youths in community settings fail to complete this dose (De Haan et al., 2013), and may miss components needed for clinical benefit, especially if therapists have to balance targeting multiple co-occurring problems. Finally, time-intensive training, supervision, and treatment integrity monitoring often employed to maximize internal validity in RCTs (Beidas & Kendall, 2010) may be difficult for community therapists and organizations to emulate. Perhaps related to these research-versus-practice differences, studies also show that treatment integrity is sometimes compromised as EBTs move from research to community implementation (e.g., Curtis et al., 2004).

These challenges may be especially burdensome in non-specialty settings (e.g., schools, primary care) where many youths receive care (Burns et al., 1995), treatment length is typically short, few training resources are available, and therapists may have limited experience and training in youth psychotherapy (Evans & Weist, 2004; Gomez et al., 2014). Thus, brief treatments that can be easily learned, adopted, and delivered by therapists with diverse backgrounds may be critical to increasing youths' access to EBTs in some service settings.

Consideration of these issues has prompted some excellent complementary approaches to treatment design, addressing the clinical complexity and comorbidity of some referred youths, providing training and support to help therapists who have diverse caseloads, and introducing strategies for personalizing treatment (e.g., Chorpita & Weisz, 2009; Ehrenreich et al., 2011; Ng & Weisz, 2016). One complementary approach involves structuring treatment around *empirically-supported principles of change*, a strategy suggested four decades ago by Goldfried (1980) and revived and elaborated upon more recently (e.g., Castonguay & Beutler, 2006; Davison, 2019; Hoffman & Hayes, 2019; Oddli et al., 2016; Rosen & Davison, 2003), including by members of the Society of Clinical Psychology EST task force (Tolin et al., 2015). As several of these authors suggest, learning specific treatment procedures may be usefully complemented by a focus on the kinds of change processes that may be set in motion by those procedures; this may be especially true for therapists to whom the treatment procedures are relatively new (Abramowitz, 2013; Marchette & Weisz, 2017) or different from their previous practices (Oddli et al., 2016). Given the current "knowledge management problem" (Chorpita et al., 2011, p. 493), whereby the breadth of psychological knowledge is too great for individual practitioners to practically digest for clinical use, focusing on principles of change applicable across problems and diagnostic categories may add usefully to existing strategies within the expanding array of specific treatment procedures for a range of youth mental health conditions.

The FIRST treatment program is aligned with these ideas. FIRST is a flexible, transdiagnostic intervention organized around well-tested principles of youth psychotherapeutic change and drafted using a process of "co-design" with clinical practitioners and treatment researchers (Weisz, Bearman et al., 2017). FIRST addresses both internalizing and externalizing dimensions, and integrates multiple treatment procedures derived from youth EBTs for anxiety, depression, and misconduct (Weisz & Bearman, 2020). Each of the five

principles is included in multiple tested and effective interventions for anxiety, depression, and conduct problems (Chorpita et al., 2005; Garland et al., 2008), and each has shown significant treatment effects in multiple studies when used as a solo intervention (Weisz et al., 2004). FIRST was designed to be focused in content (five core principles of therapeutic change), in part to provide efficiency for community therapists who seek to offer empirically-supported care but whose training resources and implementation supports may be limited. In an initial open benchmarking trial, youths with anxiety, depression, or conduct problems, referred to two community mental health clinics, were treated by clinic therapists (average 7.96 years of clinical experience) who were trained and supervised by the FIRST developers (Weisz, Bearman et al., 2017). Youths showed significant pre-to-post treatment improvement on multiple symptom measures and youth- and caregiver-nominated top problems. Therapists, youths, and caregivers also reported high satisfaction. In a recent follow-up with FIRST therapists, 92% reported that understanding the principles underlying the practices increased their clinical confidence and competence and "motivated them to continue using the intervention on their own" (Bearman et al., 2019, pp. 23–24). Thus, in this initial study, FIRST showed promise with experienced community therapists trained by the developers.

Critical questions regarding FIRST's potential for implementation remain unanswered, however. One of these is its ability to survive the "implementation cliff" (Weisz et al., 2014, p. 59) – that is, the reduction in fidelity and outcomes sometimes observed when interventions move away from developer control and oversight (e.g., Curtis et al., 2004). In the 2017 FIRST trial, although therapists were employed by partner community mental health clinics, they had direct access to the treatment development team (i.e., initial training, ongoing consultation) in ways unlikely to be replicated should FIRST be implemented in most clinical practice contexts (e.g., non-specialty settings) or on a large scale. FIRST also averaged 18.63 sessions, similar to other youth EBTs but substantially more than youths receive in community outpatient (De Haan et al., 2013; Harpaz-Rotem et al., 2004) and non-specialty settings (e.g., Gomez et al., 2014).

One framework for systematically improving fit and performance is the deployment-focused model of treatment development and testing (Weisz & Gray, 2008), which recommends iterative cycles of testing and adaptation within clinically representative contexts to ensure that interventions remain effective under likely implementation conditions. The time and financial pressures associated with scaling up interventions in practice

settings can naturally lead to efforts to manage costs – often by trimming training and support infrastructure (Weisz et al., 2014). With that in mind, we sought to examine as a proof of concept the durability of FIRST when treatment length, training, and expert oversight were increasingly abbreviated to better reflect the likely conditions of community-based implementation. Further, we used novice trainees, thus examining the functioning of this abbreviated FIRST with therapists who – unlike those in the 2017 FIRST trial but often found in non-specialty settings – did not have a strong base of prior clinical experience (see Figure 1). In Study 1, session content and sequence were guided by decision trees but treatment length was capped at six sessions to allow potential coverage of all five FIRST principles in the minimal number of sessions, while also reflecting the brief treatments, and the few sessions typically attended, in the community (De Haan et al., 2013; Gomez et al., 2014; Harpaz-Rotem et al., 2004). Training was provided to novice therapists (i.e., graduate trainees) in a community-based university clinic by a FIRST developer using the 2017 FIRST trial's two-day training. In Study 2, sessions were again limited to six, but session content and sequence were fixed to simplify implementation, training, and support that had been dedicated to individualizing treatment in Study 1. Initial training, reduced to one day, was provided by the on-site supervisor without developer involvement (see Figure 1). In Study 3, we conducted a follow-up with Study 1 and 2 trainees, after they had become practicing clinicians, to assess their perceptions of the adapted FIRST's effectiveness and difficulty level, and their post-study use of FIRST.

Using this three-study approach, we evaluated whether FIRST can be implemented by novice trainees with limited ongoing support, and whether this adapted FIRST was associated with symptom reduction. We then compared the acceptability, integrity, and clinical outcomes across the two studies and benchmarked them against the 2017 FIRST trial, allowing us to consider how these outcomes would fare with number of sessions capped and without treatment developer involvement, as might be the case if FIRST were implemented in community settings.

Study 1

Method

Study Design and Procedures

We used an open trial (pre-post) design. Youths were recruited via advertisements offering free therapy to youth aged 7–17 (based on age ranges in review of youth-focused EBTs for these target problems; e.g., Weisz et al., 2019) with anxiety, depression, and behavior problems in a Midwestern university-affiliated mental health clinic serving youths, families, and adults from the local community. Caregivers who responded to the advertisement were screened by phone by study staff (see Figure 2). Inclusion criteria were caregiver concerns about youth's anxiety, depression, and/or behavior problems, and no past diagnosis of intellectual

	2017 FIRST Trial (Weisz, Bearman et al., 2017)	Study 1	Study 2
Youth Participants	Clinic-referred ($n = 24$)	Recruited/analogue ($n = 22$)	Recruited/analogue ($n = 26$)
Therapist Participants	Therapists ($n = 14$) from two community mental health clinics	Novice graduate trainees ($n = 16$) in a university-affiliating clinic	Novice graduate trainees ($n = 14$) in a university-affiliating clinic
Treatment Length	No limit ($M = 18.63$)	Six sessions ($M = 5.68$)	Six sessions ($M = 5.00$)
Initial Training	Two days with developer	Two days with developer	One day with non-developer expert trained in Study 1
Ongoing Support	Weekly one-hour FIRST group consultation with developer[a]	Weekly one-hour group and one-hour individual FIRST supervision with non-developer CBT expert[b]	Weekly one-hour group FIRST supervision with non-developer CBT expert[b]
Individualization	Yes (Decision trees)	Yes (Decision trees)	No (Preset sequence)

Figure 1. Overview of implementation conditions in the three FIRST trials.
[a]Only one FIRST therapist was pre-licensure; this therapist received routine clinic supervision in addition to the FIRST consultation. [b]Study 1 and 2 therapists were all pre-licensure trainees and received routine clinic supervision

Study 1   Study 2   Study 3

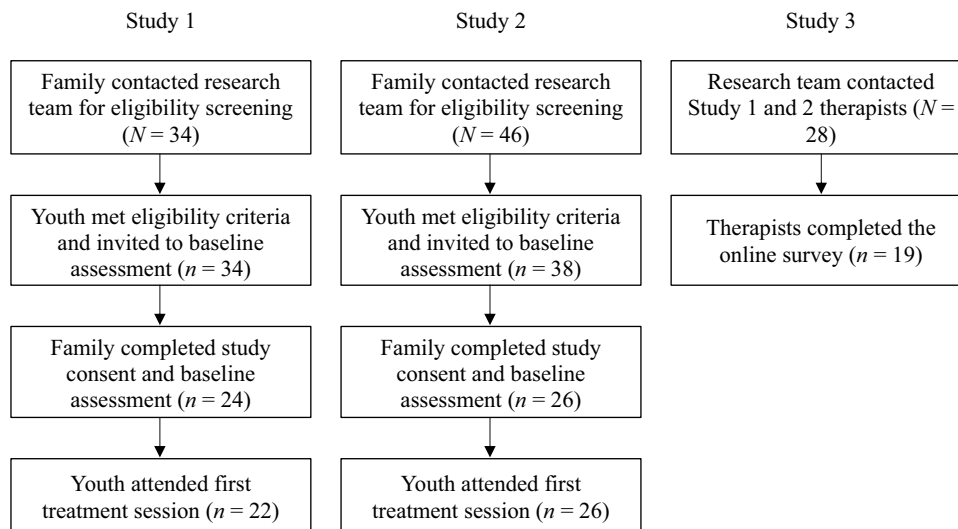| Study 1	Study 2	Study 3
Family contacted research team for eligibility screening ($N = 34$)	Family contacted research team for eligibility screening ($N = 46$)	Research team contacted Study 1 and 2 therapists ($N = 28$)
Youth met eligibility criteria and invited to baseline assessment ($n = 34$)	Youth met eligibility criteria and invited to baseline assessment ($n = 38$)	Therapists completed the online survey ($n = 19$)
Family completed study consent and baseline assessment ($n = 24$)	Family completed study consent and baseline assessment ($n = 26$)	
Youth attended first treatment session ($n = 22$)	Youth attended first treatment session ($n = 26$)	

Figure 2. Participant enrollment flowchart.

disability or mental retardation. Following the phone screen, youths and caregivers came to the clinic where they consented to study participation and completed the initial assessment. This process did not overlap with the usual intake procedure at this clinic. Youths and caregivers completed weekly symptom and top problem ratings, and a post-treatment assessment. All procedures were approved by the institution's Institutional Review Board.

Participants

Youth participants ($N = 22$) had a mean age of 10.61 ($SD = 3.17$) years. Half were female; 54.55% were Caucasian, 22.73% multiracial, 4.55% Asian/Pacific Islander; and 4.55% Latinx. Caregivers were all biological parents; 81.82% had a college degree. Presenting concerns (multiple could be endorsed) were 68.18% anxiety, 50.00% depression, and 45.45% behavior. Therapists were 16 graduate students beginning their clinical training at the university-affiliated clinic. Therapist mean age was 30.44 ($SD = 5.44$) years; 62.50% were female; 87.5% were Caucasian, 6.25% African American/Black, 6.25% Asian/Pacific Islander; and 6.25% Latinx. Training background was 56.25% psychology, 37.50% social work, 6.25% counseling; and split evenly between bachelor's and master's level. None were licensed (see Table 1).

Treatment and Therapist Training

FIRST is a principle-guided transdiagnostic approach designed to treat youth anxiety, depression, and behavior problems (Weisz, Bearman et al., 2017). FIRST includes five core principles: (1) **F**eeling calm (e.g., relaxation), (2) **I**ncreasing motivation (e.g., rewards,

consequences), (3) **R**epairing thoughts (e.g., cognitive restructuring), (4) **S**olving problems (i.e., problem solving steps), and (5) **T**rying the opposite (i.e., countering maladaptive impulses – e.g., exposure, behavioral activation). Each principle comes with session guides, suggested in-session activities, and out-of-session assignments, with content specified for the youth, caregiver, or the youth-caregiver dyad. Treatment content and sequence (e.g., shifting from one principle or treatment target to another) was guided by decision trees and youth and caregiver weekly symptom and top problem ratings (these were available to therapists and the supervisor during weekly meetings), just as in the 2017 FIRST trial. However, treatment length and ongoing support were truncated from the 2017 FIRST trial (see Figure 1) to ease implementation burden. Treatment dose was capped at six one-hour sessions, although sessions were longer ($M = 61.05$, $SD = 14.15$ minutes) if families met fewer than six times. Six sessions were selected as the minimum number of sessions to permit possible coverage of all five FIRST principles and a final review. Therapists and the on-site supervisor, a non-developer CBT expert, received an initial two-day training by one of the FIRST developers, who provided training in the 2017 FIRST trial, followed by weekly one-hour individual and one-hour group supervision to support implementation of FIRST.

Measures

Therapist Integrity in Evidence Based Interventions (TIEBI) Coding System (Bearman et al., 2012). All treatment sessions were taped. Sessions were divided into thirds (early, middle, late phase of treatment), with

Table 1. Baseline characteristics of youths, families, and therapists in the two current studies and the 2017 FIRST trial.

Overall	2017 FIRST Trial	Study 1	Study 2
Youth *M* Age (*SD*)	11.03 (2.69)	10.61 (3.17)	12.08 (1.83)
Youth % Female	50%	50.00%	42.31%
Youth % Caucasian/White	87.50%	54.54%	65.38%
African American/Black	0.00%	0.00%	7.69%
Asian American/Pacific Islander	0.00%	4.55%	7.69%
Native American/American Indian	0.00%	0.00%	3.85%
Multiracial	8.33%	22.73%%	11.54%
Youth % Latinx	8.33%	4.55%	7.69%
M Caregiver Age (*SD*)	-	41.73 (6.21)	43.46 (5.85)
Caregiver % Female	-	90.91%	92.31%
Caregiver % High School Diploma	-	18.18%	26.92%
College Degree	-	50.00%	46.15%
Post-College Degree	-	31.82%	23.08%
% Biological Child	95.83%	100.00%	92.31%
% Primary Presenting Concerns			
Anxiety	-	68.18%	57.69%
Depression	-	50.00%	46.15%
Behavior	-	45.45%	19.23%
Pretreatment Means (SD)			
BPC-Parent	9.04 (4.61)	6.33 (3.22)	5.96 (2.86)
BPC-Youth	6.12 (4.57)	6.91 (4.40)	5.76 (3.67)
TPA-Parent[a]	7.85 (1.78)	1.27 (0.51)	1.42 (0.47)
TPA-Youth[a]	7.54 (1.29)	1.32 (0.62)	1.40 (0.35)
CBCL Total Problems	60.46 (9.47)	57.50 (6.68)	60.04 (7.57)
YSR Total Problems	55.96 (10.58)	56.77 (13.14)	56.46 (8.64)
CBCL Internalizing	62.17 (11.98)	59.23 (7.53)	63.12 (8.68)
YSR Internalizing	57.26 (11.96)	56.18 (11.92)	57.15 (8.79)
CBCL Externalizing	56.58 (11.40)	55.55 (8.24)	55.80 (8.64)
YSR Externalizing	50.96 (10.58)	52.64 (13.23)	51.27 (9.37)
Therapist *M* Age (*SD*)	35.57 (8.69)	30.44 (5.44)	28.07 (6.90)
Therapist % Female	79%	62.50%	85.71%
Therapist % Caucasian/White	93%	87.50%	78.57%
Asian/Pacific Islander	-	6.25%	14.29%
African American/Black	-	6.25%	7.14%
Therapist % Latinx	7%	6.25%	7.14%
Therapist % Licensed	93%	0.00%	0.00%
Therapist % Psychology	43%	56.25%	64.29%
Counseling	7%	6.25%	35.71%
Social Work	50%	37.50%	0.00%

P = Parent/Caregiver, Y = Youth, BPC = Brief Problems Checklist, TPA = Top Problems Assessment, CBCL = Child Behavior Checklist, YSR = Youth Self Report; [a]TPA used a 0–2 scale.

one session within each phase randomly selected for coding (50% of possible recordings were coded). If three or fewer sessions were available, all were coded. Coding included therapist adherence (percent of five minute segments in which 19 items reflecting prescribed content from FIRST was present) and competence (skillfulness and thoroughness of delivery, rated from 0 = not at all to 4 = expert). To generate a mean-percent adherence score for each coded treatment session for each FIRST content item, the number of five minute segments in which that item was coded

"present" was summed, multiplied by five (number of minutes in each segment), and divided by the total session time. To generate a mean competence score for each youth's treatment session, each of the global competency codes for each "present" item were averaged. We randomly selected 32 sessions for double coding to assess agreement between coders. Reliability, across pairs of coders, was *M* ICC (1,2) =.91 (range = .68-.99) for adherence, and *M* ICC (1,2) = .87 (range = .75-.97) for competence.

Therapy Process Observational Coding System-Alliance Scale (TPOCS-A; McLeod & Weisz, 2005). The nine-item TPOCS-A assessed for youth-therapist and caregiver-therapist alliance. Independent coders rate on a six-point scale (0 = not at all to 5 = great deal) the extent to which youths and caregivers demonstrate understanding, hostility, positive affect and other alliance-related behaviors in-session. The TPOCS-A has acceptable internal consistency (α = 0.95; McLeod & Weisz, 2005). ICCs in this sample ranged from acceptable to excellent for most items: youth-therapist *M* ICC (1,2) = .64, range = .39–.88; caregiver-therapist *M* ICC (1,2) = .68, range = .43–.93. One item for each scale did not achieve acceptable ICC (i.e., ICC = .40; Cicchetti & Sparrow, 1981) and was excluded: whether the youth-therapist dyad appeared uncomfortable, ICC (1,2) = .39; whether the caregiver complied with therapy (there was no variability on this item between sessions or among coders). These ICCs were comparable to those in published studies (McLeod & Weisz, 2005). TPOCS-A scores for the 2017 FIRST trial were coded as part of another study (Woo, 2018) and are included here for comparison.

Pre- and post-treatment symptom measures. At pre- and post-treatment assessments, caregivers and youths completed the Child Behavior Checklist (CBCL) and Youth Self-Report (YSR; Achenbach & Rescorla, 2001). Caregivers reported on 118 youth problems on the CBCL, which yields internalizing, externalizing, total problem *T* scores based on youth age and gender. Youths reported on 112 youth problems on the YSR, which also yields internalizing, externalizing, and total problem *T* scores based on youth age and gender. Extensive evidence exists supporting the psychometric integrity of both measures (Achenbach & Rescorla, 2001).

Session-by-session symptom and top problem measures. At each treatment session, caregivers and youths completed the Brief Problem Checklist (BPC; Chorpita et al., 2010) and Top Problems Assessment (TPA;

Weisz et al., 2011). The BPC is a 12-item standardized measure (ratings of 0–2) including six internalizing and six externalizing items yielding a total score range from 0–24. The TPA is an idiographic measure of self-reported severity on up to three youth- and caregiver-identified problems (Weisz et al., 2011). In the current studies, the TPA was rated on a 0 to 2 scale, consistent with the BPC, CBCL, and YSR. The BPC (Chorpita et al., 2010) and TPA (Weisz et al., 2011) have demonstrated strong psychometric properties.

Data Analytic Plan

We present descriptive statistics and frequencies for acceptability and integrity outcomes. For clinical benefit, we ran multilevel models using restricted maximum likelihood in R (R Core Team, 2017) using intent-to-treat analysis with all youths initially enrolled in the study. We estimated change over time on all session-by-session and pre-post measures using all available measurement occasions. To determine the appropriate number of levels, we calculated design effects for each outcome and examined therapist-level variances; all design effects were less than 1.1 indicating the third level was unnecessary for our average cluster size ($M = 1.38$; Lai & Kwok, 2015). Therapist-level variance was also near zero for almost all outcomes. Thus, we present the results of two-level models with measurement occasion nested within youths. To facilitate comparisons with the 2017 FIRST trial, we log transformed days in treatment[1]. To benchmark our change trajectories with the 2017 FIRST trial, we obtained data from the 2017 FIRST trial and ran two-level multilevel models predicting outcome trajectories with a study x log days interaction term (negative interaction indicates faster improvement in Study 1), controlling for baseline scores. The different TPA rating scales in Study 1 (0–2) and the 2017 FIRST trial (0–10) did not permit comparisons for TPA trajectories. We calculated pre-post treatment effect sizes according to the procedures outlined in Weisz, Bearman et al. (2017; i.e., group post- and pre-treatment difference divided by the pretreatment standard deviation).

Results

Preliminary Analyses

Completers ($n = 19$) had higher baseline YSR Total scores than noncompleters ($n = 3$), $t(16.15) = 3.39$, $p = .005$. All other demographic and baseline clinical characteristics

did not differ ($p > .05$). Our sample did not differ from the 2017 FIRST trial sample on demographic and baseline clinical characteristics ($p > .004$, Bonferroni-corrected).

Treatment Integrity

TIEBI adherence ($M = 97.72\%$, $SD = 7.51$) was higher than in the 2017 FIRST trial, $t(67.71) = -2.73$, $p = .008$ (see Table 2). TIEBI competence ($M = 2.24$, $SD = 0.50$) did not differ from the 2017 FIRST trial ($M = 2.29$ $SD = 0.79$), $t(95.90) = 0.44$, $p = .660$.

Treatment Acceptability

Mean treatment dose was 5.68 ($SD = 0.89$) sessions, a third of the 2017 FIRST trial's 18.63 sessions. We saw 86.36% of cases complete the full six sessions compared to 68.20% with a planned termination in the 2017 FIRST trial, $\chi^2(1, N = 46) = 1.48$, $p = .223$. We saw 96.15% of scheduled sessions attended, compared to 92.70% in the 2017 FIRST trial, $\chi^2(1, N = 612) = 1.44$, $p = .231$. Of the scheduled sessions, 10.77% were not attended as planned; 64.29% of these sessions were rescheduled, all of which were attended. Caregivers participated in 99.19% of sessions, higher than the 76.20% in the 2017 FIRST trial, χ^2 $(1, N = 571) = 37.87$, $p < .001$. Youth alliance ($M = 3.82$, $SD = 0.69$) did not differ from the 2017 FIRST trial ($M = 3.57$, $SD = 0.85$), $t(108) = -1.68$, $p = .096$. Caregiver alliance ($M = 3.69$, $SD = 0.56$) was lower than in the 2017 FIRST trial ($M = 3.94$, $SD = 0.66$), $t(108) = 2.15$, $p = .034$ (see Table 2).

Symptom Trajectories

Session-by-Session Measures. Caregivers and youths reported improvement on all session-by-session measures (see Table 2). Caregiver (slope = −1.63, 95% CI [−2.48, −0.75], $p < .001$, ES = 1.35) and youth (slope = −1.71, 95% CI [−2.30, −1.11], $p < .001$, ES = 0.69) BPC improvements were in the medium to large range. Caregiver (slope = −0.31, 95% CI [−0.49, −0.12], $p = .001$, ES = 1.41) and youth (slope = −0.31, 95% CI [−0.51, −0.09], $p = .007$, ES = 1.15) TPA improvements were in the large range. These effects were comparable to the 2017 FIRST trial's large effects. The study x log days interaction coefficients indicated that caregiver (−1.27, $p = .015$) and youth (−1.37, $p < .001$) BPC improvement was faster in Study 1.

Pre-Post Measures. CBCL Total (change = −7.96, 95% CI [−5.99, −9.92], $p = .001$, ES = 0.90), Internalizing (change = −7.66, 95% CI [−6.35, −8.96], $p < .001$, ES = 0.81), and Externalizing (change = −9.44, 95% CI [−6.92, −11.96], $p = .001$, ES = 0.85) improvements

---

[1]Analyses were repeated using raw days in treatment and the pattern of results was nearly identical.

Table 2. Benchmarking integrity, acceptability, and clinical outcomes of current studies with the 2017 FIRST trial.

	2017 FIRST Trial			Study 1			Study 2		
Acceptability									
% Treatment completion	68.20%			86.36%			73.07%		
% Sessions attended	92.70%			94.70%			83.33%		
% Caregiver participation	76.20%			99.19%			98. 23%		
M (SD) Youth TPOCS-A	3.57 (0.85)			3.82 (0.69)			3.72 (0.81)		
M (SD) Caregiver TPOCS-A	3.94 (0.66)			3.69 (0.56)			3.59 (0.57)		
M (SD) Effectiveness rating	–			7.11 (1.27)			7.00 (1.00)		
M (SD) Difficulty rating	–			2.00 (2.83)			2.78 (1.79)		
% Post-study FIRST use	–			83.33%			33.33%		
Integrity									
Therapist Competence M (SD)	2.29 (0.79)			2.24 (0.50)			2.40 (0.49)		
% Protocol Adherence (SD)	88.50% (25.24)			97.72% (7.51)			98.61% (8.58)		
Clinical Benefit	*slope*[a]	*p*	ES	*slope*[a]	*p*	ES	*slope*[a]	*p*	ES
BPC Total – P	−0.80	<.001	0.66	−1.63	<.001	1.35	−0.73	<.001	0.98
BPC Total – Y	−0.35	.013	0.41	−1.71	<.001	0.69	−0.12	.583	0.22
TPA – P[b]	−0.78	<.001	1.91	−0.31	.001	1.41	−0.24	<.001	1.74
TPA – Y[b]	−0.72	.001	2.66	−0.31	.007	1.15	−0.09	.024	0.75
CBCL Total	−8.68	.012	0.85	−7.96	.001	0.90	−6.24	<.001	0.74
YSR Total	−12.62	<.001	1.15	−8.83	<.001	0.44	−5.74	<.001	0.67
CBCL Internalizing	−6.08	.038	0.51	−7.66	<.001	0.81	−7.95	<.001	1.27
YSR Internalizing	−11.18	.005	0.94	−6.28	.015	0.35	−6.22	.002	0.79
CBCL Externalizing	−8.73	.021	0.93	−9.44	.001	0.85	−4.52	.003	0.64
YSR Externalizing	−8.62	.008	0.94	−7.94	.003	0.33	−3.64	.004	0.30
	P M ES = 0.97			P M ES = 1.06			P M ES = 1.07		
	Y M ES = 1.22			Y M ES = 0.59			Y M ES = 0.55		
	Total M ES = 1.10			Total M ES = 0.83			Total M ES = 0.81		

2017 FIRST Trial = Weisz, Bearman et al. (2017); TPOCS-A = Therapy Process Observational Coding System-Alliance Scale; P = Parent/Caregiver, Y = Youth, BPC = Brief Problems Checklist, TPA = Top Problems Assessment, CBCL = Child Behavior Checklist, YSR = Youth Self Report.
[a]Slopes from multilevel models (change per log day) for BPC and TPA and estimated pre-post change for CBCL and YSR.
[b]TPA was rated on a 0–2 scale instead of the 1–10 used in Weisz, Bearman et al. (2017).

were in the large range (see Table 2). YSR Total (change = −8.83, 95% CI [−6.71, −10.94], $p < .001$, ES = 0.44), Internalizing (change = −6.28, 95% CI [−4.02, −8.54], $p = .015$, ES = 0.35), and Externalizing (change = −7.94, 95% CI [−5.53, −10.34], $p = .003$, ES = 0.33) improvements were in the small range. CBCL effect sizes were comparable to, and YSR effect sizes were smaller than, the 2017 trial. Study 1 CBCL and YSR trajectories did not differ from the 2017 FIRST trial.

Study 2

Method

Study Design and Procedures

We again used an open trial design. In Study 2, youths under 10 were excluded as part of a larger study evaluating youth-reported treatment adherence measures. Other than youth age, inclusion criteria, enrollment, and measures were identical to Study 1. All procedures were approved by the Institutional Review Board.

Participants

The 26 youth participants had a mean age of 12.08 (SD = 1.83) years; 57.69% male; 65.38% Caucasian, 11.54% multiracial, 7.69% African American/Black, 7.69% Asian/Pacific Islander, and 3.85% Native American/American Indian; and 7.69% Latinx. Most

caregivers were the biological parent (92.31%) and 69.23% had at least a college degree. Presenting concerns included anxiety (57.69%), depression (46.15%), and behavior (19.23%). Therapists were novice trainees starting their training at the same clinic in Study 1. The 14 therapists (2 also participated in Study 1) had a mean age of 28.07 (SD = 6.90) years; 85.71% were female; 78.57% were Caucasian, 14.29% Asian, and 7.14% were African American/Black; 7.14% were Latinx. Training backgrounds were 64.29% psychology and 35.71% counseling. A third were bachelor's and two-thirds master's level. None were licensed (see Table 1).

Treatment and Therapist Training

Following Study 1, additional adaptations were made to better align FIRST with the limited training and supervision resources of typical community clinics. The initial training and weekly individual meetings to support the use of symptom ratings and decision trees to guide treatment individualization for each youth could not be sustained long-term. As such, the on-site supervisor simplified the FIRST protocol by fixing the content and sequence (i.e., no use of decision trees or symptom ratings): (1) the first session began with establishing top problems and treatment goals, followed by brief affective education and **I**ncreasing motivation, (2) **S**olving problems, (3) **F**eeling calm, (4) **R**epairing thoughts, (5) **T**rying the opposite, and (6) a review of

the above principles and development of an individualized coping plan for each youth in the final session. All youths, regardless of their primary problems, received the same five principles in the same six-session sequence. This also allowed training to be condensed to a single day followed by weekly one-hour group FIRST supervision provided by the previously trained, non-developer, on-site supervisor. Third, elimination of the decision trees allowed the on-site supervisor to provide FIRST implementation support in a one-hour group supervision meeting (consistent with the 2017 FIRST trial), without the need for the individual meetings that had been largely dedicated to individualizing treatment for each youth in Study 1. Youths and caregivers' weekly top problem and symptom ratings were still available to therapists and the supervisor, but did not determine session sequence. Treatment sessions were on average 53.30 (SD = 10.19) minutes.

Data Analytic Plan

We calculated design effects and examined therapist-level variance for each outcome. Design effects for all outcomes were less than 1.1 except for YSR Internalizing (design effect = 1.31). Therapist-level variance was near zero for most outcomes indicating that two-level models were appropriate.[2] Thus, analyses evaluating the integrity, acceptability, and clinical benefit of Study 2 and comparisons with the 2017 FIRST trial were identical to those in Study 1. We then repeated these analyses to compare Study 1 and 2 outcomes. In the two-level models comparing Study 1 and 2, a negative interaction term indicates faster improvement in Study 2.

Results

Preliminary Analyses
Completers (n = 19) had lower CBCL Internalizing scores than noncompleters (n = 7), $t(23)$ = 2.16, p = .042. All other demographic and baseline clinical characteristics did not differ (p > .10). Our sample did not differ from the 2017 FIRST trial on demographic and baseline clinical characteristics (p > .004, Bonferroni-corrected).

Treatment Integrity
TIEBI adherence (M = 98.61%, SD = 8.58) was higher than in the 2017 FIRST trial, $t(69.93)$ = −2.97, p = .004 (see Table 2). TIEBI competence (M = 2.40, SD = 0.49)

did not differ from the 2017 FIRST trial (M = 2.29, SD = 0.79), $t(93.76)$ = −0.94, p = .352.

Treatment Acceptability
The mean dose of treatment was 5.00 sessions (SD = 1.81), less than a third of the 2017 FIRST trial's 18.63 sessions. We saw 73.07% cases complete the full six sessions compared to 68.20% with a planned termination in the 2017 trial, $\chi^2(1, N = 50)$ = 0.03, p = .853. We saw 96.30% of scheduled sessions attended, which did not differ significantly from the 92.70% in the 2017 FIRST trial, $\chi^2(1, N = 617)$ = 1.65, p = .198. Of the scheduled sessions, 11.11% were not attended as planned; 66.67% of these were rescheduled, all of which were attended. Caregivers participated in 98.23%, greater than the 76.20% in the 2017 FIRST trial, $\chi^2(1, N = 577)$ = 39.76, p < .001. Youth alliance did not differ between Study 2 (M = 3.72, SD = 0.81) and the 2017 FIRST trial (M = 3.57, SD = 0.85), t (112) = −0.90, p = .760. Caregiver alliance in Study 2 (M = 3.59, SD = 0.57) was lower than in the 2017 trial (M = 3.94, SD = 0.66), $t(113)$ = 2.98, p = .004.

Clinical Benefit
Session-by-Session Measures. Caregiver BPC improvements were in the large range (slope = −0.73, 95% CI [−1.09, −0.37], p < .001, ES = 0.98). The youth BPC slope was not significant (slope = −0.12, 95% CI [−0.54, 0.30], p = .583, ES = 0.22), and pre-post change was in the small range. Caregiver (slope = −0.24, 95% CI [−0.33, −0.14], p < .001, ES = 1.74) and youth (slope = −0.09, 95% CI [−0.17, −0.01], p = .024, ES = 0.75) TPA improvement was in the medium to large range. Study 2 caregiver effect sizes were comparable to, but BPC improvement was faster in the 2017 trial (study x log days coefficient = 0.43, p = .019). Study 2 youth effects were smaller than in the 2017 trial (see Table 2).

Pre-Post Measures. CBCL Total (change = −6.24, 95% CI [−4.22, −8.26], p < .001, ES = 0.74), Internalizing (change = −7.95, 95% CI [−5.16, −10.74], p < .001, ES = 1.27), and Externalizing (change = −4.52, 95% CI [−2.26, −6.77], p = .003, ES = 0.64) improvements were in the medium to large range. YSR Total (change = −5.74, 95% CI [−3.26, −8.22], p < .001, ES = 0.67), Internalizing (change = −6.22, 95% CI [−3.12, −9.32], p = .002, ES = 0.79), and Externalizing (change = −3.64, 95% CI [−1.68, −5.59], p = .004, ES = 0.30) improvements were in the small to medium range. The Study 2 CBCL

---

[2]Analyses were repeated using three-level models and the pattern of results was nearly identical.

Internalizing effect size was comparable to, and the improvement trajectory was faster (slope = −1.06, p = .022) than, the 2017 FIRST trial. Study 2 CBCL Total and Externalizing and all YSR effect sizes were smaller than in the 2017 FIRST trial (large), but these Study 2 trajectories did not differ from those in the 2017 trial (p > .05).

Study 1 and 2 Comparisons

Adherence did not differ between the two studies, t (145) = −0.67, p = .503, but competence was marginally higher in Study 2, $t(145) = −1.98$, p = .050 (see Table 2). The proportion of treatment completers did not differ, χ^2 (1, N = 48) = 0.60, p = .440. Caregiver participation also did not differ, $\chi^2(1, N = 259) < .001$, p = .100. Session attendance did not differ $\chi^2(1, N = 265) = 0.00$, p = 1.00. There were no differences between the two studies on caregiver, $t(139) = 0.98$, p = .331, or youth alliance, t (144) = 0.83, p = .407. Study x log days interactions indicated that caregivers reported faster improvement on the BPC (1.27, p = .001) and TPA (0.22, p = .034), and youths on the BPC (1.54, p < .001) in Study 1. Pre-post outcome trajectories did not differ (p > .05).

Study 3

Method

Procedures

We emailed an online survey to the 28 Study 1 and 2 therapists; 6 could not be reached; 4 did not respond; and 19 responded (adjusted response rate = 86.36%). Responders and nonresponders did not differ on demographic or professional characteristics (p > .05).

Measures

Therapists reported professional and practice characteristics. We adapted the Therapist Coping Cat Follow-Up Interview (Chu et al., 2015) to examine therapist perceptions of FIRST. We asked therapists for their "overall impressions of FIRST for use in routine practice" (open-ended question), "how effective was FIRST for your project cases?" on a 10-point scale (0 = not at all, 4 = moderately, 9 = extremely), "how difficult was it for you to use FIRST with your project cases?" (same 10-point scale), and "have you used FIRST since the studies; if so, about what percentage of your current child and adolescent cases do you use FIRST with?"

Results

Study 1 therapists (n = 10) averaged 4.20 (SD = 1.23) years of clinical experience since their most recent degree

(30.00% master's, 70.00% doctoral); 70.00% were licensed; 60.00% were currently treating youths; 40.00% worked in community mental health agencies, 40.00% university, professional or medical settings, and 10.00% private practice. Study 2 therapists (n = 9) had on average 3.78 (SD = 0.97) years of clinical experience post-degree (44.44% master's, 55.56% doctoral); 77.78% were licensed; 33.33% were currently treating youths; 44.44% worked in university, professional or medical settings, 22.22% private practice, and 11.11% community mental health agencies. Study 1 and 2 therapists did not differ significantly on any of these characteristics (p > .05).

Responses to the open-ended question about use of FIRST in routine practice (9 from Study 1, 9 from Study 2 therapists) were uniformly positive (e.g., "easy to learn," "useful treatment for a variety of problems," "incredibly practical, useful, and versatile," "I still use it often with cases. It's straightforward, skills-based and brief, perfect for a lot of my child cases."). Numerical effectiveness ratings were similarly high for Study 1 (M = 7.11, SD = 1.27) and Study 2 (M = 7.00, SD = 1.00) therapists, $t(16) = 0.21$, p = .839, d = 0.10; and difficulty ratings were similarly low for Study 1 (M = 2.00, SD = 2.83) and Study 2 (M = 2.78, SD = 1.79) therapists, $t(17) = −0.71$, p = .489, d = 0.33. Of therapists who were currently treating youths, 83.33% (5/6) of Study 1 and 33.33% (1/3) of Study 2 therapists reported continued use of FIRST after the study, Fisher's exact test, p = .226, OR = 10.00 (large effect).

Discussion

We examined the acceptability, integrity, and clinical benefit of FIRST, a principle-guided treatment, following adaptations intended to meet some of the needs of youths and therapists in community settings. Findings suggest that the six-session FIRST can be delivered with integrity by novice trainees with limited prior therapy experience, and that this principle-guided approach may have utility across settings with variations in training and implementation.

We focus initially on FIRST's acceptability and feasibility across training and implementation variations. In both Study 1 and 2, youths and caregivers showed high alliance with therapists, although caregivers in the current studies reported somewhat lower alliance ratings than in the 2017 FIRST trial. However, session attendance was comparable to, and caregiver involvement and treatment completion rates were even higher than, the 2017 FIRST trial. Perhaps a brief treatment with a predetermined number of sessions may be more appealing to caregivers than treatments without a predictable duration, thus enhancing treatment

motivation and participation. As such, the adapted FIRST may be a useful option when treatment needs to be brief (e.g., attrition, non-specialty settings) without sacrificing client acceptability. Study 3 findings also suggest that FIRST may be acceptable to therapists. Both Study 1 and 2 therapists perceived FIRST as being highly effective and not difficult to implement. Given that the five FIRST principles have demonstrated positive outcomes for each of the target problems (Weisz et al., 2004), it was not surprising that Study 1 and 2 therapists reported comparable perceptions of effectiveness. However, given therapist concerns with implementing prescriptive treatment protocols (Addis & Krasnow, 2000) and that therapists often tailor treatment content and sequence in routine practice (Chu et al., 2015; Palinkas et al., 2013), that Study 1 and 2 therapists did not significantly differ on perceptions of implementation difficulty was more surprising. In the case of FIRST, the focus on tailoring each principle for all three target problem types provides considerable flexibility in application, even when the principles are presented in a fixed order. This flexibility may have eased implementation difficulty in Study 2 despite fixed content and sequence. Interestingly, despite comparable perceptions of effectiveness and implementation difficulty, our small N comparison showed continued post-study use of FIRST to be somewhat higher among Study 1 than Study 2 therapists (n.s.), suggesting the *testable possibility* that more therapist training and support to implement an individually-tailored FIRST may be associated with greater long-term FIRST adoption and use.

Regarding integrity, even with reduced initial training, ongoing support, and developer involvement, therapist adherence and competence were higher in the current studies than in the 2017 FIRST trial. FIRST's focus on a few core principles applicable across multiple target problems may enhance its likelihood to survive the "implementation cliff" (Weisz et al., 2014, p. 59) when developer oversight and support are reduced. This may help explain why in Study 2, when therapists had decreased flexibility, adherence was higher than in Study 1 where treatment was individualized. With six sessions, there may also be fewer opportunities to go "off-course" and novice trainees may have limited knowledge of, or experience with, competing treatment strategies. Evidently, FIRST – whether in its original highly individualized approach or in its streamlined, simplified version – may be delivered with integrity by therapists ranging from novice trainees to veteran therapists receiving varying levels of expert support.

The findings also provide preliminary evidence for FIRST's clinical benefit across an array of implementation

support conditions. In Studies 1 and 2, youths and caregivers reported improvements on session-by-session and pre-post measures on both internalizing and externalizing problems (despite the absence of parent training in Study 2 to specifically target externalizing problems). Outcome trajectory comparisons, however, suggest that there may be a tradeoff between clinical outcomes and therapist support for treatment individualization. When we compared Study 1 and the 2017 FIRST trial, which both involved individualization but differed on the amount of supervision dedicated to support individualization, Study 1 (two-hours weekly) outperformed the 2017 FIRST trial (one-hour weekly) on both youth and parent BPC total trajectories, suggesting that more therapist support may lead to faster improvement, even with fewer sessions. In Study 2, individualization was completely dropped to permit less supervision, and improvement was slower than in Study 1 and the 2017 FIRST trial. These results may reflect the loss of therapist ability to respond optimally to individual needs, a tradeoff for simplified therapist training and treatment implementation.

In contrast to the session-by-session BPC trajectories, youth-report pre-post effect sizes were notably larger in the 2017 FIRST trial than in either Study 1 and Study 2, and larger than corresponding parent-reports. Discrepancies between youth and parent reports are not unusual (e.g., Achenbach et al., 1987; De Los Reyes et al., 2015), but the lower youth-reported change on both TPA and YSR may indicate that six sessions, whether tailored or standardized in content, are simply inadequate to fully address youth concerns. Parent-report pre-post effects, on the other hand, were fairly consistent across all three studies. This was somewhat surprising given previous findings that, compared to protocols with predetermined content and sequence, modular treatments individually tailored to client needs are associated with faster improvement (Weisz et al., 2012). It warrants mentioning that the five FIRST principles are applicable to and have demonstrated effects for each of the targeted problems (Weisz et al., 2004). As such, presenting all five principles in a standardized way may increase the chances that youths receive all the principles, in contrast to flexible administration within which some principles may be omitted entirely.

That FIRST was implemented with integrity, is acceptable to youths, caregivers, and therapists, and was associated with symptom reduction in all three studies, suggests that this principle-guided approach may have the potential to benefit youths, families, and therapists across settings that vary in therapist training and ongoing support resources. In particular, FIRST's flexibility with regard to training may make it useful in

settings where limited support infrastructure for EBT implementation requires very efficient training and therapist uptake. The overall large Study 2 effect size suggests that even with relatively modest training and provider support, youths may still experience significant symptom relief from FIRST. Of course, we do not expect that this six-session treatment would be sufficient for highly severe, complex youths often served in community clinics. However, it might hold promise as a first line treatment within a stepped care model for subclinical youths who may experience meaningful symptom relief in five to six sessions. For example, many subclinical youths receive care in non-specialty settings (Burns et al., 1995), where the adapted FIRST may also help address contextual limitations (e.g., few sessions, few training resources). In fact, several therapists in Study 3 noted that FIRST might be used with subclinical cases or as a "first step" followed by longer-term care. Also, given that the six-session FIRST may be learned and implemented with fidelity by novice therapists, paraprofessionals or providers with limited experience in non-specialty settings might offer such abbreviated treatment to youths who do not have immediate access to care (e.g., on a waitlist for more intense services). Lastly, the Study 3 findings may have implications for graduate training. By the time of Study 3, Study 1 and 2 therapists had gained experience treating youths and families outside any research study or graduate program. The fact that they reported high effectiveness, low implementation difficulty, and continued use of FIRST, suggests that training novice therapists in FIRST may support continued EBT implementation in real world practice settings post graduate school.

The current studies had several limitations. Our samples were small compared to other benchmarking studies (e.g., Weersing & Weisz, 2002), and the open trial design precludes definitive conclusions about the clinical benefit of the six-session FIRST. Our samples also differed from the 2017 FIRST trial in age and number of sessions, although when we included these as covariates in our benchmarking analyses, the pattern of results did not change. Our samples were also majority White and well-educated, and baseline problem severity was not extreme – potentially suggesting characteristics somewhat more like a research than a clinical sample (Southam-Gerow et al., 2008). That said, our two samples were actually quite similar to the 2017 FIRST trial's community clinic sample (see Table 1). Study 1 and 2 therapists differed from typical community therapists including, intentionally, amount of prior clinical experience. This reduced the relevance of our findings to currently practicing credentialed therapists, but increased the relevance to trainees or novice therapists – an important reference group for our field as we develop strategies to incorporate EBTs within training programs and train a diverse clinical workforce. Of note, in an effort to address this limitation, we surveyed the Study 1 and 2 trainees, many licensed and practicing by Study 3, on their perceptions and use of FIRST in their current practice settings.

Given the combination of strengths and limitations, the current findings complement the 2017 FIRST trial by providing evidence that a principle-guided approach can be acceptable to stakeholders, delivered with integrity, and associated with symptom reduction across variations in therapist training, and ongoing support likely to be seen in community-based care contexts. The studies also provide preliminary evidence for the six-session FIRST, which may be well-suited to training settings and possibly to non-specialty settings where therapists lack extensive training, and implementation support and treatments are brief. Our findings provide further support for a future RCT of FIRST with diverse therapists, to provide a fair and rigorous evaluation of its impact with diverse youths and families who seek care in their communities.

Disclosure Statement

Authors Bearman and Weisz are co-authors of the FIRST treatment manual used in this study. No other potential conflict of interest was reported by the authors.

Funding

ORCID

Sarah Kate Bearman http://orcid.org/0000-0002-4920-1124
John R. Weisz http://orcid.org/0000-0002-5560-6814
Kristin M. Hawley http://orcid.org/0000-0001-9542-6095

References

Abramowitz, J. S. (2013). The practice of exposure therapy: Relevance of cognitive-behavioral theory and extinction theory. *Behavior Therapy*, *44*(4), 548–558. https://doi.org/10.1016/j.beth.2013.03.003

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213–232. https://doi.org/10.1037/0033-2909.101.2.213

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. 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(2), 331–339. https://doi.org/10.1037/0022-006X.68.2.331

Bearman, S. K., Bailin, A., Terry, R., & Weisz, J. R. (2019). After the study ends: A qualitative study of factors influencing intervention sustainability. *Professional Psychology: Research and Practice*, Advance online publication. https://doi.org/10.1037/pro0000258

Bearman, S. K., Herren, J., & Weisz, J. R. (2012). *Therapy integrity in evidence based interventions: Observational coding system, coding manual*. University of Texas at Austin.

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, 17(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187.x

Burns, B. J., Costello, E. J., Angold, A., Tweed, D., Stangl, D., Farmer, E. M., & Erkanli, A. (1995). Children's mental health service use across service sectors. *Health Affairs*, 14(3), 147–159. https://doi.org/10.1377/hlthaff.14.3.147

Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: A task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology*, 62(6), 631–638. https://doi.org/10.1002/jclp.20256

Cho, E., Wood, P. K., Taylor, E. K., Hausman, E. M., Andrews, J. H., & Hawley, K. M. (2019). Evidence-based strategies in youth mental health services: Results from a national survey of therapists. *Administration and Policy in Mental Health and Mental Health Services Research*, 46(1), 71–81. https://doi.org/10.1007/s10488-018-0896-4

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

Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., & Krull, J. L. (2010). Evaluation of the brief problem checklist: Child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology*, 78(4), 526–536. https://doi.org/10.1037/a0019602

Chorpita, B. F., Rotheram-Borus, M. J., Daleiden, E. L., Bernstein, A., Cromley, T., Swendeman, D., & Regan, J. (2011). The old solutions are the new problem: How do we better use what we already know about reducing the burden of mental illness? *Perspectives on Psychological Science*, 6(5), 493–497. https://doi.org/10.1177/1745691611418240

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

Chu, B. C., Talbott 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(1), 70–79. https://doi.org/10.1037/a0038000

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability for specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127–137.

Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology*, 18(3), 411–419. https://doi.org/10.1037/0893-3200.18.3.411

Davison, G. C. (2019). A return to functional analysis, the search for mechanisms of change, and the nomothetic-idiographic issue in psychosocial interventions. *Clinical Psychological Science*, 7(1), 51–53. https://doi.org/10.1177/2167702618794924

De Haan, A. M., Boon, A. E., de Jong, J. T., Hoeve, M., & Vermeiren, R. R. (2013). A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. *Clinical Psychology Review*, 33(5), 698–711. https://doi.org/10.1016/j.cpr.2013.04.005

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900. https://doi.org/10.1037/a0038498

Ehrenreich, J. T., Southam-Gerow, M. A., Hourigan, S. E., Wright, L. R., Pincus, D. B., & Weisz, J. R. (2011). Examining similarities and differences in characteristics of anxious and depressed youth in two different clinical contexts. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(5), 398–411. https://doi.org/10.1007/s10488-010-0328-6

Evans, S. W., & Weist, M. D. (2004). Commentary: Implementing empirically supported treatments in the schools: What are we asking? *Clinical Child and Family Psychology Review*, 7(4), 263–267. https://doi.org/10.1007/s10567-004-6090-0

Garland, A. F., Accurso, E. C., Haine-Schlagel, R., Brookman-Frazee, L., Roesch, S., & Zhang, J. J. (2014). Searching for elements of evidence-based practices in children's usual care and examining their impact. *Journal of Clinical Child & Adolescent Psychology*, 43(2), 201–215. https://doi.org/10.1080/15374416.2013.869750

Garland, A. F., Hawley, K. M., Brookman-Frazee, L., & Hurlburt, M. S. (2008). Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(5), 505–514. https://doi.org/10.1097/chi.0b013e31816765c2

Goldfried, M. R. (1980). Toward the delineation of therapeutic change principles. *American Psychologist*, 35(11), 991–999. https://doi.org/10.1037/0003-066X.35.11.991

Gomez, D., Bridges, A. J., Andrews, A. I., Cavell, T. A., Pastrana, F. A., Gregus, S. J., & Ojeda, C. A. (2014). Delivering parent management training in an integrated primary care setting: Description and preliminary outcome data. *Cognitive and Behavioral Practice*, 21(3), 296–309. https://doi.org/10.1016/j.cbpra.2014.04.003

Harpaz-Rotem, I., Leslie, D., & Rosenheck, R. A. (2004). Treatment retention among children entering a new

episode of mental health care. *Psychiatric Services*, *55*(9), 1022–1028. https://doi.org/10.1176/appi.ps.55.9.1022

Herschell, A., McNeil, C., & McNeil, D. W. (2004). Clinical child psychology's progress in disseminating empirically supported treatments. *Clinical Psychology: Science and Practice*, *11*(3), 267–288. https://doi.org/10.1093/clipsy/bph082

Hoffman, S., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science*, *7*(1), 37–50. https://doi.org/10.1177/2167702618772296

Jensen-Doss, A., Hawley, K. M., Lopez, M., & Osterberg, L. D. (2009). Using evidence-based treatments: The experiences of youth therapists working under a mandate. *Professional Psychology: Research and Practice*, *40*(4), 417–424. https://doi.org/10.1037/a0014690

Lai, M. H., & Kwok, O. M. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *The Journal of Experimental Education*, *83*(3), 423–438. https://doi.org/10.1080/00220973.2014.907229

Marchette, L. K., & Weisz, J. R. (2017). Practitioner review: Empirical evolution of youth psychotherapy toward transdiagnostic approaches. *Journal of Child Psychology and Psychiatry*, *58*(9), 970–984. https://doi.org/10.1111/jcpp.12747

McLeod, B. D., & Weisz, J. R. (2005). The therapy process observational coding system-alliance scale: Measure characteristics and prediction of outcome in usual clinical practice. *Journal of Consulting and Clinical Psychology*, *73*(2), 323–333. https://doi.org/10.1037/0022-006X.73.2.323

Ng, M. Y., & Weisz, J. R. (2016). Annual research review: Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry*, *57*(3), 216–236. https://doi.org/10.1111/jcpp.12470

Oddli, H. W., Nissen-Lie, H. A., & Halvorsen, M. S. (2016). Common therapeutic change principles as "sensitizing concepts": A key perspective in psychotherapy integration and clinical research. *Journal of Psychotherapy Integration*, *26*(2), 160–171. https://doi.org/10.1037/int0000033

Palinkas, L. A., Weisz, J. R., Chorpita, B. F., Levine, B., Garland, A. F., Hoagwood, K. E., & Landsverk, J. (2013). Continued use of evidence-based treatments after a randomized controlled effectiveness trial: A qualitative study. *Psychiatric Services*, *64*(11), 1110–1118. https://doi.org/10.1007/s10567-013-0145-z

Powell, B. J., Hausmann-Stabile, C., & McMillen, J. C. (2013). Mental health therapists' experiences of implementing evidence-based treatments. *Journal of Evidence-Based Social Work*, *10*(5), 396–409. https://doi.org/10.1080/15433714.2012.664062

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification*, *27*(3), 300–312. https://doi.org/10.1177/0145445503027003003

Southam-Gerow, M. A., Chorpita, B. F., Miller, L. M., & Gleacher, A. A. (2008). Are children with anxiety disorders privately referred to a university clinic like those referred from the public mental health system? *Administration and Policy in Mental Health and Mental Health Services Research*, *35*(3), 168–180. https://doi.org/10.1007/s10488-007-0154-7

Southam-Gerow, M. A., Prinstein, M. J., Burns, B. J., Costello, E. J., Angold, A., Tweed, D., Stangl, D., Farmer, E. M. Z., & Erkanli, A. (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), 1–6. https://doi.org/10.1080/15374416.2013.855128

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science & Practice*, *22*(4), 317–338. https://doi.org/10.1111/cpsp.12122

Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology*, *70*(2), 299. https://doi.org/10.1037/0022-006X.70.2.299

Weisz, J., Bearman, S. K., Santucci, L. C., & Jensen-Doss, A. (2017). Initial test of a principle-guided approach to transdiagnostic psychotherapy with children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *46*(1), 44–58. https://doi.org/10.1080/15374416.2016.1163708

Weisz, J. R., & Bearman, S. K. (2020). *Principle-guided psychotherapy for children and adolescents: The FIRST program for behavioral and emotional problems*. Guilford Press.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., Ugueto, A. M., Langer, D. A., & Hoagwood, K. E. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology*, *79*(3), 369–380. https://doi.org/10.1037/a0023307

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., … Gray, J. (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. https://doi.org/10.1001/archgenpsychiatry.2011.147

Weisz, J. R., & Gray, J. S. (2008). Evidence-based psychotherapies for children and adolescents: Data from the present and a model for the future. *Child and Adolescent Mental Health*, *13*(2), 54–65. https://doi.org/10.1111/j.1475-3588.2007.00475.x

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*(4), 729–816. https://doi.org/10.1016/j.chc.2004.05.006

Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with

usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry*, *70*(7), 750–761. https://doi.org/10.1001/jamapsychiatry.2013.1176

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

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science*, *14*(2), 216–237. https://doi.org/10.1177/1745691618805436

Weisz, J. R., Ng, M. Y., & Bearman, S. K. (2014). Odd couple? Reenvisioning the relation between science and practice in the dissemination-implementation era. *Clinical Psychological Science*, *2*(1), 58–74. https://doi.org/10.1177/2167702613501307

Woo, R. Y. (2018). *Navigating two alliances with evidence-based practices: Relating therapist competence and adherence to youth and caregiver therapeutic alliances* [Unpublished master's thesis]. University of Texas at Austin.