Adm Policy Ment Health (2017) 44:395–404
DOI 10.1007/s10488-016-0727-4

ORIGINAL ARTICLE

Behavioral Rehearsal for Analogue Fidelity: Feasibility in a State-Funded Children's Mental Health Initiative

Shannon Dorsey[1] · Aaron R. Lyon[2] · Michael D. Pullmann[2] · Nathaniel Jungbluth[1] · Lucy Berliner[1,3] · Rinad Beidas[4]

Published online: 10 March 2016

Abstract A substantial number of evidence-based treatments (EBTs) are available, but are delivered infrequently in public mental health. To improve the quality of care, some states and systems have focused on EBT training; however, these efforts have rarely included objective measurement of clinician fidelity because of feasibility issues. The primary goal of the current study was evaluating the feasibility of the behavioral rehearsal (BR) method to assess "analogue fidelity" in a children's mental health quality improvement initiative. Results indicated low—but representative—clinician participation. Participants demonstrated greatest improvement at post-training with maintenance or decreases at 6-months (post-consultation). Implications for future use of BR are discussed.

Keywords Fidelity · Training · Implementation · Consultation · Behavioral rehearsal

Findings presented at the NIH DI Conference, 2012 and the Seattle Implementation Research Conference, 2013.

Electronic supplementary material The online version of this article (doi:10.1007/s10488-016-0727-4) contains supplementary material, which is available to authorized users.

✉ Shannon Dorsey
dorsey2@uw.edu

[1] Department of Psychology, University of Washington, 335 Guthrie Hall, Box 351525, Seattle, WA 98195, USA

[2] Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

[3] Harborview Center for Sexual Assault and Traumatic Stress, Seattle, USA

[4] Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA

Introduction

One in five youth in the United States experience a diagnosable and clinically impairing psychiatric disorder (Merikangas et al. 2010). Less than 20 % of those with need will receive services, with only a small proportion receiving evidence-based treatments (EBTs; President's New Freedom Commission on Mental Health 2003). One of the biggest challenges currently facing the mental health field is the dissemination and implementation of EBTs to community settings (McHugh and Barlow 2010). A range of implementation strategies have been developed to improve implementation efforts (Powell et al. 2015). Among these strategies, training community clinicians in EBTs is one of the most frequently used (Addis and Krasnow 2000; Herschell et al. 2004; Rakovshik and McManus 2010; Williams et al. 2011), with research suggesting that active, experiential training, followed by additional supports (e.g., EBT consultation or supervision; other organizational supports) may be required to achieve clinician behavior change and EBT fidelity (Beidas and Kendall 2010; Herschell et al. 2010).

Common Elements Approaches

One barrier to clinician training in EBTs is that most EBTs are disorder-specific. This singular focus is problematic given that clinician caseloads in community mental health are typically comprised of youth with a wide range of disorders and high rates of comorbidity (Weisz and Gray 2008). Recent efforts to increase the implementation of effective practices have sought to shift the focus from training in individual EBTs to training in the *common practice elements* (e.g., exposure, cognitive processing)

across EBTs to improve usual care (Beidas et al. 2011; Chorpita et al. 2005; Garland et al. 2010). Such approaches also include training in cross-cutting Cognitive Behavioral Therapy (CBT) techniques (e.g., using standardized assessment, CBT model explanation, assigning/reviewing homework; Sburlati et al. 2011) that support treatment delivery (e.g., Weisz et al. 2012). Common elements approaches allow for the aggregation and dissemination of research knowledge, facilitate more flexible training and implementation, offer greater coverage for youth, and seem to have greater appeal among community providers (Bernstein et al. 2015; Chorpita and Daleiden 2009; Chorpita et al. 2005; Chorpita et al. 2015).

In Washington State, a common elements training approach informed by the work of Weisz and Chorpita (Chorpita et al. 2005; Weisz et al. 2012), the CBT+ Initiative (pronounced "CBT Plus"), has been funded by the state since 2009 (see Dorsey et al. 2014). In CBT+, clinicians learn CBT for the most common clinical problems in children (i.e., behavior problems, depression, anxiety, and trauma-related anxiety). Due to limited Initiative funding (i.e., $80,000–100,000 per year), evaluation efforts have predominantly relied on clinician self-report measures (see Dorsey et al. 2014; Lyon et al. 2014), which may be problematic in light of findings that self-report ratings can be higher than objective ratings (e.g., Hurlburt et al. 2010; Miller and Mount 2001; Nakamura et al. 2014) (for exceptions, see Chapman et al. 2013; Ward et al. 2013). Given that the ultimate goal of EBT implementation is improved outcomes for clients, EBT initiatives must demonstrate that trainees can deliver interventions with fidelity to ensure that clients served by trainees will have positive outcomes. Research suggests that high-fidelity service delivery is associated with more positive outcomes (e.g., Huey et al. 2000; Washington State Institute for Public Policy 2004). Fidelity refers to the "extent to which the intervention was implemented as intended" following training and includes—at a minimum—adherence and competence (Perepletchikova et al. 2007, p. 829).

As in CBT+, few other state or system-funded quality improvement efforts have included rigorous assessment of clinician fidelity following training. In contrast, externally funded research studies evaluating training effectiveness typically include objective, observational ratings of fidelity (Proctor et al. 2011), which can be both time-consuming and costly. There is growing recognition, however, that traditional research methods—direct observation and coding of therapist sessions with clients—may not be feasible for large scale implementation research and practice without substantial resources, yet, limited alternatives currently exist (Schoenwald et al. 2011; Schoenwald and Garland 2013).

Potential Promise of Behavioral Rehearsal Methodology

One innovative methodology to measure "analogue" fidelity is Behavioral Rehearsal (BR), which is defined as "a simulated interaction between a trainee and another individual" (Beidas et al. 2014, p. 2). In BR, the trainee engages in a role play in which he or she is asked to demonstrate the technique being evaluated. Role plays may be performed with an actor trained to respond in a standardized manner, which facilitates comparison of performance across individuals and time. BR may offer a more efficient assessment of fidelity than direct observational ratings (i.e., interactions with a client) and a more effective assessment than self-report (see Beidas et al. 2014 for more information on BR methodology in mental health). BR is commonly used in the medical field to assess physician competency (referred to as standardized patient methodology; see Shah et al. 2012). In children's mental health, BR was used to assess fidelity in two externally-funded research studies (see Beidas et al. 2012; Cross et al. 2011). In addition, Nakamura et al. (2014) employed BR to test the effectiveness of a pilot train-the-trainer program for CBT in a state-funded implementation study, without the resources of external-funding (i.e., suggesting feasibility). The study included BR and self-report measures of fidelity, with findings that although both suggested skill improvement among clinicians, self-report ratings were comparatively inflated. Taken together, these studies support the potential promise of BR methodology, but more research is needed on feasibility.

Study Goals

The primary goal of the current study was to assess feasibility of BR methodology to assess training effectiveness in a state-funded quality improvement initiative. Specifically, we were interested in whether or not trainees would volunteer to do BRs, retention over multiple assessment points, and representativeness of participating clinicians. Prior to the current study, trainee fidelity was evaluated using self-report measures and voluntary evaluation participation rates were approximately 60 %, with no significant differences found between trainees who did or did not participate in the evaluation (e.g., Lyon et al. 2014). For the current study, we hypothesized that including BR would result in a decline in evaluation enrollment rates, due to increased effort required (e.g., preparation, scheduling, time to do the BR), in addition to potential trainee anxiety or discomfort associated with engaging in BR (Beidas et al. 2013). To our knowledge, BR has not been used to evaluate clinician fidelity in large state-funded implementation initiatives. Our goal was to examine the

feasibility and utility of a relatively brief BR (less than 10 min each) that might be used in place of, or to supplement, clinician self-report.

The secondary goal was to assess training effectiveness by using BR to evaluate clinician analogue fidelity for two cross-cutting, core CBT techniques: *CBT Model Explanation* and *Homework Planning*. We hypothesized that CBT model explanation is a foundational technique, as clinicians need to understand and be able to explain CBT for varying presenting problems. *Homework planning* was selected because homework—defined as between session exercises in which clients practice skills learned in session (Kazantzis et al. 2005)—is both central to CBT and associated with enhanced therapy outcomes (see Kazantzis et al. 2000; Kazantzis et al. 2010 for meta-analytic reviews). However, usual care research in children's mental health indicates that homework assignment and review are substantially underutilized (Garland et al. 2010). Our evaluation focused on assessing analogue fidelity for these two techniques across the three internalizing presenting problems targeted in CBT+ (i.e., anxiety, depression, and posttraumatic stress [PTS]).

Method

Procedure

CBT+ Training and Consultation

CBT+ includes 3 days of active and experiential training (including multiple, in-training BRs with peer and trainer feedback) on the common practice elements that comprise the four CBT approaches (e.g., psychoeducation, cognitive restructuring, exposure) and core CBT techniques (e.g., agenda setting, homework assignment) (see Dorsey et al. 2014 for more information on the CBT+ approach and development). Approximately 100 clinicians are trained each year. Post-training, participants receive 6 months of biweekly telephone consultation (12 calls) led by a CBT expert and predominantly focused on case presentations. During each call, one to two clinicians present cases and receive expert and peer feedback. Calls also include some combination of follow up on previously presented cases (e.g., clinician implementation of recommendations, client response), time for clinician questions, consultant modeling, opportunity for clinician role play, and linkage to case-specific CBT resources (http://depts.washington.edu/hcsats/PDF/TF-%20CBT/CBT_Plus_NB.html). However, microanalysis of consultation has not been conducted (see Edmunds et al. 2013 for an example). CBT+ also offers yearly advanced trainings, a listserv, and supervisor-specific supports.

Recruitment

All procedures were reviewed and exempted by the Washington State IRB. Participants were recruited in October and November of 2011, when they registered for the mid-November CBT+ training. An email invitation described the voluntary (i.e., lack of participation would not impact training or consultation) CBT+ evaluation (i.e., self-report measures and BRs at three time points—pre-training, immediately post-training, after 6-months of consultation) and incentive for participation ($10 per time point). There were no exclusionary criteria.

Behavioral Rehearsals

Two structured BRs, each with versions for the three presenting problems, were developed by the authors (CBT experts) to assess participant adherence and competency in simulated clinical interactions with a child (see Beidas et al. 2014 for procedural details). All BR versions were evaluated by outside CBT experts to ensure that each was a realistic portrayal and to equate for level of difficulty. Research assistants were trained to portray youth experiencing symptoms of anxiety, depression, or post-traumatic stress (PTS). Clinicians were asked to participate in two 8-min BRs by telephone at each time point (total of 16 min). Twenty-four hours before the scheduled BR, clinicians received an email that included a client vignette with detailed instructions for what clinicians should do for each of the two BRs. The first BR involved explaining the CBT model to the child. The second involved collaboratively planning for a homework task, either exposure (anxiety and PTS) or pleasant activity scheduling (depression) (see Online Appendices for vignettes, clinician instructions, and scoring). The presenting problem focus (i.e., depression, anxiety, PTS) for the BRs at each of the three time points was counterbalanced across participants to control for potential presenting problem order effects (e.g., *Clinician A:* Pre-training—Depression; Post-training—Anxiety; Post-consultation—PTS; *Clinician B*: Pre-training—PTS; Post-training—Depression; Post-Consultation—Anxiety). BRs were digitally audio-recorded, transcribed, and later independently coded using two checklists developed by the study team to code adherence and skill (see subsequent descriptions in measures).

Participants

Participants were 41 clinicians and supervisors employed at public mental health clinics in Washington State, out of the 100 who registered for the fall 2011 CBT+ training. Participants were predominantly female (70.7 %) and Caucasian (78 %), with fewer African American (7.3 %),

Latino/a (4.9 %), and Asian (4.9 %) participants (two were missing data on race/ethnicity). Most were between 25 and 29 (24.4 %) or 30–39 (36.6 %) years old, with 14.6 % between 40 and 49 and 17 % over 50. In terms of role, 71.8 % were clinicians, 12.8 % were supervisors, and 7.7 % reported both roles. Self-reported experience with CBT varied, 9.8 % rarely (1–5 % of clients) used CBT, 4.9 % occasionally (6–25 % of clients), 17.1 % sometimes (26–50 % of clients), 31.7 % often (51–75 % of clients) and 22 % almost always (76 % or more of clients) used CBT.

Measures

Clinician Demographics and Background

An existing Washington State CBT+ Initiative self-report measure assessed demographic and background characteristics including race/ethnicity, age, gender, agency role, orientation, and CBT experience.

CBT Model Explanation: Adherence and Skill Checklist

Therapist adherence and skill for the first BR (BR1; explaining the CBT model) was assessed by coding three elements: (1) introducing the CBT triangle diagram; (2) teaching that thoughts, feelings, and behavior are linked (e.g., a specific thought leads to related feelings and behaviors); and (3) demonstrating that *changing* either thoughts or behavior changes the other related points on the triangle, ideally using a child-appropriate example and the triangle diagram. Each element was coded on a 7-point Likert scale from 0 (not present/non-adherent) and 1 (present/adherent; low skill) to 6 (present/adherent; high skill). An overall mean score representing both adherence and skill was calculated in line with frequently used coding measures in children's mental health research (e.g., McLeod and Weisz 2010). To aid interpretation, we considered scores 1–2 to be "low"; 3–4 to be "medium"; and 5–6 to be "high" analogue fidelity. Using a 70 % criterion (Beidas et al. 2012), scores 4.2 and higher (i.e., $\geq$70 % of a perfect score of 6) were considered to represent acceptable analogue fidelity. Coders were masked to time point and participant characteristics.

Homework Planning: Adherence and Skill Checklist

Therapist adherence and skill for the second BR (Homework Planning) were informed by elements captured on the homework item of the Cognitive Therapy Scale (Young and Beck 1980). The goal was to test clinician ability to help the client plan for use of a key behavioral change strategy between sessions. Five elements were coded: (1) Introducing the technique (exposure/pleasant activity scheduling); (2) providing a rationale for expected clinical benefit; (3) identifying activities for the technique (e.g., for exposure: identifying possible tasks to try at home); (4) role play/practicing the technique with the child (e.g., for pleasant activity scheduling: a brief mood-boosting activity); and (5) collaborative facilitators/barriers discussion (e.g., determining the best time, identifying specific facilitators of completion [e.g., parental reminder], addressing barriers to completion). The scoring procedure (i.e., 0–6) and interpretation of "medium" and "high" scores were identical to the first BR.

Coder Training and Reliability

The initial codebook for both BRs was developed by the first, second, fourth, and last authors (all CBT experts), spanning the three target presenting problems. All coders ($N = 5$) were post-baccalaureate or undergraduates. Training took approximately 17 h and included readings, lecture, and demonstration of techniques by the codebook developers, followed by coding meetings with the first author (group coding, discussion, manual refinement). Between meetings, BRs were coded independently and then discussed in the following meeting (see Beidas et al. 2014 for more information on the protocol). Once the codebook was finalized, prior to coding for the study, all coders met a single-rater, random effects, absolute agreement Intraclass Correlation Coefficient [$ICC_{(2,1)}$] threshold of 0.80 compared to the first author (2 additional hours of coding). To ensure continued high reliability throughout the process, at least 20 % of BR1 and BR2 audio files were double coded. Reliability was acceptable for BR1 and very good for BR2 (Cicchetti 1994), with absolute agreement $ICC_{(2,1)}$ of 0.68 and 0.82, respectively. Once trained, coders were able to code each 8-min BR in 15–30 min. Variation in coding time was typically due to variation in audio recording quality which affected audibility.

Data Analyses Plan

Descriptive analyses, Chi squares (χ^2), and t-tests were used to assess BR feasibility. To assess analogue fidelity, Longitudinal Multilevel Models (MLM; time point score nested within clinician) using Full Maximum Likelihood estimates, robust standard errors, and autoregressive covariance structures with heterogeneous variances (Singer and Willett 2003), were run for overall mean sum scores and individual items for each BR, including a linear and a quadratic time trend. These tested for changes over time from pre- to post-training and post-training to 6-months post-consultation. The intercept term was allowed to randomly vary, with time trends fixed due to a small sample

size ($N = 41$), which led to nonconvergence and unstable estimates in alternative models that included varying time trends. MLM results were replicated using Repeated-Measure Analyses of Variance (RM-ANOVAs) for the subsample with both BRs at all three time points ($n = 18$; 41 %) to perform a check on the MLM results.

Results

BR Feasibility

The 41 participants who agreed to participate in the evaluation represent 41 % of the 100 registrants for the CBT+ training. Of 41 participants, 35 (85 %) completed BRs for at least two of the three time points. Retention of BR participants from pre-training to post-training (about a 2-week period) was 82 %. Retention from pre-training to post-consultation (about a 6.5-month period) was 68 % for clinicians still employed at their organizations at the post-consultation time point (seven participants left their organizations during consultation). Only 18 participants (44 %) completed BRs at all three time points; however, one-third of this attrition was due to clinicians leaving their organizations during consultation and, therefore, no longer eligible for initiative or study participation (state-funding requirement).

BR Sample Representativeness

We explored potential differences in trainees who volunteered to participate ($n = 41$) and those who did not ($n = 59$) (e.g., generalizability of the sample). Crosstabulations with χ^2 tests and t-tests were run to examine baseline differences, with no statistically significant differences found between the two groups. Potential areas of difference examined were demographic (i.e., age, race/ethnicity, gender) and background characteristics (i.e., role [supervisor; therapist], level of education, number of years providing therapy), CBT experience, and receipt of CBT-specific supervision.

Missing Longitudinal Data for the BR Sample

Given the high rates of attrition in the BR sample, we explored attrition bias by comparing participants who completed BRs at all three time points ($n = 18$) to those who missed at least one time point ($n = 35$). Independent samples t tests and crosstabulations with χ^2 tests revealed no statistically significant differences in participant demographics, background characteristics, or in pre-training BR item or total scores.

Model Explanation Adherence and Competency

For the first BR, Model Explanation, both the overall mean score and the *Introducing the CBT Triangle* item score significantly increased from pre- to post-training and then significantly flattened from post-training to post-consultation (see Table 1; Fig. 1). For the *Demonstrating Change* item, scores significantly increased from pre- to post-training, and also to post-consultation. Scores did not significantly change for the *Linking Thoughts, Feelings, and Behaviors* item, although they trended toward improvement ($p = 0.059$). From pre- to post-training to 6-months post-consultation, the percentage of participants who scored in the acceptable or higher analogue fidelity range on overall mean score ($\geq$4.2) changed from 20.5 to 55.2 to 52.4 %. While only one participant (2.6 %) scored in the high range of 5 or above at pre-, about one-fifth to a quarter of participants scored in the high range at post-training (22.6 %) and at 6-months post-consultation (23.8 %).

Homework Planning Adherence and Competency

For the second BR, Homework Planning, scores for the overall mean and two items, *Introduce Specific Technique* and *Provide Rationale*, significantly increased from pre- to post-training and then significantly decreased to near pre-training levels at post-consultation. One item, *Identify Activities for the Technique*, increased from pre- to post-training and then showed a slight, but statistically significant decrease at post-consultation. There were no significant changes for two items: *Role Play/Practice* and *Facilitators/Barriers Discussion.* From pre- to post-training to 6-months post-consultation, the percentage of participants who scored in the acceptable range or higher on overall mean score ($\geq$4.2) changed from 5.3 to 9.1 to 4.5 %. None scored in the high range at pre-, post-training, or 6-months post-consultation.

RM-ANOVA Replication

RM-ANOVAs repeating these analyses with the subsample that completed BRs at all three time points ($n = 18$; 41 %) resulted in nearly identical findings to the full sample analyses with one exception: the item *Introducing the CBT Triangle* significantly increased at each follow-up time point in the RM-ANOVA (pre-$M(SD) = 3.12$ (2.18), post-training $M(SD) = 3.71$ (1.53), post-consultation $M(SD) = 4.53$ (1.51), $F_{(2,32)} = 3.59$, $p < 0.039$) instead of flattening as was found in the MLM analysis with all 41 participants. All RM-ANOVA results are available upon request.

Table 1 Analogue fidelity: multilevel model estimated overall and item scores, and slope significance values

	Pre-training	Post-training	6-month post-consultation	Pre-training variance	Linear slope p	Quadratic slope p
Model introduction						
Overall mean	2.83	3.88	4.15	0.67*	<0.001	0.033*
Introducing CBT triangle	2.86	4.21	4.20	0.98	0.004	0.048*
Linking thoughts, feelings, and behaviors	3.56	4.31	4.62	0.73*	0.059	0.400
Demonstrating change	2.27	3.28	3.53	1.34*	0.006	0.141
Homework assignment						
Homework overall mean	2.29	3.12	2.51	0.05	<0.001	<0.001*
Introduce specific technique (BA/exposure)	1.36	2.39	1.56	0.26	0.011	0.018
Provide rationale for BA/exposure	1.84	2.93	1.86	0.23	0.003	<0.001*
ID activities for BA/exposure	3.82	4.7	4.22	0.12	<0.001	<0.001*
Role play or practice BA/exposure	1.53	1.97	1.37	0.64	0.354	0.297
Homework or practice plans for BA/exposure	2.95	3.43	3.43	0.40	0.302	0.437

* $p < 0.05$

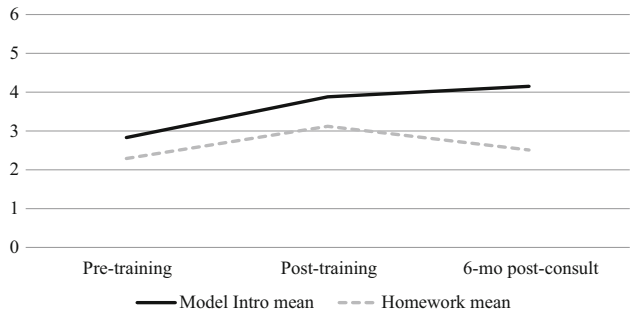

Fig. 1 Behavioral rehearsal mean scores at all time points. Scores on Model Introduction significantly increased from pre-training to post-training, then significantly flattened from post-training to 6-months post-consultation ($p < 0.05$). Scores on Homework significantly increased from pre-training to post-training, and then significantly decreased to near pre-training levels ($p < 0.05$)

Discussion

The BR methodology holds promise for measuring analogue fidelity in an effective and efficient way (Beidas et al. 2014), but to date has predominantly only been used in externally-funded research studies. This study was one of the first to test the feasibility of the BR methodology for assessing training effectiveness in a quality improvement initiative through the measurement of analogue fidelity. Our experience suggests that including BR to more rigorously assess analogue fidelity was feasible; however, fewer trainees participated and study attrition was higher in comparison to previous evaluation efforts (that did not include BR), suggesting that future efforts should potentially allocate greater resources to enrollment and retention. Using BR methodology, we determined that only half the clinicians achieved acceptable levels of analogue fidelity, and for only one of the two techniques assessed (i.e., model explanation). Feasibility and training effectiveness are discussed further below.

BR Feasibility

Fewer trainees overall (41 %) were willing to participate in an evaluation that included BR compared to past evaluations that included only self-report (e.g., 60 % enrollment), resulting in about a 30 % difference between enrollment rates. In addition, study attrition at the two follow-up time points was about 20 % higher when BR was included (see Lyon et al. 2014; Dorsey et al. 2014). Although we cannot definitively link enrollment and retention to BR inclusion, both variables were relatively stable across three prior evaluations (Dorsey et al. 2014). BR participants; however, seemed representative of trainees in that there were no significant differences between participants and non-participants across a range of characteristics (e.g., demographics, background, CBT experience). In comparison to some externally-funded training studies in which training and research participation were linked, per state mandate, trainees in the current study had access to the training and consultation regardless of evaluation participation and received only a small incentive ($10 per time point). Retention over time was relatively low, although a substantial proportion of the attrition was due to participants leaving their organizations, reflective of high turnover rates

in public mental health (e.g., Beidas et al. 2015b). Although we did not assess factors associated with initial and ongoing BR participation, other research suggests that clinicians who engage in BR may feel some anxiety or discomfort (Beidas et al. 2013). Additionally, productivity requirements likely played a role, in that clinicians were already out of the office for the 3-day training, so participating in the evaluation added more lost time just prior to and after the training. We estimated that the two BRs took approximately 20 min each (preparation; BR participation), representing about 40 min for each time point. Scheduling the assessment further from the training (2–3 weeks before/after) might improve enrollment. Alternatively, randomly or purposively sampling a subset of training registrants (e.g., based on demographics or background characteristics) with greater incentives for participation, could provide a general estimate of analogue fidelity using a smaller sample size that could then be generalized.

Given our somewhat low rates of enrollment and retention, our team identified strategies that might assist other implementation researchers in using BR even when resources are limited. These include: (1) requiring participation in BR to register for training (as long as this strategy itself does not negatively impact training participation); (2) embedding BR participation into the training using technology (see Schulman et al. 1999 for an example from medicine); (3) enlisting organizational leaders and clinical team managers to encourage and support participation; (4) providing additional incentives to clinicians and organizations who participate, such as access to webinars that are of value to participants (e.g., ethics training); and (5) finding ways to give clinicians and/or organizations feedback on performance so that BR is viewed as having professional development value.

One remaining challenge for using BR outside of externally funded research projects—when using standard research procedures—is the time involved in training actors and coders and the coding process. Without a state-university partnership, which provided access to volunteer student research assistants, faculty willing to devote additional effort, and a state funder that valued evaluation (e.g., contract included a 25 % paid research assistant position), time for these activities alone may prohibit BR use. Although other state-university partnerships exist (e.g., New York and New York University: www.ctacny.com), reliance on such partnerships would make BR an unlikely candidate for assessing training effectiveness. One option is to use BR as more of a quality control tool, without using objective, masked coders and a formal coding scheme. Trainers, consultants, or clinician's own supervisors could engage clinicians in BR and make general ratings (e.g., unacceptable, acceptable, very good), with ratings tied to specific follow-up activities (e.g., focus supervision time on challenging techniques; additional training/reading). One study with external, expert consultants demonstrated that BR inclusion in consultation predicted greater EBT delivery in the next session (Bearman et al. 2013).

Ongoing and soon-to-be-initiated research by our team will provide some answers to guide decisions about BR potential, effectiveness, cost, and particularly, how BR can be used as a quality control mechanism in everyday work settings. The first author is currently testing BR as a supervision tool in community mental health, as it can serve as both an assessment method and a mechanism for building competency if the clinician receives feedback on his/her effort (Dorsey et al. 2013). Anecdotally, supervisors and clinicians have reported BR as a worthwhile skill-building activity. This study will provide guidance on the effectiveness of BR under more feasible conditions—when used as part of supervision by community-based supervisors—in improving in clinician fidelity. Additional research on optimizing BR feasibility, testing BR validity, and developing other efficient methods of fidelity assessment are needed. Findings from the Nakamura et al. (2014) pilot study suggested that participants overestimate their own performance, in comparison to results from coded BR, which parallels findings from research comparing objectively coded therapy sessions to self-report (e.g., Hurlburt et al. 2010). Two of the authors (RB and SD) will be conducting a study that should offer some information on both accuracy and cost-effectiveness the BR methodology.

Training Effectiveness

Inclusion of BR allowed for objective evaluation of skill in two core CBT techniques for a small, but representative group of trainees. Results indicate that for both techniques, clinicians gained skills with training, moving from mean "minimal/low" to "medium" levels of analogue fidelity. After 6 months of consultation, improvement in Model Explanation was maintained, but fidelity for Homework Planning declined to near pre-training levels (i.e., to mean "low" levels). A little over 50 % achieved acceptable levels of analogue fidelity for Model Explanation at post-training and post-consultation, which is mostly in line with other studies using a similar training and supervision approach (e.g., Beidas et al. 2012; 61 %), particularly given the more complex focus on CBT for multiple presenting problems. Although disappointing, these rates may be the best that can be achieved with current approaches to implementation, which may not pay enough attention to other contextual factors that can explain as much or more of the variance in clinician practice (e.g., organizational climate, see Beidas et al. 2015a). In contrast to Model Explanation, very few clinicians achieved acceptable levels

of analogue fidelity at either time point for Homework Planning. These rates are particularly disappointing, but are not entirely out of line with other studies of technique use among EBT-trained clinicians in public mental health settings (McLeod et al. 2015).

Differences in Analogue Fidelity Between Model Explanation and Homework Planning

One question these findings raise is why analogue fidelity was higher for Model Explanation than for Homework Planning. We viewed Model Explanation, which requires understanding the CBT theoretical model of change and how to explain it to a client, as a foundational skill that underpins CBT treatment delivery. BR results suggest that, at least for this small sample, the majority of trainees acquired and maintained acceptable analogue fidelity. Future research should investigate factors associated with trainees who do not attain acceptable levels during the training and consultation period (e.g., type of supervision received, implementation climate). In contrast to maintained improvement for Model Explanation, clinician analogue fidelity for Homework Planning did not operate as expected, based on findings from prior reviews of clinician training (Beidas and Kendall 2010; Herschell et al. 2010). We see four potential reasons for lower rates of improvement specifically for Homework Planning. First, we conceptualized this technique as more complex in that it required greater tailoring both to the treatment target (depression, anxiety, PTS) and to the individual child (e.g., identification strategies to facilitate completion). Our findings of low scores at pre-training and limited improvements over time may partially explain prior research demonstrating that homework planning infrequently occurs in public mental health by both objective (Garland et al. 2010) and clinician report (Dorsey et al. 2014). Potentially, if clinicians struggle with this technique (as our results suggest), they may be less likely to use it in their clinical practice.

Second, homework planning may also be less consistent with other clinical orientations (e.g., client-centered) and therefore less likely to occur (Brookman-Frazee et al. 2009, 2010). In contrast to CBT, other orientations view what happens in session over what happens outside of session (e.g., homework) as the mechanism of change. Third, some research suggests that clinicians view homework as unrealistic for some families with high levels of outside stressors (Ringle et al. 2015). Finally, our speculation is that homework planning and review typically receive too little attention in training and consultation and need more targeted initial training and ongoing support to increase effective use. In our study, improvement with training but regression over the consultation period suggests that consultants need to focus more on building homework planning skills and providing opportunities for expert modeling and trainee behavioral rehearsal on the calls. Other work suggests homework is overlooked during supervision in community mental health (Accurso et al. 2011). As a result, some studies are attempting to better integrate coverage of core CBT techniques, like homework, into supervision (Dorsey et al. 2013). Regardless, these findings suggest that a measure of quality control (like BR) is essential to assess clinician fidelity over time.

Limitations

The primary limitations are related to assessing training effectiveness using BR given low enrollment that resulted in fewer than half of the trainees participating in the evaluation and high attrition. Although the evaluation with BR was more rigorous than prior evaluations, we may have traded greater internal validity for decreased external validity, despite the fact that participants and non-participants did not have significant differences. Second, given resources required in BR, we only examined analogue fidelity to two CBT techniques. Many other techniques and practice elements were taught and not assessed. Third, we did not document or analyze training and consultation time focused on the two techniques assessed or on BR itself (see Edmunds et al. 2013 for examples of this kind of micro-analysis). Fourth, BR does not include an evaluation of clinicians' actual fidelity with clients or client outcomes. Fifth, our evaluation did not test BR as a quality assurance or competency development tactic in the work setting, which are important next steps for research on the feasibility of BR to improve clinician fidelity. Finally, our assessment of feasibility did not include an estimate of cost, time, or reasons trainees chose not to participate (time, incentive, anxiety, etc.).

Conclusion

In general, our experience and findings suggest that BR has the potential to add value as part of an evaluation of quality improvement initiatives (e.g., identification of techniques/elements that are more challenging; training effectiveness). However, future efforts should make use of lessons learned by our team to strengthen external validity by increasing enrollment and retention. BR offers an intermediary outcome option between self-report and assessment of actual fidelity with clients; however, to make the time investment worthwhile—on both the part of the research team and participants—improving enrollment and retention is necessary.

Acknowledgments This publication was made possible in part by funding from the Washington State Department of Social and Health Services, Division of Behavioral Health Recovery, and from Grant Numbers R01 MH095749, K23 MH099179, and K08 MH095939, awarded from the National Institute of Mental Health (NIMH). Drs. Dorsey, Lyon, and Beidas are investigators with the Implementation Research Institute (IRI) (2011-2014), at the George Warren Brown School of Social Work, Washington University in St. Louis; through an award from the National Institute of Mental Health (R25 MH080916) and the Department of Veterans Affairs, Health Services Research & Development Service, Quality Enhancement Research Initiative (QUERI).

References

Accurso, E. C., Taylor, R. M., & Garland, A. F. (2011). Evidence-based practices addressed in community-based children's mental health clinical supervision. *Training and Education in Professional Psychology, 5*(2), 88.

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.

Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., & Research Network on Youth Mental Health. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(6), 518–529.

Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don't tell me: Behavioral rehearsal asa training and analogue fidelity tool. *Cognitive and Behavioral Practice, 21*(1), 1–11.

Beidas, R. S., Edmunds, J. M., Cannuscio, C. C., Gallagher, M., Downey, M. M., & Kendall, P. C. (2013). Therapists perspectives on the effective elements of consultation following training. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(6), 507–517. doi:10.1007/s10488-013-0475-7.

Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services, 63*(7), 660–665.

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 & Practice, 17*, 1–30.

Beidas, R. S., Koerner, K., Weingardt, K. R., & Kendall, P. C. (2011). Training research: Practical recommendations for maximum impact. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(4), 223–237. doi:10.1007/s10488-011-0338-z.

Beidas, R. S., Marcus, S., Aarons, G. A., Hoagwood, K. E., Schoenwald, S., Evans, A. C., et al. (2015a). Predictors of community therapists' use of therapy techniques in a large public mental health system. *JAMA Pediatrics, 169*(4), 374–382. doi:10.1001/jamapediatrics.2014.3736.

Beidas, R. S., Marcus, S., Wolk, C. B., Powell, B., Aarons, G. A., Evans, A. C., et al. (2015b). A prospective examination of clinician and supervisor turnover within the context of implementation of evidence-based practices in a publicly-funded mental health system. *Administration and Policy in Mental Health and Mental Health Services Research,*. doi:10.1007/s10488-015-0673-6.

Bernstein, A., Chorpita, B. F., Daleiden, E. L., Ebesutani, C. K., & Rosenblatt, A. (2015). Building an evidence-informed service array: Considering evidence-based programs as well as their practice elements. *Journal of Consulting and Clinical Psychology, 83*(6), 1085–1096. doi:10.1037/ccp0000029.

Brookman-Frazee, L., Garland, A. F., Taylor, R., & Zoffness, R. (2009). Therapists' attitudes towards psychotherapeutic strategies in community-based psychotherapy with children with disruptive behavior problems. *Administration and Policy in Mental Health and Mental Health Services Research, 36*(1), 1–12. doi:10.1007/s10488-008-0195-6.

Brookman-Frazee, L., Haine, R. A., Baker-Ericzén, M., Zoffness, R., & Garland, A. F. (2010). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration and Policy in Mental Health and Mental Health Services Research, 37*(3), 254–269. doi:10.1007/s10488-009-0244-9.

Chapman, J. E., McCart, M. R., Letourneau, E. J., & Sheidow, A. J. (2013). Comparison of youth, caregiver, therapist, trained, and treatment expert raters of therapist adherence to a substance abuse treatment protocol. *Journal of Consulting and Clinical Psychology, 81*(4), 674–680. doi:10.1037/a0033021.

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

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.

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

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. doi:10.1037/1040-3590.6.4.284.

Cross, W. F., Seaburn, D., Gibbs, D., Schmeelk-Cone, K., White, A. M., & Caine, E. D. (2011). Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills. *The Journal of Primary Prevention, 32*(3–4), 195–211.

Dorsey, S., Berliner, L., Lyon, A. R., Pullmann, M., & Murray, L. K. (2014). A statewide common elements initiative for children's mental health. *Journal of Behavioral Health Services and Research*. doi:10.1007/s11414-014-9430-y.

Dorsey, S., Pullmann, M. D., Deblinger, E., Berliner, L., Kerns, S. E., Thompson, K., et al. (2013). Improving practice in community-based settings: A randomized trial of supervision—study protocol. *Implementation Science, 8*(89), 1–12.

Edmunds, J. M., Kendall, P. C., Ringle, V. A., Read, K. L., Brodman, D. M., Pimentel, S. S., & Beidas, R. S. (2013). An examination of behavioral rehearsal during consultation as a predictor of training outcomes. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(6), 456–466. doi:10.1007/s10488-013-0490-8.

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

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of

psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review, 30*, 448–466.

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

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*(3), 451.

Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2010). Child and family therapy process: Concordance of therapist and observational perspectives. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 230–244.

Kazantzis, N., Deane, F. P., & Ronan, K. R. (2000). Homework assignments in cognitive and behavior therapy: A meta-analysis. *Clinical Psychology: Science and Practice, 7*(2), 189–202. doi:10.1093/clipsy.7.2.189.

Kazantzis, N., Deane, F. P., Ronan, K. R., & L'Abate, L. (Eds.). (2005). *Using homework assignments in cognitive behavior therapy*. New York: Routledge.

Kazantzis, N., Whittington, C., & Dattilio, F. (2010). Meta-analysis of homework effects in cognitive behavioral therapy: A replication and extension. *Clinical Psychology: Science and Practice, 17*(2), 144–156. doi:10.1111/j.1468-2850.2010.01204.x.

Lyon, A. R., Dorsey, S., Pullman, M., Silbaugh-Cowdin, J., & Berliner, L. (2014). Clinician use of standardized assessments following a common elements psychotherapy training and consultation program. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(1), 47–60.

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

McLeod, B. D., Smith, M. M., Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2015). Measuring treatment differentiation for implementation research: The Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale. *Psychological Assessment, 27*(1), 314–325. doi:10.1037/pas0000037.

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

Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics, 125*(1), 75–81.

Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy, 29*, 457–471. doi:10.1017/S1352465801004064.

Nakamura, B. J., Selbo-Bruns, A., Okamura, K., Chang, J., Slavin, L., & Shimabukuro, S. (2014). Developing a systematic evaluation approach for training programs within a train-the-trainer model for youth cognitive behavior therapy. *Behaviour Research and Therapy, 53*, 10–19. doi:10.1016/j.brat.2013.12.001.

Outcome Evaluation of Washington State's Evidence-Based Programs for Juvenile Offenders, January 2004. Washington State Institute for Public Policy, Report #04-01-1201.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829.

Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., et al. (2015). A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science, 10*, 21. doi:10.1186/s13012-015-0209-1.

President's New Freedom Commission on Mental Health. (2003). *Achieving the promise: Transforming mental health care in America.*

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., et al. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76.

Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. *Clinical Psychology Review, 30*(5), 496–516.

Ringle, V. A., Read, K. L., Edmunds, J. M., Brodman, D. M., Kendall, P. C., Barg, F., & Beidas, R. S. (2015). Barriers to and facilitators in the implementation of cognitive-behavioral therapy for youth anxiety in the community. *Psychiatric Services, 66*(9), 938–945.

Sburlati, E. S., Schniering, C. A., Lyneham, H. J., & Rapee, R. M. (2011). A model of therapist competencies for the empirically supported cognitive behavioral treatment of child and adolescent anxiety and depressive disorders. *Clinical Child and Family Psychology Review, 14*(1), 89–109.

Schoenwald, S. K., & Garland, A. F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment, 25*(1), 146–156.

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

Schulman, K. A., Berlin, J. A., Harless, W., Kerner, J. F., Sistrunk, S., Gersh, B. J., et al. (1999). The effect of race and sex on physicians' recommendations for cardiac catheterization. *New England Journal of Medicine, 340*(8), 618–626. doi:10.1056/NEJM199902253400806.

Shah, H., Rossen, B., Lok, B., Londino, D., Lind, S. D., & Foster, A. (2012). Interactive virtual-patient scenarios: An evolving tool in psychiatric education. *Academic Psychiatry, 36*(2), 146–150. doi:10.1176/appi.ap.10030049.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Ward, A. M., Regan, J., Chorpita, B. F., Starace, N., Rodriguez, A., Okamura, K., & Research Network on Youth Mental Health. (2013). Tracking evidence based practice with youth: Validity of the MATCH and Standard Manual Consultation Records. *Journal of Clinical Child & Adolescent Psychology, 42*(1), 44–55.

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., & 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., & Gray, J. S. (2008). Evidence-based psychotherapy for children and adolescents: Data from the present and a model for the future. *Child and Adolescent Mental Health, 13*(2), 54–65. doi:10.1111/j.1475-3588.2007.00475.x.

Williams, C., Martinez, R., Dafters, R., Ronald, L., & Garland, A. (2011). Training the wider workforce in cognitive behavioural self-help: The SPIRIT (Structured Psychosocial Interventions in Teams) training course. *Behavioural and Cognitive Psychotherapy, 39*(02), 139–149.

Young, J. & Beck, A. T. (1980). *Cognitive therapy scale rating manual*. Unpublished manuscript, University of Pennsylvania, Philadelphia, PA.