CrossMark

ORIGINAL ARTICLE

A Non-randomized Comparison of Strategies for Consultation in a Community-Academic Training Program to Implement an Evidence-Based Psychotherapy

Shannon Wiltsey Stirman[1] · Kristin Pontoski[2] · Torrey Creed[2] · Regina Xhezo[3] · Arthur C. Evans[3] · Aaron T. Beck[2] · Paul Crits-Christoph[2]

Abstract Despite the central role of training and consultation in the implementation of evidence-based psychological interventions (EBPIs), comprehensive reviews of research on training have highlighted serious gaps in knowledge regarding best practices. Consultation after initial didactic training appears to be of critical importance, but there has been very little research to determine optimal consultation format or interventions. This observational study compared two consultation formats that included review of session audio and feedback in the context of a program to train clinicians (n = 85) in community mental health clinics to deliver cognitive therapy (CT). A "gold standard" condition in which clinicians received individual feedback after expert consultants reviewed full sessions was compared to a group consultation format in which short segments of session audio were reviewed by a group of clinicians and an expert consultant. After adjusting for potential baseline differences between individuals in the two consultation conditions, few differences were found in terms of successful completion of the consultation phase or in terms of competence in CT at the end of consultation or after a 2 year follow-up. However, analyses did not support hypotheses regarding non-inferiority of the group consultation condition. While both groups largely maintained competence, clinicians in the group consultation condition demonstrated increases in competence over the follow-up period, while a sub-group of those in the individual condition experienced decreases. These findings, if replicated, have important implications for EBP implementation programs, as they suggest that observation and feedback is feasible in community mental health setting, and that employing this method in a group format is an effective and efficient consultation strategy that may enhance the implementation and sustainability of evidence-based psychotherapies.

Keywords Implementation · Evidence-based psychological treatments · Training · Consultation

✉ Shannon Wiltsey Stirman
sws1@stanford.edu; shannon.wiltsey-stirman@va.gov

Kristin Pontoski
pontoskk@mail.med.upenn.edu

Torrey Creed
tcreed@mail.med.upenn.edu

Regina Xhezo
Regina.Xhezo@phila.gov

Arthur C. Evans
Arthur.C.Evans@phila.gov

Aaron T. Beck
abeck@mail.med.upenn.edu

Paul Crits-Christoph
crits@mail.med.upenn.edu

[1] Dissemination and Training Division, National Center for PTSD and Department of Psychiatry and Behavioral Sciences, VA Palo Alto Healthcare System and Stanford University, 795 Willow Road (NC-PTSD), Menlo Park, CA 94025, USA

[2] Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, 3535 Market St., Philadelphia, PA 19104, USA

[3] Department of Behavioral Health and Intellectual DisAbility Services, 801 Market Street, 7th Floor, Philadelphia, PA 19107, USA

Introduction

Despite the development and identification of evidence-based psychological interventions (EBPIs) for a variety of disorders and problems and demonstration of their

successful transport into the community (Franklin et al. 2000; Juster et al. 1995; Simons et al. 2010), EBPIs have not been widely adopted in clinical practice. One common reason cited for this disconnect is the dearth of providers who are trained to deliver EBPIs (Weissman et al. 2006). In response to this shortage of providers, policymakers have issued mandates, provided incentives, and devoted billions of dollars to train providers from a variety of disciplines (e.g., social work, psychology, drug and alcohol counselors) in public mental health settings to utilize EBPIs (Karlin et al. 2010; McHugh and Barlow 2010). EBPIs tend to be complex, multisession treatment packages that involve provider's skillful selection and execution of a set of interventions (Carroll et al. 2010; Chorpita and Regan 2009). Training prepares providers to deliver EBPIs, and has been shown to increase both consumer access and clinician fidelity (adherence to the protocol and competence, or skill) to EBPIs (Feldstein et al. 2008; Fixsen et al. 2005; Stirman et al. 2004).

Comprehensive reviews of research on training have highlighted serious gaps in knowledge regarding best training practices (Beidas and Kendall 2010; Herschell et al. 2010; Rakovshik and McManus 2010). Much of the previously conducted research on training strategies has compared in-person training strategies to the use of treatment manuals and internet-based training strategies (Herschell et al. 2010). Findings from these studies have indicated that manuals, workshops, or web-based trainings result in poorer training outcomes than training that involves consultation or supervision. Consultation is defined as ongoing support in the form of focused interaction with a specialist in an effort to increase competence in the area of the specialist's expertise (Edmunds et al. 2013). Intensive consultation or supervision after initial training has been the only training strategy to result in benchmark levels of treatment fidelity among the majority of providers who received these strategies in previous research (Baer et al. 2004; Beidas et al. 2012; Miller et al. 2004; Sholomskas et al. 2005). Most such research has been conducted with brief interventions such as motivational interviewing for substance use disorders. The impact of consultation on clinician fidelity to longer treatment protocols for other mental health disorders, which typically include a greater number of interventions and techniques, has not been explored as extensively. However, findings to date suggest that more intensive consultation is associated with better outcomes (Beidas et al. 2013), and researchers have found no substitute for this critical element of training (Herschell et al. 2010).

While a need for consultation after initial didactic training has been identified, there remains a need to better understand the impact of the organizational framework of supervision processes on training outcomes (Ögren 2009), and to identify processes by which clinicians can be trained

to deliver psychotherapy competently (Milne 2014). Both the supervision and consultation literatures have identified the need to determine optimal processes for preparing individuals to deliver psychotherapy competently. Some have suggested that supervision and consultation should foster deliberate practice, a process that has been hypothesized to contribute to the development of expertise in psychotherapy by allowing clinicians to work toward mastery in a well-defined, specific task, to receive immediate feedback, to have opportunities to repeat their efforts, and to exploit the opportunity for improvement afforded by errors (Lewandowsky and Thomas 2009). Edmunds et al. (2013) have proposed that the processes of instruction, case review, self-evaluation, and feedback can foster skill development.

The emphasis on feedback in supervision and consultation models, combined with findings that clinicians are not always able to accurately assess their level of skill or adherence to a psychotherapy protocol (Brosan et al. 2008; Tracey et al. 2014) support the importance of feedback when training clinicians. Researchers have considered observation and feedback to be a "gold standard" for psychotherapy training (Beidas and Kendall 2010), and an essential element of training psychotherapists for clinical trials.

A critical barrier to the widespread use of individual feedback is its feasibility (Rakovshik and McManus 2010). Large-scale training programs may be limited by the number of experts available to review sessions and provide feedback (Ruzek and Rosen 2009), and smaller public sector settings lack the considerable funds required to provide such intensive training. In an attempt to address the limitations of the gold standard training method and of consultation without observation, a method of group consultation with observation and feedback was developed in the context of a community-academic training program (Creed et al. 2014). This model allows expert EBPI training consultants to review portions of sessions each week with a group of providers, and offer individualized feedback in the context of the group meeting. A potential advantage to this model is that providers are exposed to a broader sample of case material and peer examples than they would be in individual consultation, with more specific and accurate feedback than they could receive in group consultation with no observation. However, research findings have suggested that group supervision requires a sense of safety and supportive relationships to facilitate learning (Fleming et al. 2010). In the context of settings such as community mental health settings, which are often characterized by high caseloads and stressful environments, clinicians may experience discomfort when receiving feedback or allowing others to observe their sessions (Stirman et al. 2012), and this might limit the benefits of feedback.

Most published studies have not included observation in comparisons of consultation to no consultation (e.g.,

Henggeler et al. 2013; Sholomskas et al. 2005). To date, very few studies have investigated how well specific consultation strategies such as feedback prepare practitioners to deliver and sustain EBPIs in the context of service delivery. In one study, Miller and colleagues compared several motivational interviewing training strategies, including workshop only and consultation with and without session review. While consultation with and without session review produced similar changes in clinician fidelity, only the strategy that included session review resulted in client behaviors that were associated with better treatment outcomes (Miller et al. 2004). However, this study employed individual consultation strategies among a set of highly motivated clinicians who were not trained in their work setting. The impact of observation and feedback strategies remains untested in service delivery settings.

Researchers and policymakers have indicated the need for natural experiments to advance the existing literature on effective and efficient strategies to enhance implementation. Research on implementation efforts in routine care settings can provide information about the effectiveness of strategies for training and consultation when deployed with more representative populations and under the constraints typically found in typical service settings. The goal of the current study was to compare two different consultation strategies, individual observation and feedback and group consultation with feedback, in the context of a program to implement CT in a large, urban behavioral healthcare system. Because each model includes observation and individualized feedback, we hypothesized that both consultation strategies would result in significant changes in CT competence, and that the less intensive group consultation condition would be non-inferior to the more time- and cost-intensive individual consultation condition. Furthermore, we expected that a similar proportion of clinicians in each group would achieve and maintain competence in CT.

Method

Setting and Treatment

The Beck Initiative (BI) is a collaborative effort between The University of Pennsylvania (Penn), The Philadelphia Department of Behavioral Health and Developmental disAbilities (DBHIDS), and DBHIDS agencies to train providers within the network in CT (Creed et al. 2014; Stirman et al. 2009). DBHIDS is a billion-dollar mental health system with over 300 provider agencies that are heterogeneous in size, structure, populations served, and availability of resources.

CT is a psychosocial treatment that identifies and changes dysfunctional thinking, behavior, and emotional responses by helping individuals develop skills for modifying beliefs, relating to others in different ways, and changing behaviors (Beck 2005). Since 2007, the BI has offered CT training workshops and follow-up consultation. During the time that data were collected, training focused on conveying essential CT strategies and applying them to depression and problems that commonly co-occur with depression in community mental health settings (e.g., substance abuse, suicidal ideation or behavior, anxiety) through the use of a transdiagnostic, case conceptualization approach. Training in case conceptualization entails helping providers to use a cognitive theoretical approach to organizing their client's relevant life experiences, core beliefs, and thinking patterns, thereby forming a foundation for understanding the client's current problems and planning effective CT interventions.

Consultation Program

The BI program includes the following elements (termed ACCESS; Stirman et al. 2010), with the first five elements occurring during the training and consultation phase: *Assess* needs and barriers (engage stakeholders and assess their needs and current EBP fidelity through meetings, surveys, work samples, interviews) *and Adapt* training to meet their needs as required, *Convey the basics* through initial didactics, *Consult* on case material and on strategies to overcome barriers during consultation and through meetings with key personnel, *Evaluate work samples* to provide feedback, *Study outcomes* in ways that are feasible and acceptable to the agency, and *Sustain* by anticipating and addressing future barriers, maintaining communication, recertifying trained clinicians every 2 years, and making a plan for training future staff. BI training consultants use the *ACCESS* approach to serve as external facilitators and consultants.

All participating clinicians first attended a 22 h didactic workshop. The workshop was immediately followed by a baseline assessment of CT competence, and then 6 months of participation in one of the two consultation models discussed below (see Fig. 1 for a timeline). In addition, submission of audio recordings of at least 15 sessions over the course of the 6-month consultation period was required for clinicians to complete the program successfully. A subset of session recordings were rated for fidelity for the purposes of providing feedback, determining successful program completion for participants, and for program evaluation. Two years after consultation ended, clinicians submitted a session for review if they wished to retain designation as a CT provider.

All participating agencies also agreed to allow trained clinicians to engage in an internal, peer-led CT consultation meeting twice per month after the 6-month expert-led

consultation phase ended (over the 2-year follow-up phase of this study, and beyond). BI training consultants maintained a limited presence at these meetings (e.g., attended once per quarter), and provided feedback on full CT sessions when clinicians applied for recertification in CT at the 2-year follow-up. With the approval of institutional and municipal Institutional Review Boards, this study utilized the ratings that were conducted in the context of the training program to compare the two consultation strategies that were employed during the first few years of the program, which focused on training clinicians to provide CT to adult outpatients.

Consultation Conditions

Individual Feedback

For the first year and a half of the program, an individual feedback model was employed at all participating agencies. Four to eight clinicians participated at each agency. Participating clinicians submitted audio recordings of sessions to their consultant each week for 6 months. In addition to assessing a baseline level of competence, consultants conducted fidelity ratings on the session using the Cognitive Therapy Rating Scale (Young and Beck 1980) and provided the numerical ratings along with specific feedback about case conceptualization and intervention delivery during an individual, 1 h telephone meeting. A 1 h in-person group consultation meeting was also held each week for group discussion and additional didactics, but no recordings were reviewed during these meetings.

Group Consultation and Feedback

After the first 18 months of the program, the format of consultation was changed to a group consultation model for all subsequently participating agencies. In this condition, cohorts of clinicians met weekly for 2 h with a consultant. At the first consultation meeting for each cohort, clinicians established ground rules for delivery of feedback and confidentiality regarding patient information and each other's skills in learning CT, with the goal of increasing their comfort level sharing work samples in the group setting. Each participant provided session recordings, and 5–10 min segments of recordings were reviewed by the group. At the beginning of the 6-month consultation phase of training, segments were often selected by the therapist to demonstrate efforts to deliver specific interventions. Later in the consultation phase, therapists were instructed to choose segments during which they experienced challenges or successes to share with the group. At times, consultants also played random segments of sessions to assess whether CT was being used throughout a given session. The

consultant and the other clinicians provided feedback on the content of the recording. Three full sessions were rated for fidelity; one that was submitted at the beginning of the consultation period (baseline), one at mid-consultation, and one at the end of the 6-month consultation period. Clinicians were provided with written feedback and CTRS scores for those sessions.

In both consultation conditions, consultants were post-doctoral-level psychologists with expertise in CT and prior training in supervision. Consultants received 6 months of training in consultation by co-leading a consultation group with a more experienced consultant prior to facilitating consultation groups. Consultants met regularly and followed standard procedures to ensure that consultation was consistent within each consultation condition. Most consultants led more than one consultation group, and most led at least one in each format. Depending on the size of the training cohort, one to two consultants facilitated consultation meetings.

Participating Agencies and Clinicians

The sample comprised clinicians who treated adults on an outpatient basis. Consultation was provided to an average of 8 clinicians (range 6–10), typically from a single agency. Nine of the participating agencies provided general outpatient services, but had clinics that also provided specialty care for specific conditions (e.g., substance abuse, severe mental illness), and three agencies provided care to adults with substance use disorders. Clinicians in the first six agencies that participated in the program ($n = 47$) received individual consultation. The subsequent six agencies ($n = 38$) participated in group consultation. The group consultation condition included one cohort of clinicians from two different agencies, both of which provided care to a similar population. The results section includes a more detailed description of participating clinicians' characteristics.

Measures

Clinician Characteristics

Prior to their participation in the (BI), clinicians provided demographic data and information about their prior training and experience with CT, caseload, theoretical orientation, and beliefs about cause of therapeutic change.

Fidelity Assessment Instrument

Because the training program emphasized a case conceptualization approach rather than the use of a specific manual, general competence in CT, rather than adherence, was selected as the primary outcome. The CTRS (Young

and Beck 1980) is an 11-item scale that measures cognitive-behavioral therapist competence. Expert raters evaluate a complete session and assess general therapeutic skills, the therapist's ability to structure the session, and the therapist's ability to intervene using the most appropriate CT methods. BI training consultants received training in conducting the ratings using a standard set of recordings and accompanying ratings, and attended monthly meetings in which a session was rated and discussed by all training consultants to ensure consistency in rating. Per-judge reliability was assessed periodically. Two-way, random effects single-measure intra-class correlations for absolute agreement were computed for the 15 sessions rated by all training consultants, with a resulting ICC = 0.61, indicating good agreement according to Cichetti's commonly cited conventions for interpretation of ICCs (Cichetti 1994). The convention for CT clinical trial clinician (CTRS total score ≥40; Shaw et al. 1999) was used as a threshold to indicate competency in the (BI), and raters agreed about achievement of this standard for 93 % of the 15 sessions that were rated by all raters.

Analytic Strategy

Because clinicians were not randomized into training conditions, we generated a propensity score for use as a covariate in statistical models (Eckardt 2012; Harder et al. 2010; Schafer and Kang 2008). A propensity score uses background data on non-randomized participants that might plausibly impact non-random assignment to a particular condition or the outcome variable of interest, in this case CT competence, to build a model to predict the probability that they will be assigned to one condition versus the other. By using it as a covariate in our analyses, we adjusted for possible pre-existing differences between the two consultation conditions. The propensity score was calculated to include the following variables, which may have impacted assignment to the individual or group consultation condition or the outcome of interest: baseline CTRS score, whether the clinician worked with a specific population as opposed to a general mental health population, the agency where the clinician worked, CT orientation, years of experience, and whether or not the clinician was a social worker.

We compared the consultation conditions in three ways. First, to compare the proportion of group members' achievement of CT competence, we conducted logistic regressions with the propensity scores and consultation condition included in the model. Next, we tested for noninferiority of the group versus individual consultation strategies (Blackwelder 2004; D'Agostino et al. 2003; Nacasch et al. 2014), examining whether the difference between the groups at post-consultation is smaller than a predetermined clinically meaningful difference (i.e., the noninferiority margin ["delta"]). To estimate a difference that would be clinically meaningful, the deltas were determined using the criterion for statistically significant and reliable change in competence established in a recent evaluation of a training program for CT, a difference of 4.5 points on the CTRS (Branson et al. 2015). We would regard a difference of less than 4.5 as supporting the hypothesis of non inferiority. A sample size of 84 is required to be 80 % sure that the lower limit of a one-sided 95 % confidence interval. Therefore, due to attrition over the course of the follow-up, only the post-consultation data is examined for non-inferiority analyses.

Additionally, to compare patterns of change in CT competence over time and to examine the data of therapists nested within agencies, mixed-effects hierarchical regression models were employed. These models accommodate several features of the data such as repeated measurements and nested data, and allow for evaluation of fixed and random effects (Bryk and Raudenbush 1992; Raudenbush 1997). All available session ratings were used for clinicians in the group consultation condition (typically 3–5 ratings; including baseline, mid-consultation, post-consultation, and the 2-year follow-up, plus any other sessions rated during fidelity monitoring, to provide feedback, or for program evaluation). Because clinicians in the individual consultation condition had weekly ratings but the group condition did not, we restricted analyses to include the individual consultation condition's baseline, mid-consultation, post-consultation and 2-year follow-up scores, along with one additional randomly selected session from each half of the consultation period for a total of five timepoints. Change from baseline to each assessment point was calculated.

We first examined the slope of change in CTRS scores over the course of the consultation period, which was the 6 months following the baseline assessment. Three-level models with repeated measurements (level-1), nested within therapists (level-2), nested within agency (level-3) were performed. The continuous outcome (i.e., change in CTRS rating from baseline) was modeled using maximum likelihood estimation. Because we modeled the slope of change from an individual's baseline competence, change at the first timepoint in the model was always 0, and the models were therefore specified with the no intercept command. The level-1 model included an uncentered linear term computed as the number of weeks between each assessment and the training workshop. The level-2 component of the model included a dichotomous indicator for training intervention condition (individual vs. group feedback), propensity scores, baseline competence, and agency as covariates, and cross-level interactions were specified between condition, baseline skill and the level 1 terms. We

examined the deviance statistic (a log-likelihood-based goodness-of-fit statistic) and the amount of within-subject variance accounted for to identify the best way to model change over time. A linear pattern of change resulted in the best fit. Robust standard errors were used to compute the test statistics.

To examine the slope of change in competence at the 2-year follow-up, we conducted a three-level piecewise model of changes in competence over time (Singer and Willett 2003). For this analysis, the time variable was recoded so that it was zero for every participant at the post-consultation assessment. Thus, the time variable for the 2-year follow-up time period was number of weeks since the post-consultation assessment, while this time variable for the assessment points prior to the end of consultation was set at −1 multiplied by the number of weeks between the baseline and post-consultation assessment. When these two time variables were entered into the Level 1 equation predicting the outcome variable, the regression coefficient for first time variable provided an estimate of change during the post-consultation period while the regression coefficient for the second time variable represented the difference in the rate of change over time between the consultation phase and the post-consultation phase. As with the prior analysis, the continuous outcome (i.e., change in CTRS rating from post-training) was modeled using maximum likelihood estimation. The level-2 component of the model included a dichotomous indicator for consultation intervention condition (individual vs. group feedback), propensity scores, post-consultation competence, and agency as covariates, and cross-level interactions were specified between condition, post-consultation competence, and the level 1 terms. As this analysis was intended to examine change in the post-consultation phase, we report the estimates for the first time variable (i.e., change over time during the follow-up period). All models were also run without propensity scores to compare results. Since clinicians were nested within agency in this sample, we also examined the proportion of variance in change over time accounted for by agency and by individual clinician in the 3-level models by calculating ICCs (rho). All statistical analyses were conducted using SPSS version 19 (IBM, Armonk, NY, USA).

Results

Clinician Characteristics

Table 1 describes characteristics of the 85 clinicians who treated adults and enrolled in the BI between 2007 and 2009, 25 % of the clinicians listed Cognitive Behavioral Therapy as their primary theoretical orientation. Although

Table 1 Sample characteristics

Age	M = 34; SD = 10
Gender	
Female	72 %
Male	28 %
Race/ethnicity	
White	69 %
African American	25 %
Hispanic/Latino	5 %
Education and discipline	
Psychology/Counseling (Master's)	54 %
Social work (Master's/LCSW)	33 %
Family therapy (MFT)	5 %
Other (e.g., Drug and Alcohol Counselor)	5 %
Psychology/psychiatry (MD/PhD)	3 %
Years of experience	
0–5	21 %
5–10	29 %
>10	52 %

most endorsed some prior exposure to CT (72 % through reading, and 28 % through didactics), the modal number of supervised hours of CT training was 0. The mean caseload size was 40 clients (SD = 28; range 3–85).

Group Comparisons

Test of Non-inferiority

Table 2 presents adjusted and unadjusted mean competence scores for baseline, mid-consultation, and post-consultation. To estimate group means, mixed-effects hierarchical regression analyses were conducted both with and without propensity scores as covariates and results were compared. Since the pattern of results did not differ substantially and model fit statistics indicated a marginally better fit for the model that included propensity scores, adjusted means were used for the calculations used to assess non-inferiority, and results of the adjusted models examining change over time are presented.

The tests of noninferiority did not support the hypothesis that the group consultation strategy was noninferior to the individual consultation strategy. At post-consultation, the upper endpoint of the one-sided 95 % confidence interval for the observed group differences is greater than our predetermined index ($t[69] = 2.1359$, $p = 0.0362$, 95 % CI of observed delta [0.259, 7.58]). The observed effect sizes for differences between groups were $d = -0.47$ in favor of the individual consultation condition at the post-consultation timepoint, but $d = 0.68$ for the 2-year follow-up, favoring the group consultation condition. Both of

Table 2 CTRS scores at baseline, mid-consultation, post-consultation, and 2 year follow-up

	Individual (n = 47)		Group (n = 38)		Overall (n = 85)	
	M	SD	M	SD	M	SD
Unadjusted model						
Baseline	25.80	10.90	20.84	6.00	23.58	9.33
Post-consultation	39.50	7.26	37.97	9.24	38.72	8.17
Two-year follow-up	42.59	9.03	51.02	9.41	45.74	9.93
Adjusted model						
Baseline CTRS	23.62	10.90	23.54	6.01	23.58	9.33
Post-consultation	40.58	7.26	36.66	9.24	38.62	8.27
Two-year follow-up	43.40	9.03	49.62	9.14	46.50	9.06

Adjusted models included propensity scores as a covariate and post-consultation and follow-up CTRS included Baseline CTRS as an additional covariate

these effect sizes are in the "medium to large effect" range of Cohen's d scale.

Comparison of Change Over Time

Figure 1 represents patterns of change across the 6 month consultation phase and the 2-year follow-up phase.

Consultation Phase Change in scores from baseline differed significantly over the course initial consultation, $F(1, 13) = 27.84$, $p < 0.001$, $b = 0.524$, $SE(b) = 0.115$. Baseline CTRS score was not a significant predictor of change, $F(1, 364) = 1.79$, $p = 0.192$, $b = 0.036$, $SE(b) = 0.027$. Although a significant baseline by training condition interaction was observed, $F(1, 381) = 10.20$, $p = 0.002$, $b = -0.419$, $SE(b) = 0.131$, a condition by time interaction was not observed, $F(1, 16) = 0.147$, $p = 0.706$, $b = -0.082$, $SE(b) = 0.215$. A condition by time by baseline competence interaction was also not significant, $F(1, 31) = 0.006$, $p = 0.940$, $b = -0.001$, $SE(b) = 0.009$. These results indicate that the slope of change in competence did not differ in the two consultation conditions, and that differences in baseline competence did not appear to impact the slope of change in competence.

Change between post-consultation and follow-up. At post-consultation, the CTRS scores differed between groups ($F 1, 299) = 5.97$, $p = 0.0156$, $b = 0.070$, $SE(b) = 0.028$, with clinicians in the individual consultation condition scoring higher than those in the group consultation condition. However, the results of the piecewise model to examine the slope of change over the 2-year follow-up period did not indicate a main effect of time over the follow-up, $F(1,84) = 0.004$, $p = 0.951$, $b = 0.001$, $SE(b) = 0.017$; or by consultation model, $F(1, 387) = 0.005$, $p = 0.945$, $b = 0.145$, $SE(b) = 2.11$. There was a significant consultation model by time interaction, $F(1, 67) = 7.53$, $p = 0.008$, $b = 0.407$, $SE(b) = 0.148$ and a
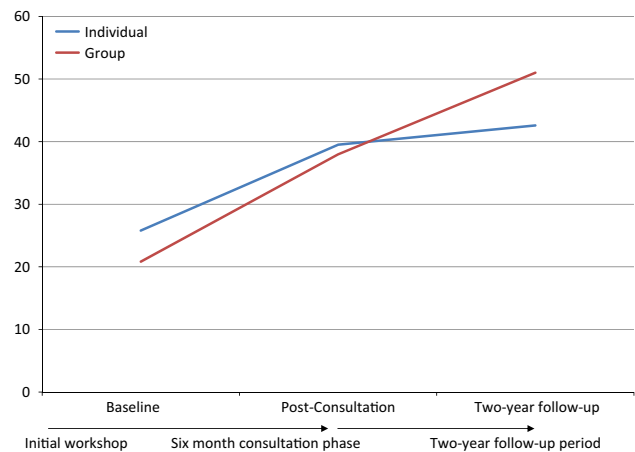

Fig. 1 CTRS scores by consultation model

significant post-consultation competence by model by time interaction, $F(1, 65) = 6.52$, $p = 0.013$, $b = -0.009$, $SE(b) = 0.003$. Closer examination of the data revealed greater increases in competence over time for clinicians in the group consultation model. To examine the interaction between competence, consultation condition, and time, we categorized clinicians as lower- and higher-scoring at post-consultation in two ways. First, we used a median split, and second, we divided into groups based on a CTRS score of 43 or below (indicating that clinicians scored less than "very good" on one or more CTRS items), and those scoring 44 or above. The pattern of change was the same each way that the groups were classified, so we present results (see Fig. 2) for the latter categorization strategy. High-performing clinicians in the individual consultation group experienced the largest reduction in CTRS scores (M = −6.00; SD = 1.69), and lower-performing clinicians in the individual consultation group experienced a slight decrease (M = −0.744; SD = 1.64). In contrast, both high- and lower-performing clinicians in the group

consultation conditions tended to experience increases in CTRS scores over time (M = 6.55; SD = 8.57 for the higher performing group; M = 7.88, SD = 2.16 for the lower-performing group).

The proportion of variance in change over time accounted for by agency, and by clinician, were both low. Intraclass Correlations (rho) for indicated that the proportion of variance accounted for by agency was (ρ = 0.009), and by clinician was (ρ = 0.001). These results suggest that neither agency nor clinician-level factors accounted for a substantial amount of variance in the change in competence over time. Therefore, we did not investigate other models that included interactions between consultation model, time, and agency.

Achievement of Competence

As Table 3 indicates, the individual and group consultation conditions did not differ significantly in terms of retention in consultation or successful completion of the training
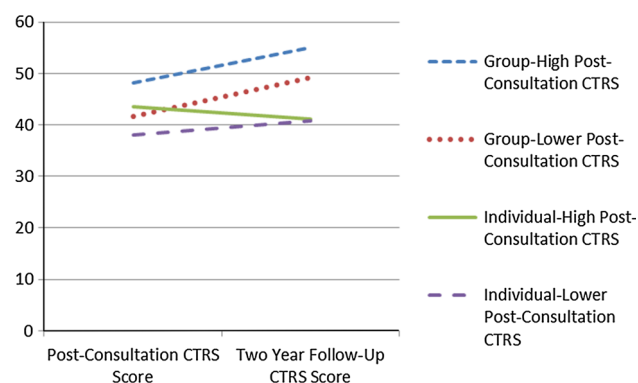

Fig. 2 Interaction between post-consultation CTRS score, consultation model, and time over a 2-year follow-up

program (defined as a CTRS score of 40 or above on a session selected from the last weeks of consultation). The 2-year follow-up data were also compared. Both recertification data (proportion of clinicians who participated in CT-oriented continuing education activities and ongoing agency-level CT consultation, and scored 40 or above on the CTRS at the 2 year follow-up) and total CTRS scores at the 2-year follow-up were examined. Slightly over half (53 %) of the clinicians were eligible for recertification after 2 years, as many of the participating clinicians had left the system or moved into a different role in which they no longer provided psychotherapy. Additionally, three agencies (two that had participated in the individual model and one that participated in the group model) did not pursue recertification. As Table 3 indicates, there was a non-significant trend indicating that clinicians in the individual consultation condition were more likely to be recertified. However, among clinicians who submitted recordings for recertification, the mean CTRS score was higher for the clinicians in the group consultation condition, and the difference was marginally significant.

Discussion

This study compared two strategies for providing post-workshop consultation in (CT) in a program to implement CT in an urban community mental health system. Because previous research has indicated the need for a more rigorous examination of the effective elements of post-didactic consultation and feedback (Beidas and Kendall 2010; Herschell et al. 2010), this study addresses an important area of interest in the field. In general, this study supports theory that suggests that clinicians' skill improves when they receive feedback and have the opportunity to make changes based on that feedback (Tracey et al. 2014).

Table 3 Comparison of consultation outcomes: adjusted and unadjusted results of logistic regressions

	Unadjusted rates			Unadjusted				Adjusted			
	Individual (n = 47)	Group (n = 38)	Overall (n = 85)	χ^2	p	Odds ratio	95 % confidence interval	χ^2	p	Odds ratio	95 % confidence interval
Retention	96 %	89 %	93 %	1.18	0.277	2.65	0.458–15.31	0.243	0.642	2.35	0.078–70.53
Successful completion	84 %	68 %	77 %	2.99	0.078	0.385	0.131–1.13	1.96	0.161	0.211	0.024–1.86
Two-year follow-up[a]	36 %/77 %	18 %/ 100 %	28 %	3.17	0.075	0.398	0.145–1.10	3.64	0.056	0.109	0.01–1.062

df = 1 for all analyses

Successful completion was defined as a CTRS score ≥40. Logistic regressions for adjusted models included propensity scores as a covariate

[a] Percentages are presented first for the overall sample and next for the portion of the sample eligible for recertification in each consultation condition

While our analyses did not support hypotheses regarding non-inferiority, we found that the majority of clinicians in both the group and individual model of consultation achieved the required level of competence in CT by the end of the 6-month consultation phase. The results of longitudinal analyses indicated that consultation model was not a significant predictor of change in CTRS skills over time during the 6-month consultation phase. Furthermore, neither clinician nor agency contributed substantially to the variance in changes in competence over time, suggesting that the consultation strategies that were tested can be effective across different community-based agencies and individuals. Previous research indicated differences in the organizational social context of agencies that participated in the BI (Stirman et al. 2013), and these findings support those of Kolko and colleagues, who found that clinicians could be successfully trained even within more negative organizational climates (Kolko et al. 2012).

This study, like others before it (Kolko et al. 2012; Swales et al. 2012), suggests that factors such as turnover may be an even greater threat to sustainment than erosion of skill. While nearly half of the clinicians were not eligible for recertification after 2 years due to turnover and other factors, competence status was generally maintained among those clinicians who had achieved competence at post-consultation and who were eligible for recertification 2 years after they completed consultation. It is important to note that because the 2-year follow-up recording was selected by the clinician for review, the 2-year data provides information about skill retention in CT, but not about the extent to which the clinicians actually deliver CT in their everyday practice after the consultation phase has ended. Thus, these results suggest that for clinicians who remain engaged in their agencies' internal, non-expert-led CT consultation, retention of competence can be expected. However, differences in the trajectory of the CTRS scores were found by consultation condition. Clinicians in the group consultation model achieved increases in skill over the 2-year follow-up, while competence scores for those who received individual consultation decreased over time. Furthermore, relative to clinicians in the group consultation condition, clinicians in the individual consultation condition tended to experience less change if they were lower-performing at post-consultation, and greater decreases in competence if they had higher scores on the CTRS at post-training.

These findings imply that clinicians in the group consultation model, who continued to review recordings of their CT sessions in a group peer-consultation format after the consultation phase was completed, may have continued to improve their skills with feedback from their peers. It is possible that this form of ongoing consultation prevented decreases in skill or regression to the mean. Those in the individual consultation group, who continued to discuss cases during peer consultation but were not trained to review recordings as a group, did not have opportunities to improve through feedback on their work samples. Previous research has provided some evidence that clinician reports of in-session behaviors and interactions may not be as accurate as observer review, and therefore it is possible that opportunities for corrective feedback are lost when consultation does not include review of, and feedback on work samples (Brosan et al. 2008).

Although a more rigorous, randomized comparison is necessary before firm conclusions can be drawn, these findings suggest that the group consultation model, which requires less time for expert consultants to review sessions and provide feedback, may still be a promising alternative to more time- and cost-intensive individual session review and feedback. In light of the contrast between the time and costs associated with expert consultation (2 h per week in the group model versus approximately 17 h per week for the individual model with a group of 8 clinicians), the group model is likely to be much more feasible in under-resourced settings, and should therefore be explored in further research. Additionally, our findings suggest that a model of training inclusive of audio review is feasible and provides an environment conducive to learning CT for clinicians from diverse backgrounds in very busy community mental health settings. At the time that the data were collected, the training program had almost no exclusion criteria for clinicians, and the participating clinicians had a range of prior clinical experience, limited prior CT training, and endorsed diverse theoretical orientations.

Although the challenges associated with the use of observational fidelity monitoring strategies are important to consider in implementation efforts (Schoenwald et al. 2011), data from this project indicates that recording sessions on a digital audio recorder and presenting them for feedback was feasible and acceptable to clinicians. In both consultation conditions, clinicians routinely recorded their CT sessions and submitted them for review and feedback, and rates of compliance were high. Previous mixed-methods research that included individuals from this sample indicated that clinicians believed that session observation was an important aspect of consultation (Stirman et al. 2012). Adding opportunities to review work samples and receive feedback in an efficient group format may be a viable and important enhancement to consultation after initial didactic training.

This study offers an important step in understanding the optimal structure of consultation models, but it has some practical and methodological limitations. A significant amount of therapist attrition and clinician selection of sessions at the 2-year follow-up precluded a thorough study

of the retention of CT skills over time or the extent to which clinicians continued to utilize CT in their day-to-day work. Because 2-year follow-up scores were only available for clinicians who had achieved competence during consultation, our sample size for these analyses is small and the analyses should be considered exploratory. These analyses also cannot determine whether clinicians who did not initially meet criteria for competence in CT would have experienced similar patterns of change. However, by examining those clinicians who successfully completed their initial consultation and remained in their agency's internal CT consultation, this study presents preliminary evidence that including audio review in an ongoing consultation format can be beneficial. Because we examined data from an ongoing training initiative, neither clinicians nor agencies were randomized to a training condition, and program evaluation data were utilized to measure outcomes. Although we adjusted for pre-existing differences between conditions that may have been present, due to the naturalistic design of this study, a randomized comparison will be necessary before more firm conclusions can be drawn about the superiority of a particular consultation strategy. While it is possible that in generating propensity scores to adjust for differences between groups, we did not include factors in the model that may have accounted for differences in baseline status or outcomes, we included a number of factors identified in prior literature that may have contributed to potential differences in outcomes (Eckardt 2012). We also cannot rule out historical confounds due to the timing of the two different interventions. However, no policy changes related to EBPIs or reimbursement policy were made within the system during the period during which the data were collected. Additionally, the consultants all had roughly the same level of experience in leading consultation groups, so differences in experience or expertise in terms of consultation and training is unlikely to account for differences in outcomes. It is possible, though, that levels of interest and enthusiasm for CT training differed between the two groups, if the awareness of the BI was higher, or its reception more positive among clinicians in the group consultation condition due to publicity within the network. Future research should include an examination of these potentially important factors. Finally, we employed the CTRS, the most commonly-used measure of competence for cognitive therapy in both research and training programs, but researchers have noted limitations to this measure, including its emphasis on general competency rather than competency in discrete CT skills and difficulty achieving high inter-rater reliability (Branson et al. 2015).

The results of this study suggest a number of potential directions for future research. Comparisons of consultation with and without an observation and feedback component, and studies to determine the minimum necessary amount of observation and feedback to achieve successful training outcomes are necessary. An examination of training at varying degrees of intensity based on baseline skill level could provide guidance for mental health systems that need to determine the most efficient strategy for training clinicians in EBPIs. Studies to assess the cost-efficiency of training and consultation strategies are also needed. Despite the limitations of this study, it capitalizes on a natural experiment to provide much-needed data on strategies for consultation. In light of the challenges of training and consultation in busy community mental health settings, the identification of effective and potentially scalable strategies can enhance efforts to make EBPIs available to underserved populations.

References

Baer, J. S., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence, 73*(1), 99–106. doi:10.1016/j.drugalcdep.2003.10.001.

Beck, A. T. (2005). The current state of cognitive therapy: A 40-year retrospective. *Archives of General Psychiatry, 62*(9), 953–959. doi:10.1001/archpsyc.62.9.953.

Beidas, R., Edmunds, J., Cannuscio, C., Gallagher, M., Downey, M. M., & Kendall, P. (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., Edmunds, J., Marcus, S., & Kendall, P. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services, 63*(7), 660–665. doi:10.1176/appi.ps.201100401.

Beidas, R., & 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. doi:10.1111/j.1468-2850.2009.01187.x.

Blackwelder, W. (2004). Current issues in clinical equivalence trials. *Journal of Dental Research, 83*(suppl 1), C113–C115.

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy, 68*, 19–26. doi:10.1016/j.brat.2015.03.002.

Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy, 36*(5), 581. doi:10.1017/S1352465808004438.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications Inc.

Carroll, K. M., Martino, S., & Rounsaville, B. J. (2010). No train, no gain? *Clinical Psychology: Science and Practice, 17*(1), 36–40. doi:10.1111/j.1468-2850.2009.01190.x.

Chorpita, B. F., & Regan, J. (2009). Dissemination of effective mental health treatment procedures: Maximizing the return on a significant investment. *Behaviour Research and Therapy, 47*(11), 990–993. doi:10.1016/j.brat.2009.07.002.

Cichetti, D. (1994). Guidelines, criteria, and rules of thumb for normed and standardized instruments in psychology. *Psychological Assessment, 6*(4), 284–290.

Creed, T. A., Stirman, S. W., Evans, A. C., & Beck, A. T. (2014). A model for implementation of cognitive therapy in community mental health: The Beck Initiative. *The Behavior Therapist, 37*(3).

D'Agostino, R. B., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues-the encounters of academic consultants in statistics. *Statistics in Medicine, 22*(2), 169–186.

Eckardt, P. (2012). Propensity score estimates in multilevel models for causal inference. *Nursing Research, 61*(3), 213–223.

Edmunds, J. M., Beidas, R. S., & Kendall, P. C. (2013). Dissemination and implementation of evidence-based practices: Training and consultation as implementation strategies. *Clinical Psychology: Science and Practice, 20*(2), 152–165. doi:10.1111/cpsp.12031.

Feldstein, A. C., Glasgow, R. E., & Smith, D. H. (2008). A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. *Joint Commission Journal on Quality and Patient Safety, 34*(4), 228–243. doi:10.2165/00148365-200806040-00001.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friendman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: The National Implementation Research Network, Louis de la Parte Florida Mental Health Institute, University of South Florida.

Fleming, L. M., Glass, J. A., Fujisaki, S., & Toner, S. L. (2010). Group process and learning: A grounded theory model of group supervision. *Training and Education in Professional Psychology, 4*(3), 194.

Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology, 68*(4), 594. doi:10.1037/0022-006X.68.4.594.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods, 15*(3), 234–249. doi:10.1037/a0019623.

Henggeler, S. W., Chapman, J. E., Rowland, M. D., Sheidow, A. J., & Cunningham, P. B. (2013). Evaluating training methods for transporting contingency management to therapists. *Journal of Substance Abuse Treatment, 45*(5), 466–474.

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*(4), 448–466. doi:10.1016/j.cpr.2010.02.005.

Juster, H. R., Heimberg, R. G., & Engelberg, B. (1995). Self selection and sample selection in a treatment study of social phobia. *Behavioral Research Therapy, 33*(3), 321–324.

Karlin, B. E., Ruzek, J. I., Chard, K. M., Eftekhari, A., Monson, C. M., Hembree, E. A., et al. (2010). Dissemination of evidence-based psychological treatments for posttraumatic stress disorder in the veterans health administration. *Journal of Traumatic Stress, 23*(6), 663–673. doi:10.1002/jts.20588.

Kolko, D. J., Baumann, B. L., Herschell, A. D., Hart, J. A., Holden, E. A., & Wisniewski, S. R. (2012). Implementation of AF-CBT by community practitioners serving child welfare and mental health: A randomized trial. *Child maltreatment*. doi:10.1177/1077559511427346.

Lewandowsky, S., & Thomas, J. L. (2009). Expertise: Acquisition, limitations, and control. *Reviews of human factors and ergonomics, 5*(1), 140–165.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*, 73–84. doi:10.1037/a0018121.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology, 72*(6), 1050–1062. doi:10.1037/0022-006X.72.6.1050.

Milne, D. (2014). Beyond the "Acid Test": A conceptual review and reformulation of outcome evaluation in clinical supervision. *American Journal of Psychotherapy, 68*(2), 213–230.

Nacasch, N., Huppert, J. D., Su, Y.-J., Kivity, Y., Dinshtein, Y., Yeh, R., & Foa, E. B. (2014). Are 60-minute prolonged exposure sessions with 20-minute imaginal exposure to traumatic memories sufficient to successfully treat PTSD? A randomized noninferiority clinical trial. *Behavior Therapy, 46*, 328–341.

Ögren, M.-L. E. C. (2009). Group supervision in psychotherapy. Main findings from a Swedish research project on psychotherapy supervision in a group format. *British Journal of Guidance & Counselling, 37*(2), 129–139. doi:10.1080/03069880902728614.

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. doi:10.1016/j.cpr.2010.03.004.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173.

Ruzek, J. I., & Rosen, R. C. (2009). Disseminating evidence-based treatments for PTSD in organizational settings: A high priority focus area. *Behaviour Research and Therapy, 47*(11), 980–989. doi:10.1016/j.brat.2009.07.008.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279–313. doi:10.1037/a0014268.

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*. doi:10.1007/s10488-010-0321-0.

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67*(6), 837.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 73*(1), 106–115. doi:10.1037/0022-006X.73.1.106.

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., et al. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: A preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology, 78*(5), 751–756. doi:10.1037/a0020569.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. London: Oxford University Press.

Stirman, S. W., Bhar, S. S., Spokas, M., Brown, G. K., Creed, T. A., Perivoliotis, D., et al. (2010). Training and consultation in

evidence-based psychosocial treatments in public mental health settings: The ACCESS model. *Professional Psychology: Research and Practice, 41*(1), 48–56. doi:10.1037/a0018099.

Stirman, S. W., Buchhofer, R., McLaulin, J. B., Evans, A. C., & Beck, A. T. (2009). The Beck initiative: A partnership to implement cognitive therapy in a community behavioral health system. *Psychiatric Services, 60*, 1302–1310. doi:10.1176/appi.ps.60.10.1302.

Stirman, S. W., Crits-Christoph, P., & DeRubeis, R. J. (2004). Achieving successful dissemination of empirically supported psychotherapies: A synthesis of dissemination theory. *Clinical Psychology: Science and Practice, 11*(4), 343–359. doi:10.1093/clipsy/bph091.

Stirman, S. W., Gutierrez-Colina, A., Toder, K., Castro, F., Esposito, G., Barg, F., et al. (2013). Clinicians' perspectives on cognitive therapy in community mental health settings: Implications for training and implementation. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(4), 274–285. doi:10.1007/s10488-012-0418-8.

Stirman, S. W., Miller, C. J., Toder, K., Calloway, A., Crits-Christoph, P., & Beck, A. T. (2012). Perspectives on cognitive therapy training within community mental health settings: Implications for clinician satisfaction and skill development. *Depression Research and Treatment, 2012*, 1–11. doi:10.1155/2012/391084.

Swales, M. A., Taylor, B., & Hibbs, R. A. (2012). Implementing dialectical behaviour therapy: Programme survival in routine healthcare settings. *Journal of Mental Health, 21*(6), 548–555.

Tracey, T. J., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69*(3), 218.

Weissman, M. M., Verdeli, H., Gameroff, M. J., Bledsoe, S. E., Betts, K., Mufson, L., et al. (2006). National survey of psychotherapy training in psychiatry, psychology, and social work. *Archives of General Psychiatry, 63*(8), 925–934.

Young, J., & Beck, A. T. (1980). *Cognitive therapy scale: Rating manual*. Unpublished manuscript.