TRAINING OBSTETRICS NURSES AND BEHAVIORAL HEALTH CARE PROVIDERS IN BEHAVIORAL ACTIVATION FOR ANTENATAL DEPRESSION

SAMUEL HENRY HUBLEY

B.A., Cornell University, 2005

M.A., University of Colorado Boulder, 2009

A dissertation submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2013

This thesis entitled:

Training Obstetrics Nurses and Behavioral Health Care Providers in Behavioral Activation for Antenatal Depression

written by Samuel Henry Hubley

has been approved for the Department of Psychology and Neuroscience

_____

Sona Dimidjian

_____

Charles Judd

Date_____

The final copy of this thesis has been examined by the signatories, and we fnd that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol # 11-0544

Hubley, Samuel Henry (Ph.D., Psychology and Neuroscience)

Training Obstetrics Nurses and Behavioral Health Care Providers in Behavioral Activation for Antenatal Depression

Dissertation directed by Associate Professor Sona Dimidjian

Abstract

Objective: A large gap exists between need for mental health treatment and access to evidence-based psychological treatments (EBPTs) in clinical practice settings where patients receive care. Developing EBPT training paradigms represents a promising approach to address this problem. This study examined training outcomes amongst a group of non-specialist and behavioral health providers who learned how to deliver behavioral activation (BA) in obstetrics and gynecology settings for the treatment of antenatal depression (AD). ***Method:*** Participants were non-specialist clinicians and behavioral health providers (n = 10 and pregnant women (n = 85 who screened positive for depression during routine prenatal care visits at 4 large health maintenance organizations in Colorado, Georgia, Minnesota and Washington. Clinician training consisted of a 2-day workshop and weekly group supervision. Training outcomes were measured at baseline, post-training, and follow-up using the Therapist Training Evaluation Outcomes Framework (Decker, Jameson, and Naugle, 2011), focusing on satisfaction with training, attitudes towards EBPTs, knowledge of BA, and BA skill acquisition. The relationship between BA skill and patients' self-reported depressive symptoms during the course of BA treatment was also assessed. ***Results:*** Results indicated that clinicians were satisfied with their training experience and significantly improved over the training period on measures of BA knowledge and skill acquisition;

however, they evidenced decreased favorable attitudes towards EBPTs. In contrast to predictions, BA skill failed to significantly predict changes in patients' self-reported depressive symptoms. ***Conclusions:*** Training non-specialists and behavioral health providers may be a feasible and effective approach to increasing access to EBPTs in routine clinical settings; however, the extent to which such training predicts patient improvement requires additional research. In addition, future work is necessary to draw causal inferences about the impact of training on knowledge and skill in implementing EBPTs.

CONTENTS

TABLES

Table

FIGURES

Figure

Introduction

Evidence-based psychological treatments (EBPTs) exist for a wide range of mental health problems. Despite considerable progress in designing, testing, and refining EBPTs, this knowledge has not widely influenced clinical practice settings where patients receive care. As a result, a large gap exists between patient need and provision of EBPTs and highlights the importance of developing effective service delivery models. To address this problem, multiple approaches have been proposed including deliver EBPTs in routine clinical practice by training non-specialists and behavioral health nascent to these settings. Interest in this approach stems from the emerging literature on training mental health specialists in EBPTs for a variety of mental health problems. Pregnant women who experience depression represent a compelling population to test the viability of training non-specialists and behavioral health providers as a means of increasing access to EBPTs given that women have ongoing contact with the medical system during pregnancy and tend to prefer psychological intervention over medication for depression treatment (Goodman, 2009).

This study examined training outcomes among a group of non-specialist and behavioral health providers who were trained in delivering behavioral activation (BA) in obstetrics and gynecology (OB/GYN) settings for the treatment of antenatal depression (AD). Embedded in a larger, multi-site effectiveness trial of BA for AD, this sub-study assessed clinician training outcomes, including satisfaction with training, attitudes towards EBPTs, knowledge acquisition, skill in implementing BA. In addition to assessing within-clinician changes over the course of the training period, we also evaluated relationships between training outcomes and patient outcomes. This study adds

to the growing EBPT training literature that seeks to define important training outcomes, outline appropriate measurement approaches, and investigate the efficacy of different approaches to training clinicians in EBPTs.

The Treatment Gap in Mental Health Care

Empirically supported treatments exist for a wide range of mental health problems (Chambless & Hollon, 1998; Chambless & Ollendick, 2001), including depression (e.g., Cuijpers, van Straten, Andersson & van Oppen, 2008), anxiety (e.g., Hoffman & Smits, 2008), and substance abuse (e.g., Lundahl, Kunz, Brownell, Tollefson, & Burke, 2010). Unfortunately, the accumulation of data supporting the efficacy of such EBPTs has not widely influenced patterns of care. The slow transfer of EBPTs to routine care settings has been the subject of many conceptual reviews (e.g., Baker, McFall, & Shoham, 2008; Westen, Novotny, & Thompson-Brenner, 2004) and empirical research.

Early work on the research-practice gap, for instance, suggested that many clinicians ranked consultations with colleagues (Cohen, Sargent, & Sechrest, 1986) and personal experience (Morrow-Bradley & Elliot, 1986) as more important than treatment research in improving care quality. Such early findings are consistent with more recent studies as well. Stewart and Chambless (2007) reported results of a study with 508 APA-members practicing clinicians. The respondents answered questions about their approach to treatment decision-making and how they react to treatment research in making treatment decisions. On a 7-point scale (1 = *Strongly Agree* and 7 = *Strongly Disagree*) that assessed extent of agreement with the view that treatment research influences their practice, respondents' mean scores were 3.09 (*SD* = 1.56). In terms of treatment decision-making, scores using the same 7-point scale indicated that past clinical

experiences were most important (M= 1.53, SD = 0.91), followed by colleagues' advice (M= 2.70, SD = 1.28), treatment research (M= 2.86, SD = 1.91), and experiences in personal therapy (M= 2.90, SD = 1.28). A post-hoc comparison indicated that past clinical experience was significantly more important than treatment research, discussions with colleagues and experiences in personal therapy in treatment decision-making (all p's $< .001$, Cohen's d's ≥ 0.76). Such results indicate that EBPT knowledge has yet to widely influence practice patterns among a large sample of APA-member clinicians.

Pignotti (2009) also reported a similar pattern of results in a survey of 400 clinical social workers, using a 7-point scale to rate the importance of different sources of information in making treatment decisions, with lower scores representing less endorsement. Among 13 options, respondents endorsed clinical experience as most important (M= 6.50, SD = 0.88). Several other reasons received high ratings including compatibility with theoretical orientation (M= 5.65, SD = 1.38), compatibility with personality (M= 5.63, SD = 1.38), and endorsement by a respected professional (M= 5.01, SD = 1.41). Favorable research in peer-reviewed journals was rated 9th out of the 13 possible options (M= 4.74, SD = 1.54). These data converge with previous findings indicating that providers report relying more on clinical experience than research when making treatment decisions.

The limited penetration of the EBPT knowledge base into clinical settings is at strong odds with the need for treatment. Such incongruity is aptly illustrated in the domain of depression. Both randomized controlled trials (e.g., DeRubeis et al., 2005; Dimidjian et al., 2006; McBride et al., 2006) and meta-analytic reviews (Cuijpers et al., 2008) suggest that efficacious EBPTs for depression exist. In particular, studies support

the efficacy of cognitive-behavioral therapy (CBT), behavioral therapy, and interpersonal therapy (IPT), with data suggesting that these EBPTs achieve efficacy results comparable to antidepressant medications, even for more severely depressed patients (Hollon & Poniah, 2010). On average, approximately two thirds of depressed patients demonstrate acute response to EBPTs (typically defined as 50% or greater reduction in depressive symptoms from baseline to post-treatment), which is comparable to the approximated 50% response rate to antidepressant medication; and among patients who respond, those who received an EBPT are half as likely to experience a relapse or recurrence of depression in the long-term compared to those who responded to pharmacotherapy and subsequently discontinued their antidepressant medication (Hollon, 2012).

The evidence suggests, however, that such EBPTs are not widely available to depressed individuals. For example, among 1,636 participants from a national sample who met criteria for a depressive or anxiety disorder according to the Composite International Diagnostic Interview (Kessler et al., 1998), less than 45% received appropriate care when defined as at least 4 visits to a mental health specialist (Young, 2001). Using a similar definition of adequate treatment (at least 4 visits with a provider prescribing a psychotropic medication or 8 visits with a provider delivering psychotherapy), an independent epidemiological survey with a nationally representative sample of people with depression reported that 64% received adequate care when seen in specialty mental health settings compared to only 41% when seen in general medical settings (Kessler et al., 2003). Moreover, between 1987 and 1997, the use of psychotherapy for depression declined from 71.1% to 60.2% among a national sample of over 65, 00 respondents, whereas the use of antidepressants for depression showed a

marked increase from 37.3% to 74.5% (Olfson, Marcus, Druss, Elinson, Tanielian, & Pincus, 2002). Such reports likely underestimate the percentage of depressed individuals receiving EBPTs given the reliance on a definition of minimal care (i.e., as few as four sessions) and the lack of any attention to the nature of the interventions delivered during such sessions.

Data on care patterns are at odds with findings from studies examining treatment preferences among depressed patients. Among a national sample of over 1, 500 primary care patients with depressive symptoms, 67% endorsed a preference for counseling over antidepressant medication (Dwight-Johnson, Sherbourne, Liao, & Wells, 2000). Taken together, EBPTs are efficacious and available, yet data indicate that on average, many depressed people do not receive adequate care, and the care they do receive appears misaligned with their treatment preferences.

Data on patterns of care among pregnant women suggest that they are no exception to these general trends. Although emerging data for EBPTs for AD are promising (Sockol, Epperson, & Barber, 2011), such treatments are rarely available in OB/GYN clinics. In a sample of over 3,000 predominately white and educated pregnant women who screened positive for depression symptoms during routine prenatal care visits at 10 obstetrics clinics, less than 15% received any form of treatment (Marcus, Flynn, Blow, & Barry, 2003). Among a sample of 375 pregnant women who met diagnostic criteria for major depression, only one third received a formal diagnosis and less than half received mental health treatment (Ko, Farr, Dietz, & Robbins, 2012). Again, the availability of specific EBPTs is likely far lower in these settings, given the lack of any attention to the nature of psychological interventions delivered. Moreover,

mental health care providers do not typically practice in specialty medical settings such as OB/GYN clinics, where most pregnant women receive their antenatal care and where they also prefer to receive mental health treatment (Flynn et al., 2009; Goodman, 2009).

Moreover, multiple studies have examined treatment preferences among pregnant women. These studies consistently demonstrate that a higher proportion of women rate psychotherapy as an acceptable form of treatment (range varies 58-92%) compared to the proportion of women who rate pharmacotherapy as an acceptable form of treatment (range = 23-35%; Alvidrez & Azocar, 1999; Flynn, Henshaw, O'Mahen, & Forman, 2009; Goodman, 2009; O'Mahen & Flynn, 2008; Sleath, Rubin, & Huston, 2003). For instance, among a sample of 509 pregnant women, only 35% of pregnant women in the 3rd trimester indicated they would likely accept a recommendation for pharmacotherapy if they became depressed, in contrast to the majority (92%) who indicated a strong preference for individual psychotherapy and a preference for receiving that care in their obstetrics clinic (Goodman, 2009). The findings from these studies converge on a clear indication of preference for psychotherapy as the treatment of choice and to receive this care in OB/GYN settings.

The Need for Effective Service Delivery Models: Delivering Care in Routine Settings and Maximizing Use of Non-Specialist Providers

The treatment gap highlights the vital importance of developing effective service delivery models. Flagship organizations in the United States have echoed this imperative as both the Surgeon General's Report on Mental Health (U.S. Department of Health and Human Services, 1999) and the American Psychological Association (2005) have specified increasing access to EBPTs as a priority topic for the field. To address this

problem, Kazdin and Blase (2011) proposed a portfolio of flexible and diverse delivery models, which included recommendations to leverage technologies, the media, and self-help resources to increase access to mental health care. Additionally, recommendations include increasing access to EBPT in which most patients seek care and providing EBPT training for non-specialists, defined as healthcare providers, such as nurses, midwives, health educators, or healthcare aides, who have not received specialized mental health training in the delivery of psychological interventions as part of their educational background.

There are several reasons to consider the value of delivery EBPT in routine clinical settings, as opposed to the traditional structure of delivery only within speciality mental health settings. Not only do providers in routine clinical settings already have existing relationships with patients, increasing access to EBPTs in these settings may also reduce the barrier of stigma associated with seeking mental health specialty care (World Health Organization, 2007). Furthermore, many patients who receive services from primary and secondary care settings have comorbid mental health illnesses and the presence of comorbid mental health can worsen the presentation and course of both acute and chronic illness (Katon, Lin, & Kroenke, 2007). Indeed several initiatives highlight the importance of developing integrated care models that address patients' medical and mental health needs in routine clinical settings (e.g., Agency for Healthcare Research and Quality, 2012; National Institute of Mental Health, 2012).

Moreover, there are several reasons to consider training non-specialists in the delivery of EBPTs. There exists a large discrepancy between the number of diagnosable cases of mental illness each year and the number of practicing specialty mental health

providers, which is estimated as 75 million to 700,000 (Hoge et al., 2007; Kessler & Wang, 2008). The labor force of non-specialists provides a sizable pool of workers who could assist in reducing this discrepancy. Moreover, non-specialists have extensive patient contact in medical settings where the majority of healthcare is provided.

Early examinations of delivery in routine settings and non-specialist care delivery focused on heterogeneous interventions for a variety of target populations and reported promising results (Christensen & Jacobson, 1994; den Boer, Wiersma, Russo, & Bosch, 2005; Durlak 1979; Montgomery, Kunik, Wilson, Stanley, & Weiss, 2010); however, the bulk of empirical work has examined primary care settings that engage non-specialists within collaborative care models that position them in the role of case managers. In this capacity, non-specialists coordinate care by facilitating referrals to specialty mental health, problem-solving logistical barriers, and monitoring treatment progress (Katon, 1995). For example, two studies recruited primary care nurses as care managers to assist physicians in depression treatment by calling patients to assess medication compliance, answer questions about side effects, and provide additional referrals if necessary (Simon, VonKorff, Rutter, & Wagner, 2000; Wells et al., 2000). In another study, 15 nurses completed six hours of manualized workshop training and ongoing weekly supervision to deliver telehealth care that consisted of promoting medication adherence, discussing side effects and developing behavioral plans focused on increasing pleasant activities; compared to usual physician care, nurse telehealth care significantly reduced depressive symptoms and improved treatment satisfaction (Hunkeler et al., 2000). Such delivery models, however, may not maximize the opportunity that non-specialists provide.

Although recent meta-analytic and qualitative reviews on training non-specialists in the treatment of anxiety and depression have concluded that non-specialists can achieve outcomes comparable to more experienced clinicians (den Boer et al., 2005; Montgomery et al., 2010), such reports included a small number of studies that varied widely in terms of methodological rigor, training quality, patient severity, and comprehensiveness of training outcomes. In fact, only four studies to date have focused on the training of non-specialists in an EBPT for depression.

In an early study of CBT-based group therapy for teaching depression management skills to older adults, professional mental health workers and laypeople received training, with each group demonstrating comparable gains on both training and patient outcomes (Thompson et al., 1983). In a later training study, Bright and colleagues (1999) compared professional master's level clinicians and non-specialists recruited from community-based self-help organizations. Following a 2-day experiential workshop, pairs of professional clinicians and pairs of non-specialist clinicians provided 10 group therapy sessions for 98 depressed outpatients. Professional and non-specialist clinicians obtained similar scores on observer-rated assessments of compliance with the treatment protocol, and patients treated by both groups of clinicians demonstrated clinically significant reductions in Beck Depression Inventory (BDI; Beck et al., 1979) and Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967) scores. However, compared to the number of patients treated by non-specialist clinicians, significantly more patients treated by professional clinicians were classified at post-treatment as non-depressed (BDI < 10).

The use of non-specialists also has been a focus of interest in the context of global mental health efforts. In an innovative, recent large study targeting postpartum depression in rural Pakistan, Rahman and colleagues (2008) trained community health care workers ("Lady Health Workers") to provide a 16-session CBT model. Training consisted of readings, video demonstrations, didactics, role-plays and discussions over the course of 2 days of workshop training, a one-day refresher workshop 3 months later, and half-day supervision that occurred once per month during study implementation (Rahman, 2007). Depressed, rural Pakistani postpartum women were randomly assigned to CBT with Lady Health Workers or enhanced routine care with Lady Health Workers trained only in preventive maternal and child education and basic health care (Rahman, Malik, Sikander, Roberts, & Creed, 2008). At 6-months postpartum, women receiving the CBT intervention were significantly less likely to meet criteria for major depression compared to women receiving enhanced usual care (23% vs. 53%; *Cluster Adjusted Odds Ratio* = .22; $p < 0.0001$). This statistically significant difference between CBT and enhanced usual care was also evident at 12-month follow-up (27% vs. 59%; *Cluster Adjusted Odds Ratio* = .22; $p < 0.0001$).

Finally, Ekers and colleagues (2011) trained mental health nurses without formal training or experience in conducting psychotherapy in delivering Behavioral Activation (BA; Hopko, Lejuez, Ruggiaro, & Eifert, 2003; Martell, Addis, & Jacobson, 2001). Training consisted of a 5-day workshop that focused on BA theory, application, and individualizing treatment for individual patients using didactics and experiential exercises and ongoing biweekly supervision during study implementation (Ekers, Dawson, & Bailey, 2012). A randomized controlled trial (N=47) compared BA to treatment as usual

among depressed patients recruited from general medical and primary care mental health practices. At post-treatment, BA delivered by non-specialists significantly outperformed treatment as usual. This study is notable given that interest in BA is informed by the perception of BA as a parsimonious and scalable EBPT.

Moreover, Ekers and colleagues (2012) conducted a cost analysis using quality-adjusted life-years (QALYs) as a metric of disease burden. QALYs, which assign a value of 1.0 for one year of perfect health and a value of 0.0 for being dead, estimate the number of years of life added by a given intervention. This analysis revealed a statistically significant difference of 0.20 QALYs in favor of BA over usual care (p = 0.042), which translated to an incremental cost-effectiveness ratio of £5756, or $8,890.14, per QALY. To put this figure in perspective, the convention in the United States for determining the value of a given intervention in terms of cost per QALY is set at $50,000, with lower figures meaning more cost-effectiveness (Grosse, 2011). Thus, training non-specialists in EPBTs may represent not only an efficacious but also a cost-effective approach to providing depression care.

Findings from these studies indicate that in resource-poor settings (e.g., rural Pakistan) and resource-strained settings (e.g., primary care settings in developed nations), non-specialists can be trained to effectively provide EBPTs for the treatment of depression and perinatal depression, specifically. Preliminary efforts aimed at expanding the workforce capable of effectively delivering EBPTs may be a feasible approach to easing the burden of care placed on the limited number of mental health specialists and increasing access to mental health care to the broader community and to specific vulnerable populations such as pregnant women.

Evaluating EBPT Training

Training non-specialists requires comprehensive and scalable methods of evaluating training efforts. To guide the study of training EBPTs for mental health problems, Decker, Jameson, and Naugle (2011) recently proposed the Therapist Training Evaluation Outcomes Framework (TTEOF) based on Kirkpatrick's (1967) classic four-level model for training evaluation. The TTEOF (see table 1) recommends a focus on both short-term outcomes (to be measured immediately after training) and long-term outcomes (to be evaluated after clinicians begin to implement the EBPT). Four key short-term outcomes are proposed: reactions to training, attitudes toward EBPTs, knowledge acquisition, and acquisition of skills to implement quality EBPTs. Two key long-term outcomes are proposed: acquisition of skills to implement quality EBPTs in clinical practice and patient outcomes.

Table 1
Therapist Training Evaluation Outcomes Framework

Level of evaluation	Time Frame	Methods
1 – Reactions to training	Short	Therapist self-report of satisfaction with training
2a – Attitudes	Short	Therapist self-report of attitudes toward EBPTs
2b – Knowledge	Short	Multiple choice knowledge questionnaire Free recall knowledge examination Open-ended quizzes Written assessments to case scenarios
2c – Skills	Short	Observer-rated role-play
3 – Behavior change in practice	Long	Therapist self-reported use of EBPT Therapist self-reported adherence to EBPT protocol Chart reviews of clinicians' notes Client- or caregiver-rated adherence Observer-rated adherence and competence of therapist delivery of EBPT
4 – Client outcomes	Long	Client- or caregiver-report of symptom or functioning Therapist report of patient functional outcomes

Note: EBPT = Evidence-Based Psychological Treatment.

The following sections reviews training studies that focused on EBPTs for depression to elaborate each level of training outcomes and describe the different measurement approaches available to assess these outcomes. For training outcomes that have not been assessed in the context of studies focused on EBPTs for depression, relevant training studies on EBPTs for other mental health problems are discussed.

Level 1 – Reactions to training. The domain of training reactions includes clinician satisfaction with training experiences, paralleling the construct of client satisfaction in the context of treatment studies. The most commonly used method for assessing reactions to training is for trainees to complete a small number of face-valid questions rated on a Likert-scale following their training experience (e.g., Seng, Prinz, & Sanders, 2006; Worrall & Fruzzetti, 2009). To date, no studies have evaluated training reactions in relation to other outcomes such as knowledge and skill acquisition or change in patient outcomes. Future studies are needed to determine the extent to which training reactions are associated with other training as well as patient outcomes.

Level 2a - Attitudes. Attitudes towards EBPTs refer to thoughts and beliefs towards a particular intervention or towards EBPTs in general. Similar to the assessment of training reactions, self-report questionnaires are the most common method for evaluation attitudes toward EBPTs.

No studies have assessed attitudes towards EBPTs for depression specifically, but multiple studies have examined beliefs towards EBPTs generally. Most commonly, the Attitudes Towards Treatment Manuals self-report measure has been used, with results suggesting that many practicing clinicians believe that treatment manuals are constraining and may threaten the therapeutic alliance (Addis & Krasnow, 2001; Addis, Wade, &

Hatgis, 1999. Although widely used, a major limitation of this approach is the omission of other important attitudinal factors. Aarons (2004) addressed this concern by developing a more comprehensive self-report measure called the Evidence-Based Practice Attitudes Scale (EBPAS) that assesses additional constructs such as EBPT appeal, openness to EBPTs, organizational requirements, and divergence from usual care. In addition to possessing strong reliability (Aarons, 2005), the EBPAS also has demonstrated strong predictive validity, with EBPAS scores measured at baseline predicting the extent to which EBPTs were implemented in community settings (Beidas et al., 2012; Henggeler et al., 2008). One study also evaluated changes in attitudes using the EBPAS and an unpublished attitudinal measure before and after training in EBPTs for childhood anxiety, depression, and behavioral disruption (Borntrager et al., 2009). In this study, the authors did not detect a statistically significant improvement in EBPAS item mean scores from baseline to post-training ($M = 2.84$, $SD = 0.41$ versus $M = 2.91$, $SD = 0.45$), but did observe statistically significant improvements in attitudes on the study-specific attitudinal measure. More studies are needed to determine the extent to which attitudes towards EBPTs change over the course of training and are associated with other training and patient outcomes.

Level 2b – Knowledge. Knowledge acquisition, also referred to as declarative knowledge (Simons, Rozek, & Serano, 2013), refers to learning important facts and principles of EBPTs. Multiple-choice tests are the most widely used method for assessing this training outcome and have been employed in several studies (e.g., Dimeff et al., 2009; Gega et al., 2007; Sholomskas et al., 2005; Sholomskas & Carrol, 2006). For example, Hubley and colleagues (accepted) examined gains in BA knowledge as an

outcome of a BA online training study. Results indicated that knowledge scores improved significantly from baseline to post-training (M = .33, SD = .16 versus M = .51, SD = .15), and were maintained at 1-week follow-up (M = .51, SD = .13). Sholomskas and colleagues (2005; 2006) also demonstrated improvements in EBPT quality occurring in tandem with increases knowledge acquisition, and another reported a statistically significant positive relationship between procedural knowledge of motivational interviewing and observer-rated competency assessments (Miller & Mount, 2001). Additional research is needed to assess the relationship between knowledge acquisition and reactions training, attitudes, or patient outcomes.

Level 2c – Skills. Increasing the degree to which clinicians administer EBPTs, and how skillfully they deliver such EBPTs (also referred to as "procedural knowledge" by Simons, Rozek, & Serano, 2013) is the primary objective of many training studies. The terms "adherence" and "competence" are used to describe these aspects of clinician performance, with adherence being defined as "the extent to which a therapist used interventions and approaches prescribed by the treatment manual and avoided the use of intervention procedures proscribed by the manual" and competence being defined as "the level of skill shown by the therapist in delivering the treatment" (Waltz, Addis, Koerner, & Jacobson, 1993, p.620). The distinction between "adherence" and "competence" has been cited as less relevant to routine clinical practice in which the main question of interest pertains to overall treatment quality (Fairburn & Cooper, 2011). As such, Fairburn and Cooper (2011) proposed to retain the lexicon of *adherence* and *competence* for academic treatment research; however, in the context of routine clinical settings, they advocate for adopting the term *"therapy quality,"* which refers to, "The extent to which a

psychological treatment was delivered well enough for it to achieve its expected effects" (pg. 374).

Efforts to assess therapy quality following training have used simulated clinical situations (i.e., structured role-plays with actors) and therapy sessions (i.e., with actual patients). Fairburn and Cooper (2011) argue that to assess therapy quality adequately, a large number of sessions are required to provide sufficient opportunity for trainees to demonstrate the full range of EBPT components if using ratings of actual therapy sessions. This can be especially difficult to accomplish prior to training. It also may pose an ethical quandary in that it may be inappropriate for providers without adequate training to treat depressed patients. For these reasons, the use of role-plays has been suggested as alternative or supplemental approach to assessing training outcomes. With role plays, researchers can exert more control over the type of presenting issues that can selectively index EBPT components of interest, and the ethical problem of under-trained providers treating patients is eliminated. Sholomskas and Carrol (2006) used role plays to assess quality of twelve-step facilitation before and after training, demonstrating that therapy quality scores significantly increased from pre-training (3.55, $SD = 1.45$) to post-training (4.70, $SD = .70$). With respect to evaluating changes in BA quality following training, Puspitassari and colleagues (2013) reported that therapy quality based on role-plays improved significantly over time, $F(2, 14) = 19.12$, $p < .001$, following online training. No studies to date have reported analyses linking EBPT skill in role-plays with other short-term training outcomes, behavior changes in practice, or patient outcomes.

Level 3 – Behavior change in practice. Studies also have evaluated EBPT quality as a long-term training outcome in the context of rating therapy sessions with

actual patients. Several approaches to assessing this outcome include therapist self-reported use of EBPTs, therapist self-reported adherence, chart reviews, client- or caregiver rated adherence, and observer-rated quality assessments. It is widely recognized that observer-rated quality assessments are the most rigorous and considered the gold-standard (Fairburn & Cooper, 2011). One recent CBT for depression training study has used this approach to assess clinician behavior change in actual practice settings. In this study, 12 community mental health center clinicians attended two days of workshop training followed by one year of phone consultations with a CBT expert (Simons et al., 2010). Within-subjects comparisons of CBT competency scores based on ratings of actual therapy sessions before and after training revealed that clinicians significantly improved their ability to competently deliver CBT, $F(2, 28) = 7.74, p = .002$.

It will be important for future studies to assess the relationship of short-term training outcomes (TTEOF Levels 1-3) and quality of EBPTs in routine practice, especially considering that few studies have assessed these relationships with any disorders, and no studies have evaluated these relationships in the context of training studies of EBPTs for depression. In one recent training study of CBT for child anxiety, baseline attitudes measured by the EBPAS (Aarons, 2004) among school mental health providers significantly predicted post-training adherence scores but not competence scores (Beidas et al., 2012). With respect to the relationship between knowledge acquisition and quality, only one study has reported data on this topic. Miller and Mount (2001) showed that procedural knowledge of motivational interviewing (i.e., written responses to prompts from hypothetical patient statements) showed low to modest correlations with observer-rated competency in motivational interviewing (range of *r*

= .10-.38). Finally, one study demonstrated significant associations between observer-rated adherence and competence to motivational interviewing and patient outcomes such as increases patients' in-session motivation change and substance use abstinence (Martino, Ball, Nich, Frankforter, & Carroll, 2008). A better understanding of how attitudes and knowledge relate to quality may help EBPT trainers optimize doses of didactic and experiential instruction during initial and ongoing training.

Level 4 – Patient outcomes. Reactions to training, changes in attitudes toward EBPTs, knowledge acquisition, and skill in delivering quality EBPTs are conceptualized as important precursors to what many consider the overall objectives of training studies—namely, to improve patient outcomes. Patient outcomes are assessed by either self-report or clinician-administered interviews. Characterizing the relationship between training outcomes and patient outcomes can inform EBPT developers and trainers to emphasize particularly potent aspects of an intervention in favor of more inert aspects, and can also lead to general guidelines for establishing knowledge and quality benchmarks that can serve as decision-making aides for appropriately matching clinician skill to wide-ranging patient needs.

Few studies on training examine the relationship between training and patient outcomes. In fact, the study by Simons and colleagues (2010) is the only training study that has assessed patient outcomes, reporting that patients treated by clinicians after CBT training showed greater reductions in self-reported depressive symptoms and statistically significant reductions in self-reported anxiety, as compared to patients treated before training. Miller and Mount (2001) conducted the only study assessing the relationship between training in an EBPT and patient outcomes.

Several studies, however, have reported on the relationship between EBPT quality and patient outcomes outside the context of training studies. A meta-analysis of 17 studies assessing the relationship between quality and outcome for a variety of interventions and patient populations failed to indicate a statistically mean effect size of r = .07 (Webb, DeRubeis, & Barber, 2010). However, when examining studies focused specifically on depression (n = 7), the authors identified a statistically significant mean effect of quality on outcome (r= .28, p < .001). Of the 7 studies included in this sub-analysis, four assessed the CBT quality-patient outcome relationship and are reviewed in further detail below.

In the first published report, Shaw and colleagues' (1999) did not find a statistically significant relationship between quality and post-treatment depression scores as measured by the BDI and HRSD-17. The authors coded 9 sessions from each of 36 patients who completed a course of CBT for clinician quality using the Cognitive Therapy Scale (CTS; Young & Beck, 1988). Using baseline depressive symptom scores as a covariate in a linear regression model, clinician quality was not a significant predictor of post-treatment depressive symptom scores (r = .28, p = .13). Only when controlling for baseline HRSD-17, therapist adherence, and quality in facilitative aspects of CBT (i.e., warmth, rapport), did a statistically significant effect of CTS scores on post-treatment HRSD-17 scores emerge.

In contrast, three more recent examinations of CBT quality and depression outcomes have demonstrated statistically significant positive associations. First, Trepka and colleagues (2004) coded one randomly selected session using the CTS amongst patients who completed 12-20 sessions of CBT. When controlling for baseline BDI

scores, the overall CTS score was associated with post-treatment BDI scores ($r = .47$, $p < .05$) and 4-month follow-up BDI scores ($r = .54$, $p < .05$). Second, Kuyken and Tsivrikos (2009) used a clinic-specific quality measure to examine a global rating of clinicians' quality in relation to post-treatment BDI-II scores. When controlling for baseline BDI-II scores, the overall quality score was significantly associated with post-treatment BDI-II scores ($r = .28$, $p < .05$).

Finally, Strunk and colleagues (2009) assessed the relationship between CBT quality and patients' depressive outcomes in a multisite randomized controlled trial of CBT for depression by rating sessions two through five with the CTS. These ratings of early quality were then entered into a multilevel model to predict both immediate change in depressive symptoms as measured by the BDI-II in sessions 3-6, and long-term change in depressive symptoms as measured by BDI-II scores at post-treatment. Results indicated a statistically significant association between quality and early reduction in self-reported depressive symptoms ($r = .28$, $p < .05$). With respect to post-treatment outcomes, there was a significant relationship between quality and clinician rated depressive symptoms ($r = .33$, $p = .02$), but not self-reported depression symptoms ($r = .24$, $p = .09$),

In sum, there is evidence to support the theory that variability in clinician quality is associated with variability in outcomes for patients with depression (Trepka et al., 2004; Kuyken & Tsivrikos, 2009; Strunk et al., 2009), although findings are equivocal and one study did not observe such a relationship (Shaw, 1999). More studies are needed to characterize the relationship between clinician quality and patient outcomes. This is especially true in the context of increasing access to EBPTs in routine clinical settings and with non-specialist providers. No studies have assessed the relationship between

EBPT quality and patient outcomes outside of RCTs conducted at academic universities, utilizing highly trained specialist mental health care providers (e.g., psychologists, psychiatrists, clinical social workers).

Present Study

This study evaluated the effects of training OB/GYN nurses and behavioral health care providers to treat AD with BA in obstetrics settings. Although the four studies reporting on training outcomes for non-specialists learning to deliver EBPTs for depression (Bright et al., 1999; Ekers et al. 2011; Rahman et al., 2008; Thompson et al., 1986) provide important starting points for increasing access to EBPTs via non-specialists, they did not comprehensively address important training outcomes outlined in the TTEOF (Decker, Jameson, & Naugle, 2011; see Table 2). No studies to date have conducted comprehensive assessments of TTEOF outcomes among samples of non-specialists or of the implementation of EBPTs in routine medical care settings such as OB/GYN clinics.

In this study, OB/GYN and behavioral health care providers received training in BA before providing treatment as part of a multi-site effectiveness study evaluating BA for AD. The effectiveness trial compared BA against augmented usual care for AD at four large health maintenance organizations (Group Health Cooperative, Washington; HealthPartners, Minnesota; Kaiser Permanente, Colorado; Kaiser Permanente, Georgia). This training study used a within-subjects design to evaluate pre- to post-training outcomes according to the TTEOF and the relationship between training and patient outcomes (Decker, Jameson, & Naugle, 2011; see Table 2).

Table 2

Review of Non-specialists Training Studies of EBPTs for Depression Organized by the Therapist Training Evaluation Outcomes Framework

Level of evaluation	Thompson et al. (1983)	Bright et al. (1999)	Rahman et al. (2008)	Ekers et al. (2011)	Present Study
1 – Reactions to training	-	-	√	-	√
2a – Attitudes	-	-	-	-	√
2b – Knowledge	√	-	-	-	√
2c – Skills	-	-	-	-	√
3 – Behavior change in practice	√	√	-	√	-
4 – Client outcomes	√	√	√	√	√

Note: EBPT = Evidence-Based Psychological Treatment.

Between-subjects comparison between study clinicians and BA experts also evaluated differences between groups on knowledge acquisition. Specifically, we addressed five primary hypotheses:

(1) Post-training satisfaction will exceed an *a priori* cut-off commensurate with "adequate" satisfaction (Level 1);

(2) Attitudes of study participants toward evidence-based practices will improve significantly from baseline to post-training and will be maintained at 9-month follow-up (Level 2a);

(3) Knowledge of BA will improve significantly from baseline to post-training and will be maintained at 9-month follow-up (Level 2b); additionally, BA knowledge will be significantly lower at baseline and statistically equivalent at post-training and 9-month follow-up compared to BA among BA experts (Level 2b);

(4) BA skills will improve significantly from baseline to 9-month follow-up (Level 2c);

(5) BA skills at post-training and follow-up will predict subsequent reductions in patients' self-reported depressive symptoms (Level 4).

Additionally, given the limited evidence, this study explores associations among short-term training outcomes (Levels 1-3) and descriptive clinical characteristics such number of patients assigned to each clinician, number of sessions completed by each clinician and number of supervision hours completed by each clinician. Finally, exploratory analyses also included clinician satisfaction, attitudes, knowledge and descriptive clinical characteristics as predictors of patient improvement in depressive severity (Level 4).

Methods

Participants

Clinician participants. Clinician participants ($N = 10$) were recruited via nominations by site investigators and self-referral following presentations to clinical managers and staff at participating OB-GYN clinics in Colorado, Georgia, Minnesota, and Washington. Inclusion criteria were: (1) receiving supervisor approval to participate as a study clinician in the BA effectiveness trial; (2) interest in learning in BA for AD; (3) availability to attend in-person workshop training; and (4) willingness to complete study assessments. The only exclusion criteria were prior formal training in BA or extensive experience in providing EBPTs for depression. All eligible clinician participants provided written informed consent and received regular work pay for time devoted to study participation (i.e., no additional compensation was provided).

Patient participants. Patient participants in the present study were pregnant women ($n = 85$), aged 18 years or older, who were receiving prenatal care at one of the

participating OB-GYN clinics listed above. As opposed to recruiting participants from selected samples of women seeking treatment or referred for treatment, patient participants were recruited via a population-based screening protocol targeting women receiving routine antenatal care at each site. To maximize generalizability, entry criteria for the effectiveness trial were minimal, and patient participants were excluded only when they possessed characteristics that would place them at risk if enrolled (e.g., evidence in the medical record of a diagnosis of bipolar disorder or psychosis, or self-report on baseline measures as being at immediate risk for self-harm). In addition, only English speaking patients were enrolled. Those who screened positive for AD by scoring 10 or higher on the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001) and consented to participation were randomly assigned to receive BA (n = 85) or enhanced care as usual (n = 78). Patient assignment to clinicians was based on clinician availability and patient location. Of the 85 patient participants assigned to the BA condition, 64 have had the opportunity to complete study assessments at baseline and at the 5 and 10-week follow-up time points. Six patient participants are missing data at either the 5 or 10-week follow-up time point. The remaining 15 patient participants have not been enrolled in the study long enough to complete the 5 and 10-week follow-up assessments; future analyses will be conducted with the final assessment follow-up at 3 months postpartum.

All patient participants provided informed consent to participate in this study, and all procedures were reviewed and approved by the IRBs at Kaiser Permanente in Colorado and Georgia, Health Partners in Minnesota, and Group Health Cooperative in

Washington. Patient participants received $25 gift cards to grocery stores or other retailers for completing assessments at each of four time points.

Procedures

Treatment. Patient participants received up to 10 sessions of BA (mean number of sessions completed =5.60, *SD* 1.38), delivered by telephone or in person at the obstetric clinic or patient's home. BA, which is a structured EBPT that aims to alleviate depression and prevent future relapse by focusing directly on behavior change, was delivered flexibly based on the core principles and strategies outlined in the primary treatment manuals (e.g., Dimidjian, Martell, Addis, & Herman-Dunn, 2008; Martell, Dimidjian, & Herman-Dunn, 2010). BA consisted of collaboratively developing a shared case conceptualization with patients using the behavioral model of depression. This individually tailored case conceptualization focused on how changes in life context can result in low levels of positive reinforcement and high levels of aversive control, and how behavioral responses of avoidance and withdrawal maintain a negative cycle of depressed mood, decreased activity, and further withdrawal and avoidance. Based on this conceptualization, clinicians guided patient participants to become more active in ways that aimed to increase contact with sources of reward and solve problems. Specific treatment strategies included self-monitoring, structuring and scheduling activities, problem solving, and obtaining social support.

Training. Training consisted of a knowledge acquisition phase and a skill promotion phase. The knowledge acquisition phase consisted of assigning background readings in BA and providing a two-day workshop led by BA experts. Prior to attending the workshop, trainees read a BA book chapter (Dimidjian, Martell, Addis, & Herman-

Dunn, 2007), two empirical articles documenting the acute and long-term efficacy BA (Dimidjian et al., 2006; Dobson et al., 2008), and the BA quality measure—the Quality of Behavioral Activation Scale (QBAS; Dimidjian, Hubley, Martell, Herman-Dunn, & Dobson, unpublished manuscript).

Workshop training used a combination of didactics, experiential exercises and individualized feedback. The first day of the workshop focused on orienting participants to study details and developing a foundational knowledge of BA strategies. Orientation to the study involved background information on AD, a review of BA efficacy data, and the rationale for treating AD with BA. The majority of the first workshop focused on BA-specific training organized around developing a BA case conceptualization and the three primary components of BA: structural and stylistic strategies, assessment strategies, and activation strategies. The first day concluded with instructions on how to structure and plan for individual therapy sessions and the entire course of BA treatment.

The second day of the workshop focused on applying BA with pregnant women, using the BA case conceptualization to guide BA strategy selection, and responding to comorbidity and emergent clinical situations (i.e., risk of child abuse, domestic violence, and suicidality). The remainder of the workshop focused on fostering trainees' skill in using the BA case conceptualization to guide selection of intervention strategies and role-playing how to implement the selected strategies.

The workshop training concluded with an open-ended question-and-answer session addressed general questions about BA and its application and study-specific logistics. This knowledge acquisition phase required 20 hours of training (approximately 16 hours of workshop training plus approximately 4 hours of background reading).

The quality promotion phase of training, for the current study analyses, consisted of 9 months of additional weekly group supervision delivered via teleconference with the BA trainers and all clinicians from each site (approximately 90 minutes per week). Supervision focused on additional didactics driven by clinician participants' needs, treatment planning, case review, and practice via role-plays. In addition, clinicians received periodic feedback on digitally recorded sessions from the study supervisors.

Measures.

Demographics. The Clinician Demographics Questionnaire is an 8-item measure developed for this study that assesses clinician participants' age, gender, ethnicity, race, education, job description, experience, and clinical setting. The Patient Demographics Questionnaire is a 7-item measure developed for this study that assesses patient participants' age, ethnicity, race, marital status, education, and income.

Level 1 – Reactions to training. The Clinician Satisfaction with Training and Supervision (CSTS) measure contains 11 items that examine satisfaction with training and supervision. The CSTS is based on the Client Satisfaction Questionnaire (Attkinson & Zwick, 1982), a measure of client satisfaction with health care services that possesses excellent psychometric properties. The CSTS contains 8 questions rated on a four-point Likert-scale and 3 open-ended questions, with higher scores indicating higher satisfaction. An example of a Likert-scale rated question is, "If this type of training were easily available to you and your colleagues, how likely are you to recommend it to a colleague?" An example of an open-ended question is, "What could we add to the training to make it more helpful?" Open-ended questions provided additional qualitative information for evaluations of clinician participants' satisfaction with and experience of training.

Psychometric data for the CSTS are limited to internal consistency statistics derived from two previous training studies that indicated adequate internal consistency (Cronbach's α = .86-.95 in Dimeff et al., 2009 and Cronbach's α = .89-.94 in Hubley et al., accepted). Internal consistency estimates in the present analysis were similarly high (Cronbach's α = .94 for both time points).

Level 2a– Attitudes. The Evidence-Based Practice Attitude Scale (EBPAS; Aarons, 2004) is a psychometrically sound 15-item questionnaire that assesses attitudes towards evidence-based practices. The EBPAS uses a 5-point Likert scale in which higher ratings indicate more favorable attitudes. An example of an EBPAS question is: "How willing are you to try new types of therapy/interventions even if you have to follow a treatment manual?" The EBPAS possesses sound psychometric properties with adequate internal consistency (Crohnbach's α = .75) and positive correlations with clinician attributes (Aarons, 2005). Internal consistency estimates in the present analysis were similarly high (Cronbach's α = .70–.80 depending on time point).

Level 2b – Knowledge. The BA Knowledge Assessment is a combination of a face-valid, free recall memory test and a 27-item multiple-choice test assessing knowledge of BA concepts. Both have been used in a prior BA training study (Hubley et al., accepted). The BA recall test requires participants to name as many BA concepts and techniques as possible. There is no overall time limit. The total score is the correct number of items identified and no penalty is assessed for incorrect responses. Two independent coders, blind to time point, followed coding instructions and a coding guide that contained every BA-specific term presented in the BA treatment manual (Martell, Dimidjian, & Herman-Dunn, 2010). The maximum score was 97. Each coder provided

an independent rating for each participant's recall test at all assessment time points. Total scores were recorded as the mean of the coders' scores. In a previous BA training study, single measure intraclass correlation coefficients indicated high inter-rater reliability that ranged from 0.95 to 0.99, depending on time point (Hubley et al., accepted). Using Shrout and Fleiss's (1979) fixed effects model, ICCs for the present analysis yielded a similarly high ICC of .93. The 27 multiple-choice test questions were derived from the original BA treatment manual (e.g., Martell, Addis, & Jacobson, 2001). A knowledge score was calculated as a sum of correct responses for all 27 items for each participant.

Level 2c – Skills. The Quality of Behavioral Activation Scale (QBAS; Dimidjian, Hubley, Martell, Herman-Dunn, & Dobson, unpublished measure) is a 14-item instrument that assesses skill in implementing BA. The QBAS was adapted from the Cognitive Therapy Rating Scale by retaining BA relevant items (e.g., "Homework") and modifying or adding additional items to be consistent with the BA approach (e.g., "Uses a behavioral case conceptualization and rationale to guide treatment"). Items are rated using a seven-point Likert-type scale with higher scores indicating more quality. Preliminary psychometric evaluations of the QBAS revealed adequate inter-rater reliability with a .72 intraclass correlation coefficient (Dimidjian et al., unpublished data).

QBAS ratings were made based on clinician participants' performance in standardized role-plays with a hypothetical patient portrayed by trained research assistants. Two research assistants participated in approximately four hours of training to play the part of the hypothetical role-play patients. Training consisted of reading a short article on BA (Hubley & Dimidjian, 2009), reading a one-page set of instructions

orienting them to the acting task, reviewing the same one-page vignettes that the study participants read, and approximately four pages of text for each hypothetical vignette that provided additional background information. The actors were instructed to portray a patient of low to moderate difficulty by limiting examples of difficult patient behavior to no more than one instance per role-play, with difficult patient behavior based on examples from the Vanderbilt Psychotherapy Rating Scale; Hollon et al., 1988). Role-plays were conducted over the telephone, were audio-recorded and took approximately 30 minutes to complete. Prior to each role-play, participants read a one-page vignette that detailed background information, depressive symptoms, depression history, and prior treatment. Instructions indicated that clinician participants were to reflect on how the details of the vignette fit the BA approach and to demonstrate skill application in the domains of assessment and intervention.

Clinician participants completed role-plays at baseline, post-training and 9-month follow-up (one participant was unable to complete the role-play at baseline due to logistical difficulties). Two BA experts, blind to time point, completed all ratings. Both experts (Christopher Martell & Ruth Dunn) co-authored the QBAS, have extensive experience with BA quality coding, served as research therapists on the seminal BA trial conducted in Seattle, Washington (Dimidjian et al., 2006), frequently provide BA training and supervision, and have co-authored core BA texts (e.g., Dimidjian, Martell, Addis, & Herman-Dunn, 2008; Martell, Dimidjian, & Herman, 2010.). Reliability between the two coders for 12 overlapping ratings in these analyses was acceptable with an ICC for the total score of .84.

Level 4 – Patient outcomes. The PHQ-9 (Kroenke, Spitzer, & Williams, 2001) is a 9-item self-report measure of depressive symptoms based on diagnostic criteria for depression. At baseline, 5 weeks post-enrollment, and 10 weeks post-enrollment, patient participants rated each item on a three-point Likert-scale, with higher scores indicating higher levels of depressive symptoms. A total score was derived based on the sum of participant responses for each item (minimum score = 0, maximum score = 27). In the case of PHQ-9 assessments with missing items (n = 5), an item average was computed and substituted for the missing item. The PHQ-9 is a well-validated instrument that corresponds to diagnostic criteria for major depression and possesses excellent psychometric properties (Kroenke, Spitzer, & Williams, 2001).

Study timeline. Clinician participants consented to the study and completed baseline assessments before two days of workshop training. After completing the two-day workshop training and background readings, clinician participants completed the CSTS, EBPAS, BA knowledge assessment, and BA skill assessment at post-training and then again at 9-month follow-up. Patient participants included in this study completed a PHQ-9 (Kroenke, Spitzer, & Williams, 2001) at baseline prior to initiating treatment, at a five-week mid-treatment interval, and at 10 weeks post-baseline. The parent study also includes a 3-month postpartum final follow-up assessment point that is not included in the present analyses.

Data Analyses. To evaluate hypothesis one that training satisfaction self-report scores would exceed an *a priori* cut-off commensurate with "adequate" satisfaction, the mean item scores of the CST were compared to an *a priori* cut-off of 2.0 using a one-sample t-test with p set at .05.

To evaluate hypotheses two through four that mean EBPAS, BAKA, and QBAS scores would improve over the course of training, repeated measures analyses of variance (ANOVA) were used with time as the dependent variable. Mauchly's Test of Sphericity was used to ensure the assumption of equal variances. Greenhouse-Geisser corrections were used if this assumption was violated. When the time variable in the omnibus repeated measures ANOVA was statistically significant with p set at .05, follow-up analyses entailed two single degree of freedom contrasts using paired sample t-tests. First, baseline scores were compared against post-training to determine whether or not scores improved significantly after training. Second, baseline scores were compared against 9-month follow-up scores to evaluate whether or not such training gains were maintained over the follow-up period. To evaluate knowledge gains compared to BA experts, independent samples t-tests were used to assess differences between study participants' and the BA experts' mean BAKA scores at each time point.

Hierarchical linear modeling (HLM) using MIXED models in SPSS version 21.0 was used to evaluate hypothesis five that short-term training outcomes will predict patients' change in PHQ-9 scores from baseline to the 5-week and 10-week assessments. HLM permits analyzing longitudinal data when repeated measures cluster within individuals and within clinicians; the inclusion of all patients, even those with incomplete data; and the specification of the best fitting covariance structure (Raudenbush & Bryk, 2002). For ease of presentation, post-training QBAS scores are used as the predictor variable in the models below.

For each patient, the following level 1 model estimates the mean depressive symptom level and the effect of linear time on depressive symptoms:

$$Y_{ijk} = ß_{0ij} + ß_{1ij}\text{Time} + e_{ijk}$$

In this model, Y_{ijk} is the PHQ-9 score for patient i, treated by clinician j, at time k. Time is linear time centered around each individual's number of completed assessments so that the median assessment point equals 0. Accordingly, $ß_{0ij}$ is the mean PHQ-9 score for the i^{th} patient (intercept) and $ß_{1ij}$ is the mean change in the PHQ-9 score for the i^{th} patient per one additional assessment interval (slope).

At level two of the multilevel analysis, level-one coefficients (intercept and slope) across all patients are modeled as a function of within-clinician differences across patients:

$$ß_{0ij} = \gamma_{00j} + u_{0ij}$$

$$ß_{1ij} = \gamma_{10j} + u_{1ij}$$

In these models, γ_{00j} is the mean PHQ-9 score across all patients for the j^{th} therapist and γ_{10j} is the mean change in PHQ-9 score across all patients for the j^{th} therapist per one additional assessment interval.

At level three of the multilevel analysis, level two coefficients are modeled as a function of between-clinician differences in post-training QBAS scores centered around the grand mean post-training QBAS score (Q_j):

$$\gamma_{00j} = \lambda_{00} + \lambda_{01}Q_j + \varepsilon_{0j}$$

$$\gamma_{10j} = \lambda_{10} + \lambda_{11}Q_j + \varepsilon_{1j}$$

The first of these models the mean PHQ-9 score of all patients treated by a given clinician as a function of that clinician's post-training QBAS score. Accordingly, given centering of Q_j, λ_{00} is the grand mean of all PHQ scores averaging across both patients within clinicians and clinicians. λ_{01} estimates the difference in the clinicians' mean

patient PHQ-9 scores as Q_j increases by one unit. In the second model, the dependent variable is the mean change over time in patients' PHQ-9 scores for a given clinician. Accordingly λ_{10} is the mean across all therapists in PHQ-9 change over time for the average patient (given centering of Q_j) and λ_{11} is the difference in the mean PHQ-9 change over time as a function of the therapist Q_j score.

Thus the expanded model reads:

$$Y_{ijk} = (\lambda_{00} + \varepsilon_{0j}) + (\lambda_{10} + \varepsilon_{1j})*\text{Time} + \lambda_{01}Q_j + \lambda_{11}Q_j*\text{Time} + e_{ijk}$$

In this expanded model, the primary coefficients of interest are λ_{10} which estimates the main effect of time in average PHQ-9 score changes across all clinicians, λ_{01}, which estimates the main effect of post-training QBAS scores on mean change in PHQ-9 scores, and $\lambda_{11,}$ which estimates whether the change in PHQ-9 scores over time depends on clinicians' post-training QBAS scores. All patient participant sessions were included in analyses using post-training outcomes as predictor variables whereas only patient participant sessions occurring after the 9-month follow-up training assessment were included in analyses using 9-month follow-up training outcomes as predictor variables.

For exploratory analyses addressing potential relationships among training variables, Pearson correlation coefficients were computed to assess associations between descriptive clinical data obtained during the post-training and 9-month follow-up interval and short-term training outcomes, with *p* set at .05. To explore the prediction of patient outcomes by clinician satisfaction, attitudes, and knowledge, these variables were entered into the HLM model in place of QBAS scores as the predictor variable.

Power analysis. To detect within-clinician changes in training outcomes across time points among a sample of 8 clinicians, and using $\alpha = .05$, this study has 36% power

to detect a medium effect size (Cohen's *d* =.50) or 65% power to detect a large effect size (Cohen's *d* =.80) in mean differences between training outcomes at different time points. To detect a significant relationship between BA skill and change in patient depressive symptoms with a sample of 64 cases and using α = .05, this study has 99% power to detect medium effect sizes (Cohen's *d* =.30) or higher.

Covariates. Several variables were considered for use as covariate independent variables including site, clinician education, clinician experience, number of patients each clinician treated, and patient comorbidity. However, given the small sample size of this study, these variables were not included as covariates.

Results

Participants.

Ten clinician participants initially enrolled in this study but two withdrew early and their data are not included. Consistent with the over-representation of women in the nursing and obstetrics professions (United States Department of Labor, 2003), the 8 remaining clinician participants were all women, had a mean age of 54.71 (*SD* = 5.59), and were primarily Caucasian (75%). All clinician participants endorsed having earned a master's degree, four out of eight (50.0%) reported having a nursing degree, three out of eight (37.5%) reported having an advanced degree in behavioral health and one (12.5%) reported a master's degree in occupational therapy. Clinician participants reported having been a provider for 19.71 (*SD* = 9.71) years. Five clinicians (62.5%) endorsed having had direct experience working with patients with depression but only three reported having ever formally delivered psychotherapy. All clinician participants denied prior training in BA. Clinicians began seeing their first patient 21.25 weeks (*SD* = 7.65)

after workshop training. On average, clinicians conducted a mean number of 28.75 sessions (*SD* = 29.01) for a mean number of 5.00 patients (*SD* = 4.63), and received 38.81 hours of supervision (*SD* = 5.69) between the post-training and 9-month follow-up assessment. Patient participants' demographics are summarized in Table 3.

Table 3
Patient Demographics

Age in years (*SD*)	29.09 (5.80)	
PHQ-9 (*SD*)	14.96 (3.71)	
Characteristics	n	%
Ethnicity:		
Hispanic/Latina	11	12.9
Not Hispanic/Latina	74	87.1
Race:		
Caucasian	44	51.8
African American	25	29.4
Asian American	4	4.7
Native American	4	4.7
Pacific Islander	2	2.4
Other	6	7.1
Marital Status:		
Married	34	40.0
Member of Unmarried Couple	27	31.8
Divorced/Separated	1	1.2
Widowed	1	1.2
Never Married	22	25.9
Education:		
Some High School	4	4.7
High School Graduate or GED	15	17.6
Vocational, Technical Training	4	4.7
Some college	35	41.2
College graduate	19	22.4
Post-graduate	8	9.4
Employment:		
Employed for Wages	54	63.5
Self-employed	6	7.1
Homemaker	6	7.1
Student	7	8.2
Unable to Work/Disability/On-leave	3	3.5
Unable to Find Work	9	10.6

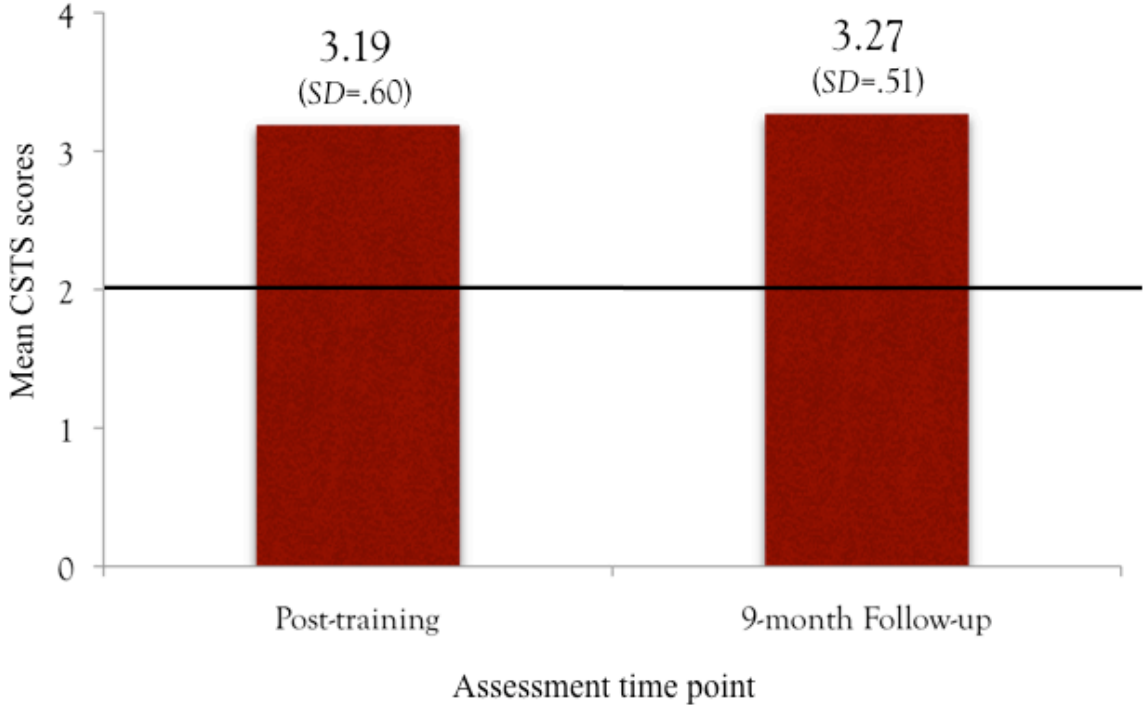
Household Income:		
< $25, 000	29	34.1
$25, 000 - $49, 000	25	29.4
$50, 000 - $74, 000	16	18.8
$75, 000 - $99, 000	9	10.6
$100, 000 - $149, 000	5	5.9
>$150, 000	1	1.2

Note: SD = standard deviation; PHQ-9 = Patient Health Questionnaire 9; GED = General Educational Development.

Clinician reactions to training (Level 1).

To examine reactions with training, mean CSTS scores were compared to an *a priori* cut-off of 2.0, which was commensurate with adequate satisfaction. Post-training and 9-month follow-up scores exceeded the 2.0 cut-off with means of 3.19 (*SD* = 0.60) and 3.27 (*SD* = 0.51), respectively with a score of a "3" indicating "good satisfaction" (see Figure 1).

Figure 1. Mean satisfaction scores.



In their qualitative responses to open-ended questions, clinicians highlighted both positive and constructive reactions to their training experience. Several clinicians expressed appreciation for the multi-method, interactive style of the workshop training, the value of "roundtable" role-plays during group supervision in which clinicians took turns in responding to clinical situations, and the supportive and constructive nature of individual feedback from the trainers. Clinicians also reported ideas to improve both the workshop training and ongoing supervision, including less information in the initial workshop with content spread over a period of time longer than 2 days. A desire for smaller supervision groups also emerged as a common theme at the 9-month follow-up assessment.

Change in Clinician Attitudes (Level 2a).

A repeated measures ANOVA showed statistically significant changes in EBPAS scores over time $F(2, 14) = 4.57$, $p = .030$ (see Table 4). Contrary to prediction, however, mean EBPAS scores did not improve between baseline and post-training,

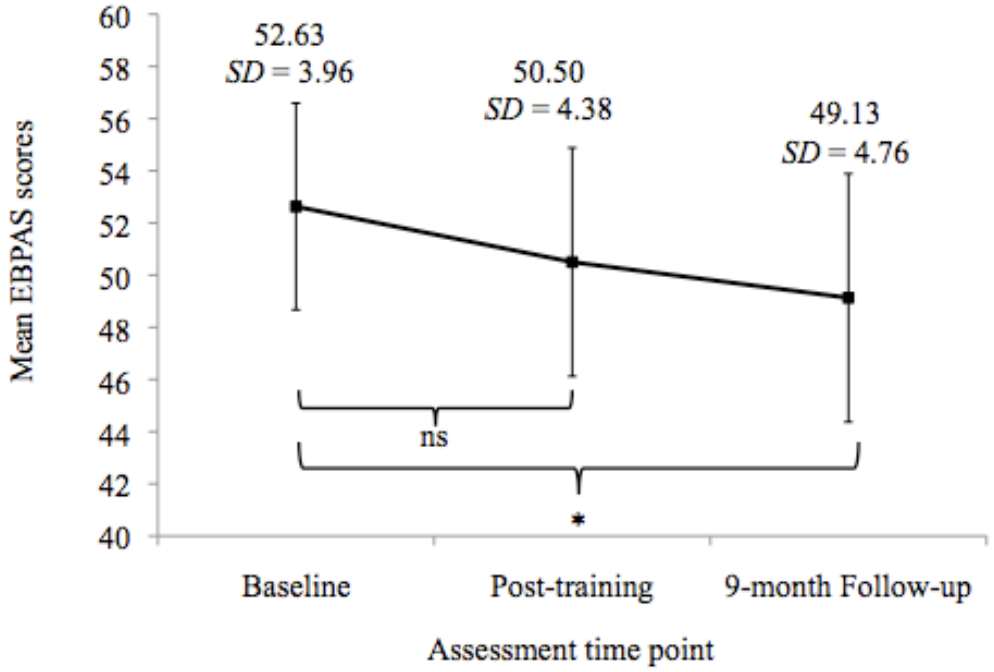
Table 4
Short-Term Training Outcomes

Level of evaluation	Baseline *M* *(SD)*	Post-training *M* *(SD)*	Follow-up *M* *(SD)*	*F*	η^2
1 – Reactions to training	-	3.31 (.60)	3.25 (.51)	-	-
2a – Attitudes	52.63[a] (3.96)	50.50 (4.38)	49.13[a] (4.76)	4.57*	.40
2b – Knowledge, multiple-choice test	12.88[a,b] (4.58)	19.38[a] (2.77)	18.00[b] (4.11)	20.46***	.75
2b – Knowledge, recall test	3.00[a,b] (3.43)	7.94[b,c] (2.44)	11.00[a,c] (4.53)	22.13***	.76
2c – Skills in BA quality based on structured role-play	41.71[a] (10.55)	49.88 (14.68)	53.50[a] (16.26)	4.43*	.42

Note: $N = 8$ for each estimate except Skills in BA Quality measured at baseline ($n = 7$). *M* = Mean scores. *SD* = Standard deviation. BA = Behavioral Activation. Means with the same superscript were significantly different ($p < .05$). η^2 = eta squared for baseline to follow-up comparisons* = $p < .05$, *** = $p < .001$.

$t(1,7) = -1.76$, $p = .122$, Cohen's $d = .51$, and actually declined significantly from baseline to 9-month-follow-up, $t(1,7) = -3.50$, $p = .010$, Cohen's $d = .80$ (see Figure 2). Upon further inspection of changes in mean scores for individual EBPAS items, statistically significant declines from baseline to 9-month follow-up emerged for items 8 and 15 (Item 8 = "I would try a new therapy/intervention even if it were very different from what I am used to doing," and Item 15 = "If you received training in a new therapy or intervention that was new to you, how likely would you be to adopt if you felt you had enough training to use it correctly?").

Figure 2. Evidence-based Practice Attitudes Scale scores

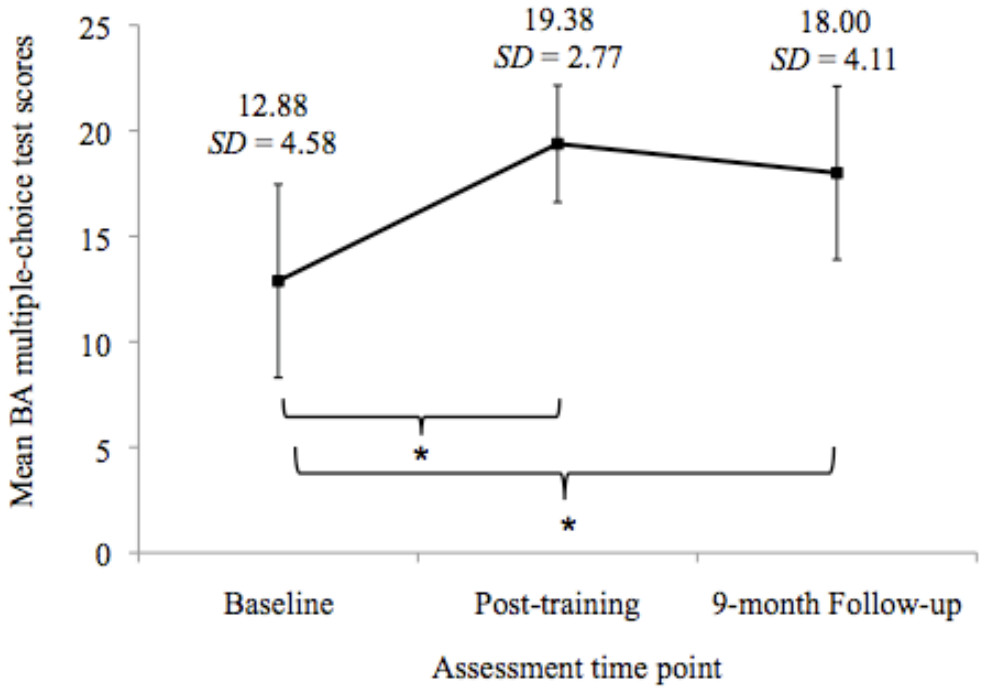


* = $p < .05$. *SD* = standard deviation.

Change in clinician knowledge (Level 2a).

On average, participants correctly answered less than half of the knowledge test items at baseline, indicating a low level of BA specific knowledge (see Table 4). A significant effect for time using repeated measures ANOVA, $F(2, 14) = 20.46$, $p < .001$, indicated that the sample as a whole improved in BA knowledge over the training period. Planned contrasts compared baseline knowledge to knowledge measured at post-training and 9-month follow-up and indicated that clinicians' knowledge improved significantly between baseline and post-training, $t(7) = 5.58$, $p < .001$, with a large within-subject effect size (*Cohen's d* = 1.97; see Figure 3). Participants maintained their knowledge gains, as 9-month follow-up scores remained significantly higher than baseline scores $t(7) = 5.00$, $p = .002$) with a large within-subject effect size (*Cohen's d* = 1.77).

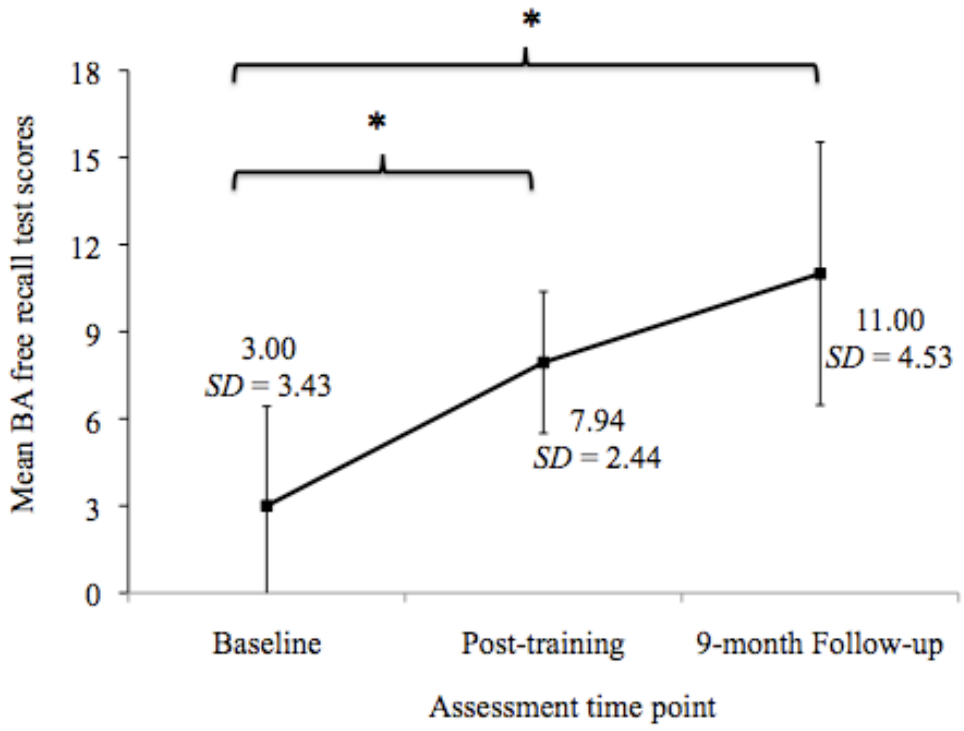
Figure 3. Behavioral Activation multiple-choice test scores



* = $p < .05$. *SD* = standard deviation.

As shown in Table 4, all participants correctly generated, on average, only 3.00 (*SD* = 3.43) BA terms or concepts at the baseline recall test, further demonstrating participants' overall lack of familiarity with BA. A significant effect for time, $F(2, 14) = 22.13$, $p < .001$, indicated that the sample as a whole improved in their ability to generate BA terms and concepts over the training period. Planned contrasts compared recall knowledge at baseline to post-training and 9-month follow-up indicated that clinicians' recall knowledge improved significantly after training, $t(7) = 3.66$, $p = .008$, with a large within-subject effect size (*Cohen's d* = 1.29; see Figure 4). Participants not only maintained their knowledge gains at 9-month follow-up compared to baseline, $t(7) = 6.67$, $p < .001$ with a large within-subject effect size (*Cohen's d* = 2.36), but recall scores also

Figure 4. Behavioral Activation free recall test scores



* = $p < .05$. *SD* = standard deviation.

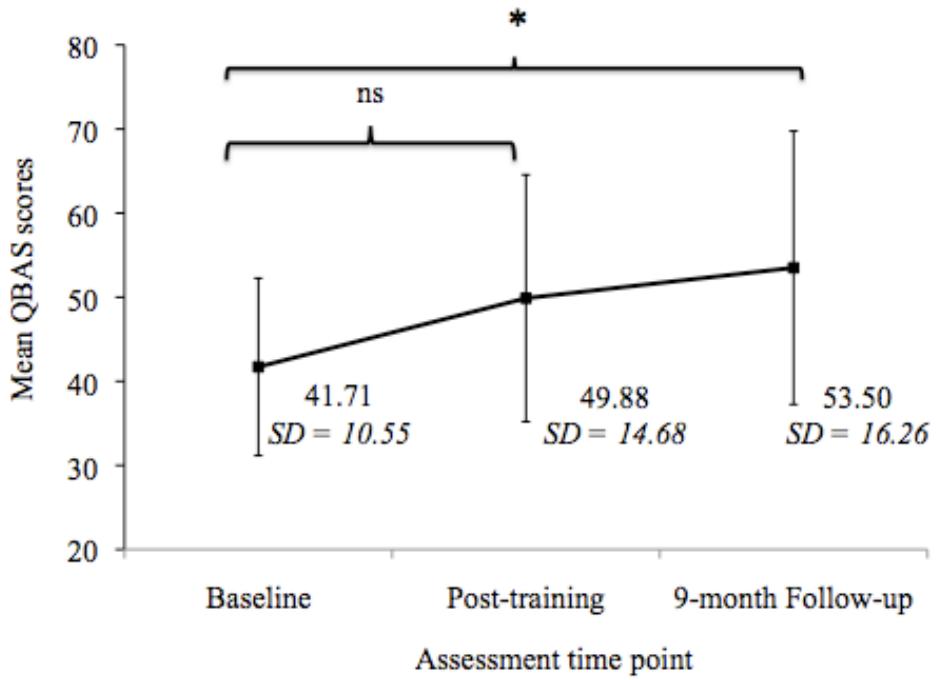
continued to significantly improve from post-training to 9-month follow-up $t(7) = 2.85$, $p = .025$, with a large within-subject effect size (*Cohen's d* = 1.01).

As predicted, independent samples t tests indicated that BA expert knowledge scores (Multiple-choice test $M = 24.50$, $SD = .71$; Recall test $M = 16.75$, *SD = 2.47)* were significantly higher than clinician participants' scores at baseline. Contrary to predictions, BA experts continued to significantly outperform clinician participants based on the multiple choice test at post-training $t(8) = 3.43$, $p = .009$, Cohen's $d = 2.94$, and at 9-month follow-up, $t(7.97) = 4.67$, $p = .003$, Cohen's $d = 2.70$. Similar patterns for knowledge were evident on the BA recall test with BA experts significantly outperforming clinician participants at post-training, $t(8) = 4.56$, $p = .002$, Cohen's $d = 3.60$, and, at the level of a non-significant trend, at 9-month follow-up, $t(3.07) = 2.42$, $p = .092$, Cohen's $d = 1.21$.

Change in clinician skills in delivering BA (Level 2c)

Results indicated that clinician participants significantly improved in the skill with which they delivered BA during the structured role plays, $F(2, 12) = 4.43$, $p < .05$, and items scores at post-training (M=3.56, $SD = 1.05$) and 9-month follow-up (M=3.82, $SD = 1.16$) both exceeded the "satisfactory" threshold. Planned contrasts comparing baseline skill to post-training skill and to the 9-month follow-up indicated a non significant increase in skill between baseline and post-training, $t(6) = 2.31$, $p = .06$ with a large within-subject effect size (*Cohen's d* = .87), and a statistically significant improvement from baseline to 9-month follow-up $t(6) = 2.47$, $p < .05$, with a with a large within-subject effect size (*Cohen's d* = .93; see Figure 5).

Figure 5. Quality of Behavioral Activation Scale (QBAS) scores



* = $p < .05$. *SD* = standard deviation.

Prediction of patient outcome by clinician skill.

A statistically significant change in mean PHQ-9 scores, $\lambda_{10} = -2.80$; $p < .001$, 95% *CI* = (-3.54-(-)2.07), indicated that patient participants' PHQ-9 scores decreased by an average of 3.54 PHQ-9 points per assessment interval (mean PHQ scores at baseline, 5 weeks, and 10 weeks were 14.82 (*SD* = 3.67), 11.43 (*SD* = 4.85), and 8.80 (*SD* = 5.42), respectively. The random error term associated with between-clinician variability in mean change in PHQ-9 scores was not statistically significant, $\varepsilon^2_{1j} = .40$; $p = .69$, 95% *CI* = (-1.53-2.32), suggesting that variability in mean change in PHQ-9 scores could not be explained by between-clinician factors. Indeed, contrary to prediction, there was not a significant main effect of BA skill measured at post-training on change in PHQ-9 scores (*r type effect size* = .04, $p = .72$), indicating that mean change in PHQ-9 scores did not

vary as a function of clinicians' post-treatment BA skill scores. Similarly, QBAS scores measured at follow-up did not significantly predict change in PHQ-9 scores over time (r *type effect size* = .08, p = .46).

Exploratory associations among training outcomes.

Table 5 depicts Pearson correlation coefficients between training outcomes. No training outcomes measured at baseline were significantly associated with other training outcomes measured at post-training or 9-month follow-up. In contrast, several significant associations emerged between training outcomes measured at post-training and training outcomes measured at 9-month follow-up: post-training multiple-choice test scores and attitudes at 9-month follow-up, (r = .74, p < .05), post-training QBAS scores and 9-month follow-up multiple-choice test scores, (r = .79, p = .02) and recall test scores, (r = .86, p < .01), and post-training reactions and 9-month follow-up QBAS scores, (r = .88, p < .01).

We also noted several significant associations between descriptive clinical characteristics and subsequently measured 9-month follow-up outcomes: number of patients assigned to each clinician and 9-month follow-up CSTS scores (r = .79, p < .05), 9-month follow-up EBPAS scores (r = .78, p < .05), and 9-month follow-up multiple-choice test scores (r = .85, p < .01); sessions clinicians conducted and 9-month follow-up CSTS scores (r = .74, p < .05), 9-month follow-up EBPAS scores (r = .78, p < .05), and 9-month follow-up multiple-choice test scores (r = .80, p < .05); and between number of supervision hours received and, 9-month follow-up EBPAS scores (r = -.77, p < .05).

Table 4

Correlation Matrix of Short-Term Training Outcomes

	CSTS PT	CSTS FU	EBP BL	EBP PT	EBP FU	MCT BL	MCT PT	MCT FU
CSTS PT	____							
CSTS FU	-.29	____						
EBP BL	-.24	.41	____					
EBP PT	.12	-.15	.66	____				
EBP FU	-.23	.36	.81*	.70	____			
MCT BL	.01	.62	.23	.16	.56	____		
MCT PT	-.37	.70	.44	.10	.74*	.70	____	
MCT FU	.33	.60	.38	.25	.59	.78*	.72*	____
RT BL	.35	.43	-.03	.04	.34	.66	.56	.70
RT PT	.23	.17	.59	.39	.63	.24	.51	.68
RT FU	.27	.50	.29	.11	.56	.65	.77*	.94**
QB BL	.25	.12	-.25	-.42	-.06	.37	.27	.50
QB PT	.57	.37	.09	-.09	.15	.31	.40	.80*
QB FU	.88**	-.14	-.07	.17	.04	.22	-.01	.61
# of pts.	-.50	.2	.65	.30	.78*	.75*	.85*	.59
# of ses	-.52	.74*	.67	.36	.78*	.73*	.80*	.55
sup hour	.43	-.28	-.46	.63	-.78*	-.59	-.59	-.34

Note: CSTS = Clinician Satisfaction with Training Scale. EBP = Evidence-Based Practice Attitudes Scale. MCT = Multiple-Choice Test. RT = Recall test. QB = Quality of Behavioral Activation Scale. BL = Baseline. PT = Post-Training. FU = 9-month Follow-up. # of pts. = number of patients assigned to clinician. # of ses = number of sessions completed by clinician. sup hour = number of supervision hours received by clinicians. * = $p < .05$. ** = $p < .01$.

Table 4

Correlation Matrix of Short-Term Training Outcomes

	RT BL	RT PT	RT FU	QB BL	QB PT	QB FU	# of pts	# of ses	sup hour
CSTS PT									
CSTS FU									
EBP BL									
EBP PT									
EBP FU									
MCT BL									
MCT PT									
MCT FU									
RT BL	———								
RT PT	.19	———							
RT FU	.70	.78*	———						
QB BL	.08	.52	.63	———					
QB PT	.55	.69	.86**	.62	———				
QB FU	.37	.60	.62	.59	.79*	———			
# of pts.	.36	.29	.48	-.01	.08	-.25	———		
# of ses	.27	.28	.43	-.01	.01	-.27	.99*	———	
sup hour	-.25	-.12	-.23	.23	.25	.25	-.76	-.80	———

Note: CSTS = Clinician Satisfaction with Training Scale. EBP = Evidence-Based Practice Attitudes Scale. MCT = Multiple-Choice Test. RT = Recall test. QB = Quality of Behavioral Activation Scale. BL = Baseline. PT = Post-Training. FU = 9-month Follow-up. # of pts. = number of patients assigned to clinician. # of ses = number of sessions completed by clinician. sup hour = number of supervision hours received by clinicians. * = $p < .05$. ** = $p < .01$.

Exploratory prediction of patient outcomes by clinician satisfaction, attitudes, and knowledge.

HLM analyses revealed no significant effects for number of patients assigned to each clinician (*r type effect size* = .09, *p* = .48), number of sessions conducted by each clinician (*r type effect size* = .01, *p* = .99), number of supervision hours received by each clinician (*r type effect size* = .02 , *p* = .98), reactions to training (*r type effect size* = .01, *p* = .90), attitudes (*r type effect size* = .05, *p* = .72), multiple-choice test scores (*r type effect size* = .05, *p* = .63), or recall test scores (*r type effect size* = .19, *p* = .27) on patient outcomes.

Discussion

In light of the widely acknowledged gap between the need for mental health treatment and availability of EBPTs, national and international organizations such as the National Institute of Mental Health (2013) and the World Health Organization (2013) are prioritizing the importance of urgently shifting attention from efficacy testing to implementing novel service delivery methods for EBPTs. Among many promising approaches, increasing access to EBPTs in medical settings where the majority of people receive their healthcare is especially exciting. Findings from this study demonstrate that training non-specialists and behavioral health providers to provide EBPTs in non-specialist settings is associated with positive reactions and gains in both knowledge and skill acquisition; however, findings also highlight potential areas for improvement in both measurement and training.

From a feasibility perspective, clinician participants reported positive reactions as mean satisfaction scores measured at post-training and 9-month follow-up indicated that clinician participants were satisfied with their training experience. Reactions to training are important because they reflect trainees' subjective experiences during training, which are valuable data points for trainers to consider in the process of comparing and refining different training approaches.

In contrast however, clinician attitudes as measured by the EBPAS (Aarons, 2004) declined over the training period. Upon further inspection, the largest declines occurred among EBPAS items measuring willingness and likelihood to use new interventions. There are several possible explanations for this unexpected finding. First, some clinicians may have felt challenged by stepping outside of their typical roles to treat depressed pregnant women. Given such challenges in the context of already demanding workloads, their willingness to consider learning additional EBPTs for mental health issues may have decreased over time. Second it is important to note that clinicians delivered BA in the context of settings that were not organized to support broadly mental health care provision. It may be necessary to modify organizational culture to promote positive attitudes to EBPTs among individual clinicians. In fact, providers who work in settings with written practice policies for managing mental illness reported higher EBPAS scores compared to providers who worked in settings without written practice policies (Aarons, 2004).

Finally, it is also possible that clinicians without extensive experience in mental health treatment reported more positive attitudes towards EBPTs initially, and that after training and experience implementing an EBPT, their attitude scores at 9-month follow-

up (49.13, *SD* = 4.76) converged towards scores reported by more experienced clinicians. Indeed, the EBPAS was designed for mental health professionals, not for non-specialists and behavioral providers, which may further account for initially higher EBPAS scores at baseline. Future studies can help clarify attitudes non-specialists and behavioral health providers hold towards EBPTs, how those attitudes change over the course of training, and the degree to which attitudes are promoted or constrained by other factors such as organizational support or culture.

The results of this study also demonstrate that non-specialists and behavioral health providers can acquire EBPT knowledge and skill. Clinician participants demonstrated gains in BA knowledge acquisition as measured both by their performance in multiple-choice testing and a free recall task. Consistent with predictions, there was a statistically significant improvement in multiple-choice and free recall scores from baseline to post-training and the difference between follow-up and baseline scores remained statistically significant. An improvement in multiple-choice scores of approximately 50% from baseline to post-training is consistent with previous assessments of knowledge acquisition in training studies of EBPTs (e.g., Dimeff et al., 2009; Hubley et al., accepted). However, findings also underscore the potential limitations of brief training. BA experts outperformed clinician participants on both tests of BA knowledge scores at post-training and at 9-month follow-up. It is possible that even after workshop training, clinical experience, and ongoing supervision, novice BA providers such as the clinician participants in this study may not be able to achieve the same level of knowledge as BA experts. Many have criticized training paradigms reliant on brief workshops without ongoing supervision (e.g., Beidas & Kendall, 2010) and argue for an

overhaul to more optimally modify clinicians' knowledge and patterns of clinical practice so that trainees approximate levels of knowledge and skills evidenced by experts. While it is undoubtedly true that modifications to existing training models have the potential to help non-specialist and behavioral health clinicians more effectively acquire EBPT knowledge and skill, it has also been argued some EBPTs may be too complex for widespread implementation by non-specialists and behavioral providers (Comer & Barlow, 2013). Accordingly, mental health specialists who undergo intensive training both in graduate school and throughout their professional development may be required to deliver more complex EBPTs, especially for low base rate and treatment resistant disorders.

Using the training model of brief workshop training coupled with ongoing supervision, clinician participants in this study significantly improved in their skill acquisition as demonstrated by ability to implement quality BA in the context of standardized role-plays. As predicted, a statistically significant difference from baseline BA skill scores emerged at 9-month follow-up, but not at post-training. The finding that clinicians did not demonstrate significant improvements in BA skill immediately post-training is consistent with Beidas and Kendall's (2010) conclusion that workshop training alone does not typically engender significant improvements in EBPT quality; rather, workshop training plus ongoing clinical contact and supervision may be necessary to improve the quality with which EBPTs are delivered. The present study, however, did not reveal significant associations with amount of clinical contact or supervision with 9-month follow-up BA quality scores. Future studies can shed light on the relative

contributions of ongoing clinical contact versus ongoing supervision to the acquisition of skill in implementing EBPTs in standardized role-plays and actual patient encounters.

Finally, although skill acquisition increased over time, indicated by increased BA skill measured at post-training and follow-up, such gains did not predict patient change in depressive symptoms; similarly exploratory analyses of other training outcomes (satisfaction, attitudes, and knowledge) also failed to predict patient outcomes. Several possible interpretations may account for such null findings. First, it is possible that quality implementation of EBPTs is irrelevant to patient outcomes. In their meta-analyses of the quality-outcome relationship, Webb and colleagues (2011) noted significant heterogeneity in the distribution for the competence-outcome relationships across studies ($Q = 37.15$, $p < .01$) and cautioned against such conclusive interpretations that competence is unrelated to patient outcome based on this mean effect size estimate. Second, it has been widely recognized that the association between competence and outcome is generally examined as a secondary aim in the context of an RCT. As the aim of most RCTs is to demonstrate variability between EBPTs as opposed to variability within EBPTs, the range of variability in competence scores observed in RCTs typically clusters around a competence "ceiling" leaving fewer opportunities to examine the effect of lower levels of quality on outcome (Whisman, 1993). In this study, it appeared that there was sufficient variability in post-training QBAS scores to address this problem (range = 36-69); however, small sample sizes of quality estimates may have precluded finding statistically significant effects of BA quality on patient outcomes.

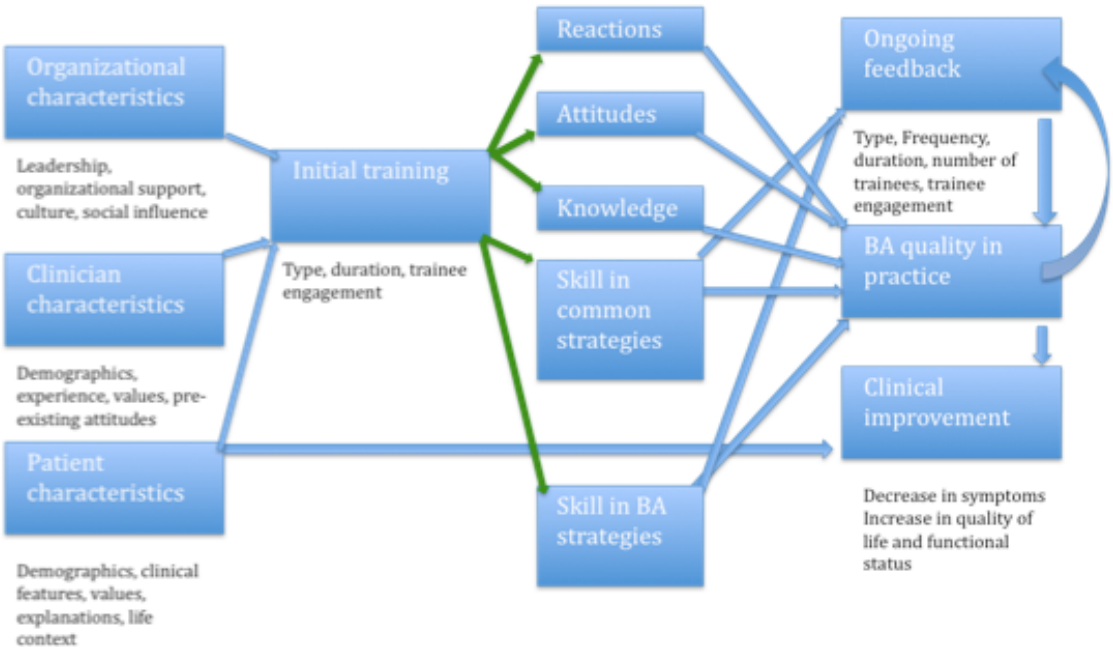
Furthermore, the power calculations indicated that the sample size of 85 patient participants would provide 99% power, in fact, there were not 85 distinct BA quality

scores matched to each case. Rather, all cases for a given clinician used the BA quality score earned by that clinician at the post-training role-play assessments opposed to a quality score earned by that clinician during actual treatment of each patient. This may be problematic for a number of reasons. Using quality scores from standardized role-plays implies that the ability to implement quality EBPTs is a trait-like characteristic that can be accurately captured from one time point to determine overall quality using a single "snapshot." However, other studies have noted considerable variability in within-clinician quality scores across and within patients (Barber, 2006, 2007). Thus, reliance on quality scores derived from standardized role-plays may have provided an inappropriate assessment of overall skill, limited variability of BA quality scores, and limited power as the predictor variable in the quality-outcome analysis. Future analyses would benefit using BA quality assessed during actual treatment sessions with patients.

Overall, results of this study suggest that workshop training paired with ongoing supervision can lead to improvements in short-term training outcomes that can be maintained over an extended follow-up interval. However, the ways in which training outcomes are linked to patient outcomes remains somewhat unclear given that the magnitude of the relationship between short-term training outcomes and clinical improvement was quite low (*r type effect sizes* ranged from .01-.12). It is therefore important to consider what additional factors might influence patient improvement. A comprehensive model including organizational, clinician, and patient level variables is proposed and used to summarize relationships that were supported in the current study and variables that merit consideration in future research (see Figure 6). For example, variability in organizational support, such as provision of physical space and release from

other clinical duties that are required to do deliver EBPTs to pregnant women, may be important determinants of attitudes towards implementing EBPTs in routine clinical

Figure 6. Model of organizational, clinician, and patient characteristics hypothesized to interact with training and patient outcomes.



Note: Green arrows represent relationships supported by this study. Blue arrows represent relationships requiring further study.

settings and, ultimately, improved patient outcomes. Such organizational level variables should be measured and included as potential moderators in future studies.

Clinician characteristics also may play a role to the extent that trainees engage with the initial training and with the learning experience overall. For example, older and more experienced clinicians may have developed idiosyncratic interpersonal and clinical approaches to treating patients to which they are strongly wedded and may have difficulty "shifting gears" to adopt new interventions. And finally, it is important to understand the ways in which patient characteristics may influence not only approaches to structuring training and providing ongoing feedback, but also how the clinical change

process unfolds. For instance, life events such as loss of a relationship or a miscarriage may influence clinical change more powerfully than a clinicians' ability to correctly recall key EBPT terms and concepts. Understanding how different organizational, clinician, and patient characteristics interact to influence clinical change is an important avenue for future research.

Limitations.

There are several important limitations to note. First, small sample sizes may have limited detection of statistically significant differences between knowledge scores of clinician participants and BA experts and statistically significant relationships between training and patient outcomes. However, the large effect sizes observed between clinician participants and BA experts on knowledge assessments (Cohen's *d* ranged from 1.21-3.60) and the small effect sizes observed for the relationship between short-term training and patient outcomes (*r type effect sizes* ranged from .01-.12) suggests that even with larger samples, these would have remained inconsistent with our predictions.

Second, the training component of the study was uncontrolled. Thus, the possibility remains that within subject change on training outcomes is accounted for simply by the passage of time or repeated measurement. For instance, it is possible that clinician participants' multiple-choice test and QBAS scores may have improved due to a "practice effect" given that they answered the same multiple-choice test questions and were asked to implement the same BA strategies at each assessment time point. This is unlikely however given the magnitude of effect sizes (range = .87-2.36) observed in this study and the 9-month interval between the follow-up assessment and the baseline and post-training assessments. Additionally, it is the expectation in training studies for

participants to demonstrate newly acquired knowledge and clinical skill in the same context until a level of proficiency is achieved.

Moreover, the single group design for the training study does not allow us to make interferences about the relative value of different training components or models. With respect to the scalability of the training model used in this study, it is important to note that the clinicians had weekly consultation with expert BA clinicians. Such ongoing consultation may not be feasible in many clinical practice settings, and it would be valuable to examine the necessity of this training component in future studies. Although it may not be ethically feasible to randomize non-specialists and behavioral health providers to training without any follow-up supervision (given that these providers are unlicensed and inexperienced in the treatment of mental health disorders), future studies could compare approaches to training that differ in terms of supervision frequency (i.e., number of hours or regularity of contact) as well as other dimensions of training such as content (i.e., the whole EBPT package versus core components), training approach (i.e., emphasis on passive versus active didactic instruction), or delivery method (i.e., in-person workshops versus online training).

Third, important limitations in measurement should be noted. As discussed previously, this study did not include assessment of long-term quality outcomes in actual sessions (Level 3 of the TTEOF). Including such assessments in future training studies is important not only for the reasons mentioned above (i.e., increasing variability in BA quality as a predictor of patient outcomes), but also to document the extent to which improvements in QBAS scores transfer from standardized role-plays to actual therapy sessions. Similarly, future studies would benefit from more frequent quality assessments

throughout the training period to assess change in clinician skill and the relationship between EBPT quality and patient outcomes.

Fourth, it is unclear to what extent these findings will generalize to larger, more diverse samples of non-specialists and behavioral health providers as clinician recruitment was based on nominations and interest in participating in this study. Additionally, only five of the eight clinicians in this study were non-specialists and although we were interested in comparing training outcomes between the non-specialists and behavioral health providers, the IRB denied requests to report on individual- or between-clinician outcomes. It will be important for future studies to make comparisons between different groups of non-specialists and behavioral health providers.

Conclusions

Findings from this study are compatible with recent calls to develop novel service delivery methods to increase access to EBPTs (Kazdin & Blase, 2011). In conjunction with Ekers and colleagues' (2011) BA training study with mental health nurses, this study demonstrates the feasibility of training non-specialists and behavioral health providers in BA. Tapping such a largely unused source of health care providers to provide EBPTs is a scalable delivery model capable of making significant public health impacts. For example, Rahman and colleagues' (2008) study showed that 40 "Lady Health Workers" trained in a CBT intervention for perinatal depression could treat nearly 500 women through pregnancy to 9 months postpartum and achieve outcomes similar to more highly trained mental health specialists. Such "disruptive innovations" seek to provide concise and cost-effective services that prioritize meeting needs for the majority of a population (Rotheram-Borus, Swendeman & Chorpita, 2012). Training non-specialists to deliver

EBPTs may represent this type of disruptive innovation required to trigger a true paradigm shift in clinical training and practice.

Second, a primary source of interest in BA lies in the perception that BA is a relatively less complex than other EBPTs for depression (e.g., CBT and IPT). Given that the majority of people with depression do not receive adequate care (Kessler et al., 2003), it is important to identify which EBPTs are the easiest to learn and may be most amenable to widespread dissemination and implementation. Such perceptions hold particular relevance for training non-specialists and behavioral health providers who typically have less training in specialty mental health. Emerging empirical evidence has demonstrated that preliminary approaches to BA training using an online training format (Hubley et al., accepted) and modular training (Puspitassari et al. 2013) elicited positive reactions from trainees and expected training outcomes, although neither of these studies assessed the hypothesis that BA is easier to learn compared to other EBPTs. In further support of BA's dissemination potential, Ekers and colleagues (2011; 2012) have shown that mental health nurses without extensive EBPT training can learn and deliver BA for depressed primary care patients with outcomes similar to more experienced clinicians. Taken together, this accumulating evidence provides preliminary support for the notion that BA is an easy-to-learn, easy-to-implement EBPT, which may be especially important in the context of addressing the treatment gap via training non-specialists and behavioral health providers.

It is clear that too many people with mental illness receive no treatment. For those that do, too few receive state-of-the-art treatments with the strongest empirical support. Optimizing training approaches for non-specialists and behavioral health

providers in the delivery of EBPTs in routine practice settings is a promising approach to meaningfully reducing the gap between treatment need and treatment received.

References

Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61-74.

Aarons, G. A. (2005). Measuring provider attitudes toward evidence-based practice: Consideration of organizational context and individual differences. *Child and Adolescent Psychiatric Clinics of North America, 14*(2), 255-271.

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68*(2), 331-339.

Addis, M. E., Wade, W. A., & Hatgis, C. (1999). Barriers to dissemination of evidence‐based practices: Addressing practitioners' concerns about manual‐based psychotherapies. *Clinical Psychology: Science and Practice, 6*(4), 430-441.

Alexander, C. L., Arnkoff, D. B., & Glass, C. R. (2010). Bringing psychotherapy to primary care: Innovations and challenges. *Clinical Psychology: Science and Practice, 17*(3), 191-214.

Alvidrez, J., & Azocar, F. (1999). Distressed women's clinic patients: Preferences for mental health treatments and perceived obstacles. *General Hospital Psychiatry, 21*(5), 340-347.

Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning, 5*(3), 233-237.

Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology toward a scientifically principled approach to mental and behavioral

health care. *Psychological Science in the Public Interest, 9*(2), 67-103.

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M.B. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research, 16*(02), 229-240.

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*(5), 493-503.

Beidas, R. S., Mychailyszyn, M. P., Edmunds, J. M., Khanna, M. S., Downey, M. M., & Kendall, P. C. (2012). Training school mental health providers to deliver cognitive-behavioral therapy. *School Mental Health, 4*(4), 197-206.

Borntrager, C., Chorpita, B., Higa-McMillan, C., & Weisz, J. (2009). Provider attitudes toward evidence-based practices: Are the concerns with the evidence or with the manuals? *Psychiatric Services, 60*(5), 677-681.

Bright, J. I., Baker, K. D., & Neimeyer, R. A. (1999). Professional and paraprofessional group treatments for depression: A comparison of cognitive-behavioral and mutual support interventions. *Journal of Consulting and Clinical Psychology, 67*, 491-501.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7-18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*(1), 685-716.

Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5*(1), 8-14.

Cohen, L. H., Sargent, M. M., & Sechrest, L. B. (1986). Use of psychotherapy research by professional psychologists. *American Psychologist, 41*(2), 198-206.

Comer, J. & Barlow, D. (2013). The occasional case against broad dissemination and implementation. *American Psychologist 68*(5), 1-20.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909-922.

Decker, S.E., Jameson, M.T., & Naugle, A. E. (2011). Therapist training in empirically supported treatments: A review of evaluation methods for short- and long-term outcomes. *Administrative Policy in Mental Health, 38,* 254-286.

den Boer, P., Wiersma, D., Russso, S., & Bosch, R. (2005). Paraprofessionals for anxiety and depressive disorders. *Cochrane Database of Systematic Reviews*.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., . . . Brown, L. L. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62*(4), 409-416.

Dimeff, L. A., Koerner, K., Woodcock, E. A., Beadnell, B., Brown, M. Z., Skutch, J. M., . . . Harned, M. S. (2009). Which training method works best? A randomized controlled trial comparing three methods of training clinicians in dialectical behavior therapy skills. *Behaviour Research and Therapy, 47*(11), 921-930.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Gollan, J. K. (2006). Randomized trial of behavioral activation, cognitive therapy,

and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658-670.

Dimidjian, S., Martell, C. R., Addis, M. E., & Herman-Dunn, R. (2008). Behavioral activation for depression. In D. Barlow (Ed.), *Clinical Handbook of Psychological Disorders: A step-by-step treatment manual*, 328-364. (4th ed.). New York, NY: Guilford Press.

Dimidjian, S., Hubley, S., Martell, C., Herman, R., & Dobson, K. (unpublished manuscript). Quality of Behavioral Activation Scale. Boulder, Colorado: University of Colorado Boulder.

Dobson, K. S., Hollon, S. D., Dimidjian, S., Schmaling, K. B., Kohlenberg, R. J., Gallop, R. J., . . . Jacobson, N. S. (2008). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology, 76*(3), 468-477.

Durlak, J. A. (1979). Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin, 86*(1), 80-92.

Dwight‐Johnson, M., Sherbourne, C. D., Liao, D., & Wells, K. B. (2000). Treatment preferences among depressed primary care patients. *Journal of General Internal Medicine, 15*(8), 527-534.

Ekers, D., Dawson, M., & Bailey, E. (2012). Dissemination of behavioural activation for depression to mental health nurses: Training evaluation and benchmarked clinical outcomes. *Journal of Psychiatric and Mental Health Nursing, 20*(2),186-192.

Ekers, D., Godfrey, C., Gilbody, S., Parrott, S., Richards, D. A., Hammond, D., & Hayes, A. (2011). Cost utility of behavioural activation delivered by the non-specialist. *The British Journal of Psychiatry, 199*(6), 510-511.

Ekers, D., Richards, D., McMillan, D., Bland, J. M., & Gilbody, S. (2011). Behavioural activation delivered by the non-specialist: Phase II randomised controlled trial. *The British Journal of Psychiatry, 198*(1), 66-72.

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*(6), 373-378.

Flynn, H. A., Henshaw, E., O'Mahen, H., & Forman, J. (2009). Patient perspectives on improving the depression referral processes in obstetrics settings: a qualitative study. *General Hospital Psychiatry, 32*(1), 9-16.

Gega, L., Norman, I., & Marks, I. (2007). Computer-aided vs. tutor-delivered teaching of exposure therapy for phobia/panic: Randomized controlled trial with pre-registration nursing students. *International Journal of Nursing Studies, 44*(3), 397-405.

Goodman, J. H. (2009). Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. *Birth Issues in Perinatal Care, 36*(1), 60-69.

Grosse, S. D. (2008). Assessing cost-effectiveness in healthcare: History of the $50,000 per QALY threshold. *Expert Review of Pharmacoeconomics and Outcomes Research, 8*(2), 165-178.

Henggeler, S. W., Chapman, J. E., Rowland, M. D., Halliday-Boykins, C. A., Randall, J., Shackelford, J., & Schoenwald, S. K. (2008). Statewide adoption and initial implementation of contingency management for substance abusing adolescents. *Journal of Consulting and Clinical Psychology, 76*(4), 556.

Hofmann, S. G., & Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *The Journal of Clinical Psychiatry, 69*(4), 621.

Hoge, M., Morris, J., Daniels, A., Stuart, G., Huey, L., & Adams, N. (2007). *An action plan for behavioral health workforce development.* Washington, DC: Department of Health and Human Services.

Hollon, S. D. (2012). Cognitive and Behavior Therapy in the Treatment and Prevention of Depression. *FOCUS: The Journal of Lifelong Learning in Psychiatry, 10*(4), 506-509.

Hollon, S. D., & Ponniah, K. (2010). A review of empirically supported psychological therapies for mood disorders in adults. *Depression and Anxiety, 27*(10), 891-932.

Hopko, D. R., Lejuez, C., Ruggiero, K. J., & Eifert, G. H. (2003). Contemporary behavioral activation treatments for depression: Procedures, principles, and progress. *Clinical Psychology Review, 23*(5), 699-717.

Hubley, S., & Dimidjian, S. (2009). Behavioral Activation. In I. Weiner & W. Craighead (Eds.). *The Corsini Encyclopedia of Psychology and Behavioral Science.* New York, New York: John Wiley & Sons, Inc.

Hubley, S., Woodcock, E., Dimeff, L., & Dimidjian, S. (accepted). Disseminating behavioral activation via online training: Preliminary steps.

Hunkeler, E. M., Meresman, J. F., Hargreaves, W. A., Fireman, B., Berman, W. H., Kirsch, A. J., . . . Getzell, M. (2000). Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. *Archives of Family Medicine, 9*(8), 700-708.

Katon, W., Lin, E. H., & Kroenke, K. (2007). The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. *General Hospital Psychiatry, 29*(2), 147-155.

Katon, W., Von Korff, M., Lin, E., Walker, E., Simon, G. E., Bush, T., . . . Russo, J. (1995).

Collaborative management to achieve treatment guidelines impact on depression in primary care. *The Journal of the American Medical Association, 273*(13), 1026-1031.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science, 6*(1), 21-37.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . Wang, P. S. (2003). The epidemiology of major depressive disorder. *The Journal of the American Medical Association, 289*(23), 3095-3105.

Kessler, R. C., & Walters, E. E. (1998). Epidemiology of DSM‐III‐R major depression and minor depression among adolescents and young adults in the national comorbidity survey. *Depression and Anxiety, 7*(1), 3-14.

Kessler, R. C., & Wang, P. S. (2008). The descriptive epidemiology of commonly occurring mental disorders in the united states. *Annual Review of Public Health, 29*, 115-129.

Kirkpatrick, D.L. (1967). Evaluation of training. In R.L. Craig & L. R. Bittel (Eds.). *Training and development handbook.* New York, New York: McGraw-Hill.

Ko, J. Y., Farr, S. L., Dietz, P. M., & Robbins, C. L. (2012). Depression and treatment among US pregnant and nonpregnant women of reproductive age, 2005–2009. *Journal of Women's Health, 21*(8), 830-836.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ‐9. *Journal of General Internal Medicine, 16*(9), 606-613.

Kuyken, W., & Tsivrikos, D. (2008). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics, 78*(1), 42-48.

Laschinger, H. K. S., Finegan, J., Shamian, J., & Wilk, P. (2003). Workplace empowerment as a predictor of nurse burnout in restructured healthcare settings. *Longwoods Review, 1*(3), 2-

11.

Lundahl, B. W., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. L. (2010). A meta-analysis of motivational interviewing: Twenty-five years of empirical studies. *Research on Social Work Practice, 20*(2), 137-160.

Marcus, S. M., Flynn, H. A., Blow, F. C., & Barry, K. L. (2003). Depressive symptoms among pregnant women screened in obstetrics settings. *Journal of Women's Health, 12*(4), 373-380.

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in Context: Strategies for Guided Action*: New York, New York: Norton.

Martino, S., Ball, S. A., Nich, C., Frankforter, T. L., & Carroll, K. M. (2008). Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence, 96*(1), 37-48.

McBride, C., Atkinson, L., Quilty, L. C., & Bagby, R. M. (2006). Attachment as moderator of treatment outcome in major depression: A randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *Journal of Consulting and Clinical Psychology, 74*(6), 1041-1054.

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*(04), 457-471.

Montgomery, E. C., Kunik, M. E., Wilson, N., Stanley, M. A., & Weiss, B. (2010). Can paraprofessionals deliver cognitive-behavioral therapy to treat anxiety and depressive symptoms? *Bulletin of the Menninger Clinic, 74*(1), 45-62.

Morrow-Bradley, C., & Elliott, R. (1986). Utilization of psychotherapy research by practicing

psychotherapists. *American Psychologist, 41*(2), 188-197.

National Institute of Mental Health. (2013). Dissemination and Implementation Research in Health. Retrieved from http://grants.nih.gov/grants/guide/pa-files/PAR-13-054.html.

O'Mahen, H. A., & Flynn, H. A. (2008). Preferences and perceived barriers to treatment for depression during the perinatal period. *Journal of Women's Health, 17*(8), 1301-1309.

O'Malley, S. S., Suh, C. S., & Strupp, H. H. (1983). The Vanderbilt Psychotherapy Process Scale: A report on the scale development and a process-outcome study. *Journal of Consulting and Clinical Psychology, 51*(4), 581.

Olfson, M., Marcus, S. C., Druss, B., Elinson, L., Tanielian, T., & Pincus, H. A. (2002). National trends in the outpatient treatment of depression. *The Journal of the American Medical Association, 287*(2), 203-209.

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), 829-841.

Pignotti, M., & Thyer, B. A. (2009). Use of novel unsupported and empirically supported therapies by licensed clinical social workers: An exploratory study. *Social Work Research, 33*(1), 5-17.

Puspitasari, A., Kanter, J. W., Murphy, J., Crowe, A., & Koerner, K. (2013). Developing an online, modular, active learning training program for behavioral activation. *Psychotherapy, 50*(2), 256-267.

Rahman, A. (2007). Challenges and opportunities in developing a psychological intervention for perinatal depression in rural Pakistan–a multi-method study. *Archives of Women's Mental Health, 10*(5), 211-219.

Rahman, A., Malik, A., Sikander, S., Roberts, C., & Creed, F. (2008). Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: A cluster-randomised controlled trial. *The Lancet, 372*(9642), 902-909.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Rotheram-Borus, M. J., Swendeman, D., & Chorpita, B. F. (2012). Disruptive innovations for designing and diffusing evidence-based interventions. *American Psychologist, 67*(6),463-476.

Seng, A. C., Prinz, R. J., & Sanders, M. R. (2006). The role of training variables in effective dissemination of evidence-based parenting interventions. *International Journal of Mental Health Promotion, 8*(4), 20-28.

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., . . . Imber, S. D. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67*(6), 837-846.

Sholomskas, D. E., & Carroll, K. M. (2006). One small step for manuals: Computer-assisted training in twelve-step facilitation. *Journal of Studies on Alcohol, 67*(6), 939-945.

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.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420-428.

Simon, G. E., VonKorff, M., Rutter, C., & Wagner, E. (2000). Randomised trial of monitoring, feedback, and management of care by telephone to improve treatment of depression in primary care. *British Medical Journal, 320*(7234), 550-554.

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. L., 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,* 751-756.

Simons, A. D., Rozek, D. C., & Serrano, J. L. (2013). Wanted: Reliable and valid measures for the science of cognitive behavioral therapy dissemination and implementation. *Clinical Psychology: Science and Practice, 20*(2), 181-194.

Sleath, B., Rubin, R. H., & Huston, S. A. (2003). Hispanic ethnicity, physician-patient communication, and antidepressant adherence. *Comprehensive Psychiatry, 44*(3), 198-204.

Sockol, L. E., Epperson, C. N., & Barber, J. P. (2011). A meta-analysis of treatments for perinatal depression. *Clinical psychology review, 31*(5), 839-849.

Stewart, R. E., & Chambless, D. L. (2007). Does psychotherapy research inform treatment decisions in private practice? *Journal of Clinical Psychology, 63*(3), 267-281.

Strunk, D. R., Brotman, M.A., DeRubeis, & Hollon, S.D. (2009). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology, 78*(3),429-437.

Thompson, L. W., Gallagher, D., Nies, G., & Epstein, D. (1983). Evaluation of the effectiveness of professionals and nonprofessionals as instructors of “coping with depression” classes for elders. *The Gerontologist, 23*(4), 390-396.

Trepka, C., Rees, A., Shapiro, D. A., Hardy, G. E., & Barkham, M. (2004). Therapist competence and outcome of cognitive therapy for depression. *Cognitive Therapy and Research, 28*(2), 143-157.

United States Department of Labor. (2003). Quick facts on registered nurses. Retrieved from http://www.dol.gov/wb/factsheets/QF-nursing.htm

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620-630.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*(2), 2000-2011.

Wells, K. B., Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unutzer, J., . . . Rubenstein, L. V. (2000). Impact of disseminating quality improvement programs for depression in managed primary care: A randomized controlled trial. *Journal of the American Medical Association, 283*(2), 212-220.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological bulletin, 130*(4), 631-663.

World Health Organization (2013). The Mental Health Action Plan. Retrieved from http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf

Worrall, J. M., & Fruzzetti, A. E. (2009). Improving peer supervisor ratings of therapist performance in dialectical behavior therapy: An Internet-based training system. *Psychotherapy: Theory, Research, Practice, Training, 46*(4), 476-479.

Young, J., & Beck, A. T. (1980) *Cognitive Therapy Scale rating manual.* Philidelphia, PA: University of Pennsylvania.

Young, A. S., Klap, R., Sherbourne, C. D., & Wells, K. B. (2001). The quality of care for depressive and anxiety disorders in the United States. *Archives of General Psychiatry, 58*(1), 55.