Accepted Manuscript

Investigating Bang for your Training Buck: A Randomized Controlled Trial Comparing Three Methods of Training Clinicians in Two Core Strategies of Dialectical Behavior Therapy

Linda A. Dimeff, Melanie S. Harned, Eric A. Woodcock, Julie M. Skutch

PII: S0005-7894(15)00016-7
DOI: doi: 10.1016/j.beth.2015.01.001
Reference: BETH 538

To appear in: Behavior Therapy

Received date: 21 February 2013
Accepted date: 16 January 2015

Please cite this article as: Dimeff, L.A., Harned, M.S., Woodcock, E.A. & Skutch, J.M., Investigating Bang for your Training Buck: A Randomized Controlled Trial Comparing Three Methods of Training Clinicians in Two Core Strategies of Dialectical Behavior Therapy, Behavior Therapy (2015), doi: 10.1016/j.beth.2015.01.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Investigating Bang for your Training Buck: A Randomized Controlled Trial Comparing Three Methods of Training Clinicians in Two Core Strategies of Dialectical Behavior Therapy

Linda A. Dimeff, Melanie S. Harned, Eric A. Woodcock, & Julie M. Skutch

Behavioral Tech Research, Inc., Seattle, WA
Abstract

The present study examined the efficacy of online training (OLT), instructor-led training (ILT), and a treatment manual (TM) in training mental health clinicians in two core strategies of Dialectical Behavior Therapy (DBT): chain analysis and validation. A randomized controlled trial compared OLT, ILT, and TM among clinicians naïve to DBT (N = 172) who were assessed at baseline, post-training, and 30, 60, and 90 days following training. Primary outcomes included satisfaction, self-efficacy, motivation, knowledge, clinical proficiency, and clinical use. Overall, ILT outperformed OLT and TM in satisfaction, self-efficacy, and motivation, whereas OLT was the most effective method for increasing knowledge. The conditions did not differ in observer-rated clinical proficiency or self-reported clinical use, which both increased to moderate levels after training. In addition, ILT was particularly effective at improving motivation to use chain analysis, whereas OLT was particularly effective at increasing knowledge of validation strategies. These findings suggest that these types of brief, didactic trainings may be effective methods of increasing knowledge of new treatment strategies, but may not be sufficient to enable clinicians to achieve a high level of clinical use or proficiency. Additional research examining the possible advantages of matching training methods to types of treatment strategies may help to determine a tailored, more effective approach to training clinicians in empirically supported treatments.

Keywords: dissemination, online training, empirically supported treatment, Dialectical Behavior Therapy
Investigating Bang for your Training Buck: A Randomized Controlled Trial Comparing Three Methods of Training Clinicians in Two Core Strategies of Dialectical Behavior Therapy

The considerable divide between research on empirically supported treatments (ESTs) and clinical practice is a widely recognized problem in current mental health care (e.g., APA Presidential Task Force on Evidence-Based Practice, 2006; Insel, 2009). Although numerous ESTs now exist for a variety of psychological disorders and problems, these powerful treatments have remained largely inaccessible to consumers of mental health services (e.g., Freiheit, Vye, Swan, & Cady, 2004). Given the persistent underutilization of ESTs, policymakers and funding agencies have become increasingly focused on the urgent need to promote the use of ESTs in clinical practice. Beginning in 2008, for example, the National Institute of Mental Health (NIMH) announced a five-year strategic plan that includes a specific effort to fund research that will “help close the gap between the development of new, research-tested interventions and their widespread use by those most in need” (U.S. Department of Health & Human Services, 2008; p. 28). Similarly, organizations (e.g., Department of Veterans Affairs, 2008) and states (e.g., Nakamura et al., 2011) have begun to launch initiatives designed to increase consumers’ access to ESTs.

One of the primary obstacles to disseminating ESTs has been a shortage of clinicians who are trained to deliver these treatments (e.g., Weissman et al., 2006). The increasing demand for evidence-based care has therefore led to the development of clinician training programs that are now being rolled out at the national, state, and individual treatment developer levels (McHugh & Barlow, 2010). Although this increased attention to training clinicians in ESTs is a sign of progress in many ways, it also brings with it a new set of problems. In particular, these dissemination programs often utilize training methods of unknown efficacy. Indeed, research on
how best to train mental health professionals in ESTs is a relatively new field, with two recent reviews finding fewer than ten studies of therapist training that utilize methodologically rigorous randomized controlled trial (RCT) designs (Beidas & Kendall, 2010; Herschell, Kolko, Baumann, & Davis, 2010).

Within leading dissemination programs, the most commonly utilized didactic training methods include the distribution of training materials (e.g., treatment manuals), workshops, and online training (McHugh & Barlow, 2010). The extant research indicates that online trainings and in-person workshops, particularly when they include active learning methods (e.g., practice opportunities, role-plays), are generally superior to treatment manuals that rely on passive learning strategies (Beidas & Kendall, 2010; Dimeff et al., 2009; Dimeff, Beadnell, Woodcock, & Harned, 2011; Sholomskas & Carroll, 2006). Studies comparing online training and workshops have yielded mixed results, with the majority finding no differences in knowledge gains across the two training methods (Beidas, Edmunds, Marcus, & Kendall, 2012; McDonough & Marks, 2002; Sholomskas et al., 2005; Weingardt, Villafranca, & Levin, 2006), whereas one study found that online training outperformed workshops in increasing clinician knowledge (Dimeff et al., 2009).

Although some research has evaluated the utility of these common training methods, no studies have yet examined the possible advantages of matching training methods to particular types of content. Didactic training in ESTs typically covers a wide variety of treatment strategies, including protocol-based strategies (i.e., those administered in a structured, step-by-step manner in specific, prescribed circumstances) as well as principle-based strategies (i.e., those that are implemented flexibly according to the overarching theory and proposed mechanisms of the treatment). In addition, didactic training in ESTs often requires learning
strategies from different theoretical backgrounds, such as change-based interventions from behavior therapy and acceptance-based interventions derived from Western contemplative and Eastern Zen practices. These different types of strategies may require different types of learning and, therefore, different approaches to training.

The present study had two aims. First, given the paucity of research on methods of training clinicians in ESTs, we sought to replicate existing research utilizing a methodologically rigorous RCT design. To that end, we examined the relative efficacy of three methods (online training, OLT; instructor-led training, ILT; and a treatment manual, TM) of training clinicians in two core strategies of DBT (Linehan, 1993): chain analysis and validation strategies. This study design is consistent with prior research evaluating methods of training clinicians in DBT skills (Dimeff et al., 2009), and extends this work to evaluate the efficacy of these training methods for novel content areas within DBT. Consistent with Kirkpatrick’s (1998) multi-level model of evaluating training programs, the present study assessed learners’ satisfaction with the training and barriers encountered (Level One), knowledge, self-efficacy, and motivation to use the training content (Level Two), and clinical proficiency and use (Level Three). We hypothesized that OLT and ILT would result in comparable outcomes across all three levels at post-training and follow-up, and that both conditions would be superior to TM.

Second, we sought to extend prior research by conducting exploratory analyses to examine whether the efficacy of these training methods varied depending on the type of treatment strategy being taught. Chain analysis is a change-based strategy similar to functional analysis that involves assessing each link in the chain of events that led up to and followed a specific problem behavior. This information is then used to generate solutions that could prevent the behavior from recurring. In DBT, chain analysis is a semi-structured procedure that is used
in most instances when a client engages in serious dysfunctional behavior. Validation is an acceptance-based strategy that involves communicating to the client that his or her responses make sense and are understandable. In DBT, validation is used as needed depending on the characteristics of the situation and the specific client. For example, therapists may use validation to model more compassionate ways of viewing one’s behavior or as a strategy for increasing clients’ willingness to work on changing a problem behavior. These two core DBT strategies therefore differ along several key dimensions. Whereas chain analysis is a change-oriented strategy that is protocol-based, validation is an acceptance-oriented strategy that is principle-based. Thus, we explored whether any differences arose in learning these two treatment strategies between training conditions.

Method

Recruitment and Screening

All procedures were approved by the Western Institutional Review Board. The study was conducted between October 2007 and January 2009. The target population for this training was clinicians providing counseling and/or mental health services to individuals with BPD with and without substance abuse problems. To maximize the generalizability of the findings, broad inclusion criteria were used that reflected the heterogeneity of clinicians that provide these services, including students in training, chemical dependency counselors (e.g., with AA or BA degrees), and licensed mental health professionals (i.e., with masters or doctoral degrees). Participants were recruited via professional listservs, a database of clinicians who had participated in previous studies, and direct contact with drug abuse and mental health treatment agencies within the Pacific Northwest. Interested individuals completed a phone screen to determine study eligibility. Inclusion criteria included: (1) minimal previous experience using
the strategies taught in the training (i.e., having not applied DBT validation strategies and chain analysis strategies in their clinical practice); (2) no extensive training in DBT (i.e., had not read the DBT manuals or received three or more days of DBT training); (3) access to a computer with a sound card and Internet; (4) treating at least one individual with substance abuse problems; and (5) available and willing to participate in a two-day instructor-led workshop in Seattle, WA if assigned to the ILT condition.

**Randomization**

The 200 eligible participants were randomized by the participant coordinator to one of the three conditions using a minimization random assignment procedure (White & Freedman, 1978) that matched participants on clinical experience (<2 years, 2-6 years, and >6 years), highest education level (High School or Associate’s degree, Bachelor’s degree, Master’s degree or RN, and PsyD, PhD or MD), and prior exposure to DBT: none/minimal (e.g., attended an in-service training, grand rounds, or lecture) or some/moderate (e.g., read part of the primary DBT manuals, attended a one- or two-day workshop).

**Training Conditions**

To ensure the consistency of content across conditions, the training materials for all three conditions were developed by the same expert team led by the first author. Course content and learning objectives were first developed for both OLTs and were then applied to the development of the TMs and ILTs. Each set of training materials was developed to cover the two topic areas in equal proportion, with two-thirds of the training focused on chain analysis and one-third focused on validation strategies. Greater time was needed to teach chain analysis given that a thorough understanding of this technique also required a detailed review of behavioral theory and intervention.
Online Training (OLT; $n=55$). Participants in this condition received access to two online training courses teaching DBT chain analysis (http://behavioraltech.org/ol/details_chain.cfm) and DBT validation strategies (http://behavioraltech.org/ol/details_validation.cfm). The courses were developed utilizing best practices from product development models, including formative evaluation and iterative design that involved extensive testing with the target audience and DBT experts. The final courses were designed in Adobe Flash and were interactive, media-rich, and incorporated expert commentaries, practice exercises, simulated therapy, printable handouts and study guides, and knowledge checks. The validation strategies OLT involved approximately four hours of instruction, whereas the chain analysis OLT was approximately eight hours in length. Participants had access to both courses throughout the training and follow-up periods.

Instructor-led Training (ILT; $n=55$). Participants in this condition attended a two-day (12-hour) workshop in Seattle, WA conducted by two DBT experts (the first and second authors). The workshop was developed to be consistent with the training methods and principles of gold-standard ILTs and included both passive learning (e.g., lecture, video) and active learning methods (e.g., role-play, experiential exercises). To be consistent with the OLTs, approximately two-thirds of the ILT was devoted to DBT chain analysis and one-third to DBT validation strategies. All participants attended the same, one-time workshop. Participants were provided a binder of printed PowerPoint slides to capture their notes during the workshop.

Treatment Manual (TM; $n=62$). Participants in this condition received a text-based treatment manual composed of training materials on DBT chain analysis (133 pages) and DBT validation strategies (59 pages). The manual was created specifically for this trial by the first author and Dr. Marsha Linehan, the DBT treatment developer. The manual contained expert
descriptions and explanations of each concept, clinical scenarios, and practice examples and was accompanied by a suggested study guide.

**Assessment Procedures**

**Baseline assessment.** At baseline, participants completed a performance-based role play over the telephone and were then directed to an online assessment system to complete the remainder of the measures. Of the randomized participants, 28 (14.0%) did not complete the baseline assessment; this, however, was not associated with training condition, $\chi^2(2) = 3.62, p = .16$. Upon completion of the baseline assessment, participants were considered enrolled in the trial, were informed of their training condition, and were issued their study materials. Study materials included a hyperlink, username and password for the OLT, a text-based treatment manual for the TM condition, or admission tickets and directions to the ILT. Participants were not compensated for completing the baseline assessment.

**Follow-up assessments.** All participants completed the post-training assessment shortly after they completed the training; however, procedures for scheduling the post-training assessment differed depending on experimental condition. OLT and TM participants were instructed to schedule and complete the post-training assessment once they had completed their training materials, or at the end of the 30-day training phase (with the possibility of one 2-week extension if requested), whichever came first. ILT participants were scheduled to complete their post-training assessment the week following the two-day workshop. At 30 and 60 days post-training, all participants completed a brief survey assessing their use of chain analysis and validation in clinical practice over the past 30 days. A final assessment was conducted at 90 days post-training that included the full battery of outcome measures. Participants were paid $50 for completing the training, $50 each for the post-training and 90-day follow-up assessments,
and $25 each for the 30- and 60-day follow-up assessments for a maximum total payment of $200. All participants had access to their training materials during the follow-up period.

**Participant Flow and Retention**

Participant flow from initial screening through the 90-day follow-up assessment is provided in Figure 1. Of the 200 randomized participants, 172 (86.0%) completed the baseline assessment. Between baseline and the post-training assessment, significantly more OLT participants (36.4%) dropped out of the study than ILT and TM participants (5.5% and 6.5%, respectively; \( \chi^2(2) = 26.11, p < .0001 \)). The higher attrition in OLT was likely the result of an unexpected delay of approximately 6.5 months (due to production setbacks in its development) between the baseline assessment and the start of the OLT training. In contrast, ILT and TM participants received access to their training shortly after completing the baseline assessment. Despite this differential attrition, no between-condition differences were found on key demographic variables (age, gender, ethnicity, education, profession, work setting, and clinical experience) among those participants who completed the post-training assessment (\( p \)'s > .10). In addition, OLT participants who did versus did not drop out of the study before post-training did not differ on baseline scores on any outcome measure (\( p \)'s > .08). Power analyses indicated that 165 participants (55 per condition) would be sufficient to detect medium effect sizes (\( d = 0.54 \)) using a two-tailed test (\( \alpha = 0.05 \)) at the minimum power level recommended for RCTs (power = 0.80; Cohen, 1988).

**Measures**

Unless otherwise noted, all measures were administered at the baseline, post-training, and 90-day follow-up assessments.
**Participant characteristics.** All participants completed a 13-item measure at baseline which assessed demographic, theoretical and professional characteristics.

**Training completion.** A single item was used to assess training completion at the post-training assessment for the OLT and TM conditions (“What is your best estimate for the percentage of the training you completed?”). Given the nature of the ILT, training completion in this condition was assessed via sign-in/sign-out data for each of the two days (similar to the standards used in Continuing Education workshops).

**Barriers.** This self-report measure was adapted from an existing measure (Dimeff et al., 2009) and administered at the post-training assessment to assess potential barriers to learning and using a new treatment. Items (n=20) assessed organizational (e.g. “Conflict between agency policies and DBT philosophy or practice”), systemic (e.g. “Problems with getting insurance reimbursement for using/applying DBT”), and attitudinal barriers (e.g. “My own non-behavioral theoretical orientation”) to using the information learned during the training. Barrier items were rated on a 4-point Likert scale ranging from 1 (Not applicable/No Impact) to 4 (Significant Impact) and averaged to create a total score for analysis.

**Satisfaction.** A 9-item satisfaction measure from Dimeff et al. (2009) assessed effectiveness of the training style and presentation methods at the post-training assessment. Items were generic (not specific to training condition or content area), rated on a 5-point Likert scale from 1 = “Poor” to 5 = “Excellent,” and averaged to create a total score (Cronbach’s alpha = .91). Example items include: “To what extent did the training hold your interest?” and “To what extent was the information provided in the training clear and understandable?”

**Self-efficacy.** An adapted 33-item version of the self-efficacy subscale of the Behavioral Anticipation and Confidence Questionnaire (BAQ; Dimeff et al., 2009) measured participants’
self-reported confidence in their ability to implement chain analysis and validation strategies. Items from the original BAQ were adapted by DBT experts to reflect the content areas of chain analysis and validation strategies. A total score, as well as two subscales specific to each content area, were analyzed. The chain analysis subscale included 11 items (e.g., “I feel confident in my ability to conduct a chain analysis with a difficult client”) and the validation strategies subscale included 10 items (e.g., “I feel confident in my ability to generate validation statements across all six validation levels used in DBT”). Items were rated on a 5-point Likert scale ranging from 1 (Not Confident) to 5 (Very Confident). Internal consistency was excellent for both subscales and the total score (Cronbach’s alphas ranged from .93 to .97, depending on time point). In addition, the total scores at post-training and 90-day follow-up were significantly correlated ($r = .70, p < .001$), suggesting the measure had excellent test-retest reliability.

**Motivation.** An adapted 16-item version of the motivation subscale of the BAQ (Dimeff et al., 2009) measured participants’ self-reported motivation to learn and use chain analysis and validation strategies. Items from the original BAQ were adapted by DBT experts to reflect the content areas of chain analysis and validation. A total score and two subscales specific to each content area were analyzed. The chain analysis subscale included five items (e.g., “I am motivated to use chain analysis in my clinical practice” and “I am motivated to use chain analysis procedures to better understand and treat my client’s drug problem”) and the validation strategies subscale included four items (e.g., “I am motivated to practice validation strategies” and “I am motivated to teach my colleagues validation strategies”). Items were rated on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Internal consistency was excellent for both subscales as well as the total score (Cronbach’s alpha ranged from .87 to .93, depending on time point). In addition, the total scores at post-training and 90-day follow-up
were significantly correlated ($r = .72, p < .001$), suggesting the measure had excellent test-retest reliability.

**Knowledge.** A 55-item chain analysis knowledge test and a 41-item validation strategies knowledge test were developed to assess knowledge of the course content for each content area. All items were piloted iteratively with content-naïve mental health professionals as well as DBT experts. Items not answered correctly by DBT experts or those answered correctly significantly more frequently than chance by content-naïve clinicians were removed or modified to ensure construct validity and reliability. Items included a combination of multiple choice and matching questions. Mean scores for the proportion of items answered correctly were calculated for each measure as well as a total (combined) score. The total knowledge scores at post-training and 90-day follow-up were significantly correlated ($r = .71, p < .001$), suggesting the measure had excellent test-retest reliability.

**Clinical proficiency.** Structured role plays were used to assess participants’ proficiency in applying the course content in simulated clinical scenarios. Three brief 20-minute role plays were developed and evaluated for equivalency with feedback from a panel of DBT experts. The role plays required participants to conduct a chain analysis of a fictional client’s most recent problem behavior (drug use, suicide attempt, or non-suicidal self-injury) while weaving in validation strategies as needed. Research assistants (RAs) were trained to act like BPD clients and were blind to participants’ condition and assessment point. Role play scenarios were randomly ordered for each participant and were conducted and digitally recorded via telephone. A coding instrument was developed based on relevant items from the established DBT adherence measure (Linehan & Korslund, 2003). For chain analysis, 16 items assessed six components of conducting a chain analysis: orienting/commitment, use of collaborative language, hypothesis
testing, summarizing links in the chain, solution generation/troubleshooting, and connecting the problem behavior to the client’s goals. For validation, eight items assessed each of the six levels of validation as well as cheerleading and validation of wise mind. Two coders were trained to reliability and were checked for drift on a monthly basis by a DBT expert (the first author) who served as the gold standard coder. Corrective feedback was provided to actors on an ongoing basis based on review of the digital recordings. A global Clinical Proficiency score was computed, in addition to scores for the chain analysis and validation strategies scales separately, that ranged from 0 (Poor/Absent) to 3 (Very good/Highly proficient). A random sample (20% of all role plays) was coded for reliability and inter-rater reliability was excellent for both chain analysis (mean ICC = .91) and validation strategies (mean ICC = .87).

Clinical use. Participants were asked to report the frequency of their clinical use of chain analysis and validation strategies at the 30, 60, and 90 day follow-up assessments. The items read: “In the past 30 days, how many times did you teach or apply DBT validation strategies [DBT chain analysis] in your clinical work?” The items were rated on a 6 point ordinal scale: 1 (0 times), 2 (1-4 times), 3 (5-9 times), 4 (10-15 times), 5 (15-20 times), and 6 (more than 20 times).

Analysis Strategy

Analyses were performed using Hierarchical Linear Modeling (HLM) using PROC MIXED models in SAS 9.3. Predictors in the analyses were condition, time, and a time by condition interaction. Given that the focus of the HLM analyses was on evaluating change over time, only participants who completed baseline and at least one outcome assessment were included (n=145, see Figure 1). Between condition differences were reflected in the time by condition effect for outcomes measured at both baseline and at least one outcome assessment.
time point, and the condition effect for outcomes measured only at follow-up time points. The best-fitting covariance structures were selected and specified based on the Akaike's Information Criterion (AIC; Burnham & Anderson, 2004; Vallejo, Ato & Valdes, 2008). Type III F-tests (omnibus tests to determine if differences existed between conditions) were calculated, and when significant, pairwise Scheffé-adjusted comparisons of predicted means from regression estimates were used to identify which conditions differed at each time point.

Results

Sample Description

No significant differences were found between conditions on any demographic variable (see Table 1). The modal study participant was a Caucasian female approximately 40 years of age with a Masters degree and 6.5 years of clinical experience who was practicing in an outpatient community mental health clinic.

Training Completion and Barriers Encountered

Training completion. A significant Condition effect was found for percentage of training completed at the post-training assessment ($F (2, 142) = 3.95, p < .05$). Participants in the ILT condition completed more of their training than those in the TM condition ($t (142) = 2.78, p < .01$). In general, participants had completed a high percentage of their training by the post-training assessment: TM = 89%, OLT = 92%, and ILT = 98% complete.

Barriers. The three conditions did not significantly differ in terms of barriers at post-training ($F (2, 142) = 1.82, p = .17$). Across conditions, the most common barriers were: (1) no regular clinical supervision available to support skills learned in training ($M = 2.59, SD = 1.13$), (2) no formal commitment from my organization to learning and implementing course content
(\(M = 2.17, SD = 1.04\)), and (3) no relief time from work provided for learning and implementing course content (\(M = 2.08, SD = 1.01\)).

**Primary Outcome Analyses**

Table 2 presents descriptive data for the longitudinal outcome total scores (chain analysis and validation strategies content areas combined) at each time point, by condition, and the associated statistics from the HLM analyses. Between-condition Cohen’s \(d\) effects sizes are shown in Table 3.

**Satisfaction.** Cross-sectional analyses at post-training showed a significant between-condition difference on satisfaction with the training style and presentation methods (\(F(2, 142) = 9.99, p < .0001\)). ILT participants rated the training as more satisfactory than those in OLT (\(t(142) = 2.82, p < .01\)) and TM (\(t(142) = 4.38, p < .0001\)). Overall, participants reported the training styles and methods to be on average between “good” and “very” satisfactory: ILT (\(M = 4.01, SD = 0.64\)), OLT (\(M = 3.57, SD = 0.76\)), and TM (\(M = 3.41, SD = 0.74\)).

**Self-efficacy.** A significant Time effect indicated that all conditions reported an increase in self-efficacy over time (\(F(2, 279) = 140.63, p < .001\)). A significant Condition effect (\(F(2, 164) = 13.95, p < .001\)) indicated differences between conditions across time points. A non-significant Time by Condition interaction effect (\(F(4, 277) = 0.76, p = .55\)) indicated that self-efficacy did not change differentially by condition across time. Post-hoc comparisons indicated that ILT participants rated their self-efficacy as significantly higher at baseline, post-training, and final follow-up compared to TM participants (\(t(283) = 3.41, p < .001; t(296) = 4.53, p < .001; t(326) = 4.61, p < .001\); respectively). Moreover, ILT participants reported greater self-efficacy compared to OLT participants at baseline, post-training, and final follow-up (\(t(283) = 2.42, p < .05; t(339) = 3.18, p < .01; t(356) = 3.31, p < .001\); respectively). No differences were found
between OLT and TM conditions at any time point (p’s > .30). Given the significant baseline differences, the Condition effect is best interpreted as indicating that ILT participants began the study with higher self-efficacy and their scores remained higher than both other conditions over time.

**Motivation.** A significant effect for Time indicated the sample as a whole experienced a decrease in motivation over time \(F(2, 142) = 16.51, p < .0001\). Moreover, a significant Time by Condition interaction indicated differential change over time by condition \(F(4, 141) = 2.80, p < .05\). Post-hoc comparisons revealed participants in the ILT condition reported higher motivation at the post-training and final follow-up assessments compared to the TM \(t(143) = 2.34, p < .05; t(139) = 3.16; p < .005\;\text{respectively}\). Additionally, ILT participants reported higher motivation compared to OLT participants at post-training \(t(148) = 2.09, p < .05\), but not at final follow-up \(t(141) = 1.51, p = .13\). Motivation started and remained high for all conditions throughout the study despite the significant decrease over time (means ranged from 3.74 - 4.24 on a 1-5 Likert scale).

**Knowledge.** At baseline, participants answered between 37-38% of questions correctly. A significant effect for Time indicated the sample as a whole generally improved on knowledge across time \(F(2, 287) = 344.49, p < .0001\). Additionally, a significant Time by Condition interaction effect \(F(4, 286) = 4.14, p < .01\) indicated differential knowledge gains by condition over time. Post-hoc comparisons indicated that OLT participants demonstrated greater knowledge at the post-training and final follow-up assessments compared to TM \(t(342) = 3.97, p < .0001; t(363) = 2.45; p < .05\;\text{respectively}\) and ILT \(t(337) = 2.57, p < .05; t(357) = 2.42; p < .05\;\text{respectively}\). Knowledge test scores ranged from 53% to 64% correct across conditions at the post-training and final follow-up assessments.
Clinical proficiency. A significant effect for Time indicated general improvement for the sample in ability to perform the treatment strategies in role plays \((F(2, 156) = 57.24, p < .0001)\). The Time by Condition interaction was not significant \((F(4, 155) = 0.47, p = .76)\), indicating there was no differential change between the conditions over time. Means at the post-training and final follow-up assessments indicated moderate proficiency following training.

Clinical use. A significant effect for Time indicated increased clinical use of the treatment strategies following training \((F(2, 130) = 4.26, p < .05)\). The Condition effect was not significant \((F(2, 127) = 0.73, p = .48)\), indicating there was no overall difference between conditions over the follow-up period. Descriptive data indicate that participants used chain analysis and/or validation strategies on average between 1-9 times during each 30-day follow-up period.

Exploratory Outcome Analyses

Exploratory analyses were conducted to compare training conditions separately by content area taught (chain analysis vs. validation strategies). For the longitudinal outcomes, between-condition Cohen’s \(d\) effects sizes are shown in Table 3 and results of HLM analyses are presented in Table 4 by content domain. These analyses indicated that all three conditions resulted in improved self-efficacy across time points for each content area. ILT was superior to TM and OLT at every time point regardless of the content area being taught, with the exception of baseline for validation strategies (ILT was superior to TM, but not OLT). Similarly, ILT was superior in increasing motivation to use chain analysis compared to TM (at both post-training and 90-day follow-up) and OLT (at post-training only). The Time by Condition interaction effect was not significant for motivation to use validation strategies. OLT was particularly effective at increasing knowledge of validation strategies, with OLT outperforming ILT and TM.
at both outcome points. The Time by Condition interaction effect was not significant for knowledge of chain analysis. There were no between-condition differences for either chain analysis or validation strategies in clinical proficiency and no clear pattern of results emerged for self-reported use of the strategies in clinical practice.

**Discussion**

The present study examined the effectiveness of three methods of training mental health professionals in two core strategies of DBT: chain analysis and validation. Interactive online training courses (OLT) were compared to a two-day instructor-led workshop (ILT) and a written treatment manual (TM). Results suggested that all three training methods were satisfying and improved self-efficacy, knowledge, clinical proficiency, and clinical use. Overall, ILT outperformed the other training methods in terms of satisfaction and improving motivation, whereas OLT was the most effective method of increasing knowledge. The training methods yielded comparable improvements in clinical proficiency and clinical use. The pattern of differences on several outcomes varied depending on the treatment strategy being taught.

In contrast to hypotheses, ILT was the preferred training method in terms of learner satisfaction and resulted in higher motivation, particularly regarding chain analysis, than both OLT and TM. Moreover, there were no differences between OLT and TM for these outcomes. Although this is consistent with some prior research finding that clinicians prefer face-to-face over computer-based training (e.g., Beidas et al., 2012; McDonough & Marks, 2002), it is inconsistent with prior research indicating that online training is as or more effective than manuals and workshops in terms of satisfaction (Dimeff et al., 2009). This discrepancy appears to be due to somewhat lower satisfaction with the OLTs in the present study ($M=3.57$) compared to the Dimeff et al. (2009) study ($M=3.81$), as satisfaction with the ILTs was comparable across
the two studies ($M's = 3.99-4.01$). The OLTs in the present study utilized more text and were somewhat less interactive than the OLT used in Dimeff et al. (2009), which may explain the difference in user ratings. Alternatively, the discrepant findings may be due differences in the two study samples, with the present study including a higher proportion of chemical dependency counselors and clinicians working in outpatient community mental health centers. Importantly, however, the higher self-efficacy at baseline and throughout the study as well as the greater improvements in motivation found in ILT did not translate into superior knowledge gains, clinical proficiency, or use of newly learned skills in clinical practice. Similar to prior studies (e.g., Miller & Mount, 2001), these findings suggest that clinicians who attend brief workshops may overestimate their knowledge of and ability to implement newly learned treatment strategies. Alternatively, it is possible that clinicians in OLT and TM may have underestimated their knowledge and clinical proficiency, resulting in lower self-efficacy and motivation.

In terms of knowledge gains, OLT outperformed ILT and TM, particularly in increasing knowledge of validation strategies. The superiority of online training in increasing knowledge of ESTs is consistent with prior studies that have examined OLTs that utilize cutting-edge instructional design technology with extensive use of active learning methods (Dimeff et al., 2009). This study replicates this prior finding and suggests that well-designed OLTs may have an advantage over traditional training methods. However, the greater knowledge among clinicians completing the OLT did not lead to superior clinical proficiency or use, suggesting that knowledge alone is not sufficient to substantially enhance clinician behavior. Indeed, all three training methods resulted in significant and comparable gains in clinical use and proficiency, with clinicians increasing to moderate levels on both outcomes. Consistent with the emerging consensus in the field, these results suggest that relatively brief didactic trainings may need to be
augmented with supervision or consultation for clinicians, perhaps particularly those with no prior experience in a given EST, to reach a high level of competence and to further increase knowledge gains (e.g., Beidas et al., 2012; Harned et al., 2014; Herschell et al., 2010; Miller, Yahne, Moyers, Martinez & Pirritano, 2004; Sholomskas et al., 2005). Given that lack of available supervision is a primary barrier to implementation cited by clinicians in this study and others (e.g., Dimeff et al., 2009), more work is clearly needed to develop effective methods of making supervision and consultation more widely available. Overall, these findings suggest that didactic training methods may best be used as a first step for acquiring knowledge of an EST, and that online training may be the most effective method of doing this. Of note, online training is also an accessible and cost-effective method of receiving training in ESTs (Weingardt, 2004), which may make it a particularly attractive training option.

Finally, the effectiveness of the three training methods varied somewhat depending on the type of treatment strategy being taught. Specifically, ILT was particularly effective at improving motivation to use chain analysis, whereas OLT was particularly effective in increasing and maintaining knowledge of validation strategies. These findings suggest that face-to-face workshops may be particularly well-suited for teaching unfamiliar and complex strategies such as chain analysis. In the present study, teaching of chain analysis included a detailed review of behavioral learning theory (i.e., operant and classical conditioning) and it is possible that this type of theoretical or academic content may be more compelling when presented by live trainers versus online and text-based training methods, perhaps particularly among clinicians with lower levels of education such as many of those in the present study. On the other hand, online trainings may be the preferred method for teaching more familiar and less complex strategies such as validation, which may primarily require reading and viewing multiple clinical examples.
In addition, it is possible that protocol-based strategies such as chain analysis may best be learned via lecture and expert modeling such as in ILT, whereas principle-based strategies such as validation may require more opportunity to apply the content in clinical simulations and receive feedback such as in OLT. Importantly, given that the two treatment strategies being taught differed on multiple dimensions, it is not possible to draw firm conclusions about how best to match different types of treatment strategies to specific training methods. Instead, these preliminary findings should be used to inform future research on this important topic with the goal of tailoring training methods to different types of didactic content.

The present study had several limitations that naturally emerge when conducting hybrid efficacy-effectiveness research and that should be considered when interpreting the findings. First, the higher rate of attrition in OLT, although likely caused by factors unrelated to the OLT (i.e., production setbacks that delayed access to the course), resulted in a greater amount of missing data in this condition that may have biased the findings. Moreover, it is possible that baseline scores may have changed during the delay between the baseline assessment and receiving access to the OLT. Second, the results may have been impacted by the percentage of training completed, which was higher in ILT than TM. In addition, training completion was assessed via self-report in TM and OLT, but not in ILT. Third, the sample consisted of clinicians who volunteered to participate in this research and therefore may be particularly motivated to learn and use the treatments strategies being taught. Moreover, nearly half of the sample consisted of clinicians with less than a master’s degree, which, although representative of providers in community mental health settings, may have impacted the findings. Fourth, given the specificity of the outcomes to the content areas of chain analysis and validation strategies, no established measures existed that could be used in the present study. Thus, although all
measures demonstrated adequate reliability and/or face validity, they were developed or adapted for use in the present study and their broader psychometric qualities are unknown. Fifth, although the performance-based role plays were carefully developed and monitored to maximize their generalizability to real-world clinical interactions, they remain a proxy measure for what a clinician might do in actual clinical practice. Finally, an inherent limitation in this type of research is that the quality of the training depends not only on the training method and content, but on non-specific factors such as instructor charisma (ILT), writing quality (TM), and visual appeal (OLT). Thus, additional research examining multiple applications of these types of trainings methods is needed to reach broader conclusions.
Acknowledgements

This research, as well as the development of the online training courses, was funded by two Small Business Innovation Research (SBIR) grants from the National Institute on Drug Abuse (5R44DA016493-04 and 5R44DA018049-03) awarded to the first author. Portions of this research were presented at professional conferences: Dialectical Behavior Therapy Strategic Planning Meeting in Seattle, WA, and the 43rd and 44th Annual Conference of the Association for Behavioral and Cognitive Therapies (New York, NY and San Francisco, CA, respectively). We acknowledge and thank the following individuals: Cecilia Spitznas, Ph.D., NIDA Program Officer; Blair Beadnell, Ph.D., Statistical Consultant; Andrew Paves, primary research assistant who aided the first author in developing the text materials; Marsha Linehan, Ph.D., DBT Treatment Developer; Sharon Y. Manning, Ph.D., former BTECH Continuing Education Director for her oversight of content across training methods to ensure comparability; Brian Knudson and the NogginLabs, Inc. team, and Steve Lee and the Allen Interactions, Inc. team for working side-by-side our group in designing and building the online training courses; Kelly Koerner, Ph.D. for her work in designing and writing curriculum for early versions of the online training courses; our BTECH Research colleagues for their assistance with this research, including Angela Kelley-Brimer, Ignacio Contreras, Tim Kelly, Marni Levy, Maria Monroe-DeVita, Ph.D., Tiffany Mueller, and Brian Rash; as well as the many clinics who supported our subject recruitment efforts. Drs. Dimeff and Harned are compensated for providing training and consultation in DBT. Please address correspondence to Linda A. Dimeff, Ph.D., Evidence-Based Practice Institute, LLC, 814 29th Avenue, Seattle, WA 98122. Email: linda@ebpi.org. Phone: 206-384-7371.
References


Department of Veterans Affairs (2008). *Uniform mental health services in VA medical centers and clinics*. Veterans Health Administration Handbook 1160.01. Washington, DC.


Table 1

Sample Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>OLT (n=55)</th>
<th>ILT (n=55)</th>
<th>TM (n=62)</th>
<th>Full Sample (N=172)</th>
<th>Between-Condition Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age M (SD)</td>
<td>42.0 (12.6)</td>
<td>39.4 (11.7)</td>
<td>38.2 (11.5)</td>
<td>39.8 (11.9)</td>
<td>( F (2, 215) = 1.5, p=.22 )</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \chi^2 (2) = 4.4, p = .11 )</td>
</tr>
<tr>
<td>Female</td>
<td>76.4%</td>
<td>67.3%</td>
<td>83.9%</td>
<td>76.2%</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23.6%</td>
<td>32.7%</td>
<td>16.1%</td>
<td>23.8%</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \chi^2 (10) = 6.8, p = .74 )</td>
</tr>
<tr>
<td>Caucasian</td>
<td>76.4%</td>
<td>78.2%</td>
<td>80.6%</td>
<td>78.5%</td>
<td></td>
</tr>
<tr>
<td>Native American</td>
<td>1.8%</td>
<td>3.6%</td>
<td>0.0%</td>
<td>1.7%</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.8%</td>
<td>7.3%</td>
<td>3.2%</td>
<td>4.1%</td>
<td></td>
</tr>
<tr>
<td>Asian American</td>
<td>7.3%</td>
<td>3.6%</td>
<td>8.1%</td>
<td>6.4%</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.5%</td>
<td>1.8%</td>
<td>3.2%</td>
<td>3.5%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7.3%</td>
<td>5.5%</td>
<td>4.8%</td>
<td>5.8%</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \chi^2 (10) = 5.6, p = .85 )</td>
</tr>
<tr>
<td>High School</td>
<td>5.5%</td>
<td>5.5%</td>
<td>3.3%</td>
<td>4.7%</td>
<td></td>
</tr>
<tr>
<td>AA/AS</td>
<td>10.9%</td>
<td>7.3%</td>
<td>1.6%</td>
<td>6.4%</td>
<td></td>
</tr>
<tr>
<td>BA/BS</td>
<td>27.3%</td>
<td>32.7%</td>
<td>37.7%</td>
<td>32.7%</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>47.3%</td>
<td>43.6%</td>
<td>47.5%</td>
<td>46.2%</td>
<td></td>
</tr>
<tr>
<td>MD/PhD</td>
<td>5.5%</td>
<td>7.3%</td>
<td>6.6%</td>
<td>6.4%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3.6%</td>
<td>3.6%</td>
<td>3.3%</td>
<td>3.5%</td>
<td></td>
</tr>
<tr>
<td>Profession</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \chi^2 (16) = 15.9, p = .46 )</td>
</tr>
<tr>
<td>Psychologist</td>
<td>0.0%</td>
<td>5.5%</td>
<td>4.8%</td>
<td>3.5%</td>
<td></td>
</tr>
<tr>
<td>Psychiatrist</td>
<td>0.0%</td>
<td>1.8%</td>
<td>0.0%</td>
<td>0.6%</td>
<td></td>
</tr>
<tr>
<td>Psychiatric Nurse Practitioner</td>
<td>1.8%</td>
<td>1.8%</td>
<td>0.0%</td>
<td>1.2%</td>
<td></td>
</tr>
<tr>
<td>Chemical Dependency</td>
<td>25.5%</td>
<td>27.3%</td>
<td>12.9%</td>
<td>21.5%</td>
<td></td>
</tr>
</tbody>
</table>
Counselor
Social Worker (MSW) 16.4% 16.4% 16.1% 16.3%
Mental Health Counselor 21.8% 21.8% 19.4% 20.9%
(MA/MS/MFT)
Mental Health Counselor 3.6% 1.8% 4.8% 3.5%
(BA/BS)
Student 3.6% 9.1% 11.3% 8.1%
Other 27.3% 14.5% 30.6% 24.4%

**Work Setting**

<table>
<thead>
<tr>
<th>Work Setting</th>
<th>0.0%</th>
<th>1.8%</th>
<th>4.8%</th>
<th>2.3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient CMHC</td>
<td>45.5%</td>
<td>41.8%</td>
<td>41.9%</td>
<td>43.0%</td>
</tr>
<tr>
<td>Residential/Day Treatment</td>
<td>10.9%</td>
<td>7.3%</td>
<td>3.2%</td>
<td>7.0%</td>
</tr>
<tr>
<td>Facility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug Treatment Program</td>
<td>7.3%</td>
<td>7.3%</td>
<td>8.1%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Inpatient</td>
<td>7.3%</td>
<td>9.1%</td>
<td>8.1%</td>
<td>8.1%</td>
</tr>
<tr>
<td>Corrections Facility</td>
<td>3.6%</td>
<td>0.0%</td>
<td>9.7%</td>
<td>4.7%</td>
</tr>
<tr>
<td>School/Training</td>
<td>1.8%</td>
<td>3.6%</td>
<td>4.8%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Other</td>
<td>23.6%</td>
<td>29.1%</td>
<td>19.4%</td>
<td>23.8%</td>
</tr>
</tbody>
</table>

**Experience M (SD)**

<table>
<thead>
<tr>
<th>Months worked as a treatment provider</th>
<th>76.7 (78.6)</th>
<th>77.6 (80.2)</th>
<th>83.4 (99.9)</th>
<th>79.4 (86.7)</th>
</tr>
</thead>
</table>

\[ F (2, 745) = 0.1, \quad p = .91 \]

**Prior DBT Experience**

<table>
<thead>
<tr>
<th>Prior DBT Experience</th>
<th>84.6%</th>
<th>88.9%</th>
<th>83.9%</th>
<th>85.3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>None/minimal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some/moderate</td>
<td>15.4%</td>
<td>11.1%</td>
<td>16.1%</td>
<td>14.7%</td>
</tr>
</tbody>
</table>

\[ \chi^2 (2) = 0.6, \quad p = .72 \]

*Note:* Some categories do not total 100.0% because of rounding. Prior DBT experience is defined as: none/minimal (e.g., attended an in-service training, grand rounds, or lecture) or some/moderate (e.g., read part of the primary DBT manuals, attended a one- or two-day workshop). OLT = Online training. ILT = Instructor-led training. TM = Treatment manual.
Table 2

Means, Standard Deviations, and HLM Results for Longitudinal Outcomes: Total Scores

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Baseline</th>
<th>Post-Training</th>
<th>90 Day Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLT (SD) n</td>
<td>ILT (SD) n</td>
<td>TM (SD) n</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>2.24 (0.67)</td>
<td>2.54 (0.62)</td>
<td>2.13 (0.60)</td>
</tr>
<tr>
<td>Motivation</td>
<td>4.24 (0.50)</td>
<td>4.21 (0.48)</td>
<td>4.17 (0.50)</td>
</tr>
<tr>
<td>Knowledge</td>
<td>0.38 (0.08)</td>
<td>0.37 (0.08)</td>
<td>0.38 (0.09)</td>
</tr>
<tr>
<td>Proficiency</td>
<td>0.83 (0.62)</td>
<td>0.85 (0.55)</td>
<td>0.83 (0.59)</td>
</tr>
</tbody>
</table>

30 Day Follow-up

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Baseline</th>
<th>Post-Training</th>
<th>90 Day Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLT (SD) n</td>
<td>ILT (SD) n</td>
<td>TM (SD) n</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>2.40 (1.18)</td>
<td>2.16 (1.11)</td>
<td>2.73 (1.43)</td>
</tr>
</tbody>
</table>

Note: OLT = online training. ILT = instructor led training. TM = treatment manual.

*ab*In a row, when means share the same subscript this means that they are significantly different in HLM analyses (*p* < .05)
Table 3

**Cohen’s d Between-Condition Effect Sizes**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>OLT vs. TM</th>
<th></th>
<th>ILT vs. TM</th>
<th></th>
<th>OLT vs. ILT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>90 Day FU</td>
<td>Pre</td>
<td>Post</td>
<td>90 Day FU</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.17</td>
<td>0.15</td>
<td>0.07</td>
<td>0.67*</td>
<td>0.96*</td>
<td>0.84*</td>
</tr>
<tr>
<td>Chain analysis</td>
<td>0.07</td>
<td>0.17</td>
<td>0.02</td>
<td>0.47*</td>
<td>0.73*</td>
<td>0.77*</td>
</tr>
<tr>
<td>Validation</td>
<td>0.17</td>
<td>0.36</td>
<td>0.24</td>
<td>0.52*</td>
<td>0.92*</td>
<td>0.70*</td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.14</td>
<td>0.06</td>
<td>0.34</td>
<td>0.08</td>
<td>0.52*</td>
<td>0.67*</td>
</tr>
<tr>
<td>Chain analysis</td>
<td>0.07</td>
<td>0.14</td>
<td>0.34</td>
<td>0.00</td>
<td>0.59*</td>
<td>0.82*</td>
</tr>
<tr>
<td>Validation</td>
<td>0.05</td>
<td>0.25</td>
<td>0.47</td>
<td>0.12</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.00</td>
<td>0.62*</td>
<td>0.46*</td>
<td>0.12</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>Chain analysis</td>
<td>0.00</td>
<td>0.49</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Validation</td>
<td>0.00</td>
<td>0.58*</td>
<td>0.45*</td>
<td>0.00</td>
<td>0.20</td>
<td>0.09</td>
</tr>
<tr>
<td>Proficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.00</td>
<td>0.09</td>
<td>0.05</td>
<td>0.04</td>
<td>0.00</td>
<td>0.18</td>
</tr>
<tr>
<td>Chain analysis</td>
<td>0.28</td>
<td>0.18</td>
<td>0.06</td>
<td>0.16</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>Validation</td>
<td>0.23</td>
<td>0.06</td>
<td>0.05</td>
<td>0.09</td>
<td>0.10</td>
<td>0.31</td>
</tr>
</tbody>
</table>

*Note:* Raw data were used to calculate effect sizes. Interpretation of Cohen’s d: .20 (small), .50 (medium), and .80 (large) (Cohen, 1988). Pre = baseline. FU = follow-up. OLT = online training. ILT = instructor led training. TM = treatment manual.

* Between-condition differences are significant at p < .05 in the HLM analyses.
### Table 4

*Means, Standard Deviations, and Exploratory HLM Results for Longitudinal Outcomes: Content Domain Specific*

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Post-Training</th>
<th>90 Day Follow-up</th>
<th>30 Day Follow-up</th>
<th>60 Day Follow-up</th>
<th>90 Day Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLT</td>
<td>ILT</td>
<td>TM</td>
<td>OLT</td>
<td>ILT</td>
<td>TM</td>
</tr>
<tr>
<td><strong>Self-Efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chain analysis</td>
<td>1.92 (0.75)</td>
<td>2.24 (0.76)</td>
<td>1.90 (0.70)</td>
<td>2.79 (0.61)</td>
<td>3.43 (0.80)</td>
<td>2.90 (0.67)</td>
</tr>
<tr>
<td>Validation</td>
<td>2.28 (0.65)</td>
<td>2.52 (0.69)</td>
<td>2.17 (0.67)</td>
<td>3.16 (0.76)</td>
<td>3.55 (0.70)</td>
<td>2.91 (0.71)</td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chain analysis</td>
<td>3.95 (0.66)</td>
<td>3.90 (0.71)</td>
<td>3.90 (0.69)</td>
<td>3.84 (0.61)</td>
<td>4.14 (0.65)</td>
<td>3.75 (0.68)</td>
</tr>
<tr>
<td>Validation</td>
<td>4.20 (0.63)</td>
<td>4.25 (0.66)</td>
<td>4.17 (0.63)</td>
<td>4.09 (0.53)</td>
<td>4.26 (0.58)</td>
<td>3.93 (0.70)</td>
</tr>
<tr>
<td><strong>Knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chain analysis</td>
<td>0.47 (0.10)</td>
<td>0.47 (0.09)</td>
<td>0.47 (0.10)</td>
<td>0.67 (0.09)</td>
<td>0.62 (0.10)</td>
<td>0.62 (0.11)</td>
</tr>
<tr>
<td>Validation</td>
<td>0.24 (0.10)</td>
<td>0.24 (0.10)</td>
<td>0.24 (0.10)</td>
<td>0.59 (0.21)</td>
<td>0.51 (0.20)</td>
<td>0.47 (0.21)</td>
</tr>
<tr>
<td><strong>Proficiency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chain analysis</td>
<td>0.65 (0.62)</td>
<td>0.58 (0.63)</td>
<td>0.48 (0.60)</td>
<td>1.52 (0.77)</td>
<td>1.46 (0.73)</td>
<td>1.38 (0.80)</td>
</tr>
<tr>
<td>Validation</td>
<td>1.00 (0.79)</td>
<td>1.11 (0.71)</td>
<td>1.18 (0.79)</td>
<td>1.61 (0.72)</td>
<td>1.48 (0.68)</td>
<td>1.56 (0.90)</td>
</tr>
</tbody>
</table>

**Notes:**
- *p < 0.05, **p < 0.01, ***p < 0.001.
- Condition x Time: Interaction effect between Condition and Time.
Note: For outcomes where baseline scores were covaried (i.e., self-efficacy), between condition effects are represented by the Condition $F$ value. For all other outcomes, between condition effects are represented by the Time x Condition $F$ value. The number of participants who completed the post-training assessment varied somewhat by measure: OLT ($n = 31$-$35$), ILT ($n = 50$-$52$), TM ($n = 55$-$58$). The number of participants who completed the 90 day follow-up assessment varied somewhat by measure: OLT ($n = 28$-$33$), ILT ($n = 43$-$46$), TM ($n = 50$-$51$).

*p < .05, **p < .01, ***p < .001

$\text{ab}$ In a row, when means share the same subscript this means that they are significantly different ($p < .05$)
Figure Captions

Figure 1

*Participant flow from randomization through final follow-up.*
Figure 1.

```
Assessed and Eligible (N=200)
  ↓
Allocation
  ↓
Online Training (n=66)
  ↓
Completed Baseline (n=55)
  ↓
Received Training (n=35)
  ↓
Completed Post-training (n=35)
  ↓
Discontinued Participation (n=20)
  • Too Busy (n=11)
  • Unreachable (n=8)
  • Technical Issues (n=1)
Lost to Follow-up
  • 30 day follow-up (n=2)
  • 60 day follow-up (n=2)
  • 90 day follow-up (n=0)
Discontinued Participation (n=2)
  • Unknown (n=1)
  • Too Busy (n=1)
Analyzed (n=55)
  Excluded (HLM Analyses; n=20)
  Excluded (Other Analyses; n=0)

Instructor-Led Training (n=67)
  ↓
Completed Baseline (n=55)
  ↓
Received Training (n=52)
  ↓
Completed Post-training (n=52)
  ↓
Discontinued Participation (n=3)
  • Unreachable (n=2)
  • Illness (n=1)
Lost to Follow-up
  • 30 day follow-up (n=0)
  • 60 day follow-up (n=2)
  • 90 day follow-up (n=4)
Discontinued Participation (n=1)
  • Too Busy (n=1)
Analyzed (n=55)
  Excluded (HLM Analyses; n=3)
  Excluded (Other Analyses; n=0)

Treatment Manual (n=67)
  ↓
Completed Baseline (n=62)
  ↓
Received Training (n=58)
  ↓
Completed Post-training (n=58)
  ↓
Discontinued Participation (n=4)
  • Unreachable (n=4)
Lost to Follow-up
  • 30 day follow-up (n=1)
  • 60 day follow-up (n=1)
  • 90 day follow-up (n=6)
Discontinued Participation (n=0)
Analyzed (n=62)
  Excluded (HLM Analyses; n=4)
  Excluded (Other Analyses; n=0)
```
Highlights

- Clinicians were trained in two core strategies of Dialectical Behavior Therapy.
- A randomized controlled trial compared online training, a workshop, and a manual.
- The workshop was superior in terms of satisfaction and motivation.
- Online training was superior in increasing knowledge.
- The efficacy of the training methods varied depending on the strategy being taught.