12-2010

The Predictive Power of Therapeutic Alliance on Treatment Retention

James B. Anderson
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Clinical Psychology Commons, and the Counseling Psychology Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/497

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
THE PREDICTIVE POWER OF THERAPEUTIC ALLIANCE ON TREATMENT RETENTION

by

James B. Anderson

A Dissertation Submitted to the Faculty of The Graduate College in partial fulfillment of the requirements for the Degree of Doctor of Philosophy Department of Psychology Advisor: Scott T. Gaynor, Ph.D.

Western Michigan University Kalamazoo, Michigan December 2010
THE PREDICTIVE POWER OF THERAPEUTIC ALLIANCE ON TREATMENT RETENTION

James B. Anderson, Ph.D.
Western Michigan University, 2010

Therapeutic alliance has long been recognized as an important contributor to successful psychotherapy, and research has demonstrated that it has a moderate and reliable impact on treatment outcome across studies employing a variety of treatment strategies. The current study seeks to build on the alliance literature by examining the degree to which therapeutic alliance is predictive of treatment retention. Two trained raters assessed the alliance between therapists and their patients by watching and rating therapeutic alliance in videotaped psychotherapy sessions. They watched tapes of therapists using two different types of treatment: cognitive modification and supportive therapy. Alliance was assessed via the Working Alliance Inventory-Observer Form (WAI-O) and the Session Evaluation Questionnaire (SEQ). Results were analyzed using a series of 2 \times 2 ANOVAs based on treatment status (completer or dropout) and therapy type (cognitive modification or supportive therapy). While no statistically significant differences were detected on the WAI-O, data from a subscale of the SEQ indicated that treatment completers were significantly more aroused than dropouts. Data also indicated that SEQ total scores and multiple SEQ subscale scores were rated as significantly higher in supportive therapy sessions than in cognitive modification sessions, and that participants and therapists tended to appear to like each other more in supportive
therapy than in cognitive modification treatment. The present paper discusses potential implications for these findings, and provides suggestive evidence that the level a psychotherapy participant’s arousal may be predictive of treatment completion and premature dropout.
ACKNOWLEDGMENTS

First, I’d like to thank my advisor, Scott Gaynor. Without him, this project would have never gotten done. I’d also like to thank my research assistants, Emily Hill and James Mellor. Their tireless work coding videotapes made this project possible.

Additionally, I thank my committee (Richard Spates, Helen Pratt, and Cynthia Pietras) for their scholarly contributions to the project’s methodology and analysis of the results. Their involvement made this project much better. Finally, I’d like to thank my family, particularly my parents (John Anderson and Elizabeth Palmen), and my wife (Dikla Eckshtain). Their emotional support kept me going through the difficult times that occasionally arose throughout the process of this project.

James B. Anderson
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii

LIST OF TABLES .................................................................................................................. v

LIST OF FIGURES................................................................................................................ vi

INTRODUCTION ..................................................................................................................... 1
  Literature Review ............................................................................................................. 1
  Statement of Purpose ....................................................................................................... 12

METHODS ................................................................................................................................ 13
  Participants ...................................................................................................................... 13
  Design .............................................................................................................................. 15
  Measures .......................................................................................................................... 18
    Working Alliance Inventory-Observer Form ............................................................... 18
    Session Evaluation Questionnaire ............................................................................... 19
    Likability Ratings ........................................................................................................ 20
    General Procedure ...................................................................................................... 20

RESULTS .................................................................................................................................. 25
  Rater Agreement ............................................................................................................ 25
  Therapeutic Alliance and Treatment Status ................................................................. 27
  Therapeutic Alliance and Treatment Type .................................................................... 31
  Treatment by Status Interactions ................................................................................... 35
Table of Contents—Continued

Change in Alliance Between First and Last Sessions Rated.......................... 35
Binomial Analyses of Differences in Alliance by Status.............................. 37
DISCUSSION ........................................................................................................... 39
REFERENCES ........................................................................................................... 61
LIST OF TABLES

1. Demographic Information ........................................................................................................ 14
2. Selected Participants .................................................................................................................. 17
3. Rater Agreement ....................................................................................................................... 26
4. Statistically Significant Status and Treatment-Based Differences .................................... 36
5. Between-Group Binomial Analyses ...................................................................................... 38
6. Between-Individual Binomial Analyses .................................................................................. 40
LIST OF FIGURES

1. Participant Selection and Matching Process ......................................................... 16
INTRODUCTION

Literature Review

The importance of the relationship between therapist and client has long been recognized as a key to successful intervention. Carl Rogers advocated a humanistic, client-centered approach to therapy throughout his career that began in the middle of the 20th century. Central to Rogers’ approach was the creation of the relationship between therapist and client. According to Rogers (1946),

If the counselor can create a relationship permeated by warmth, understanding, safety from any type of attack, no matter how trivial, and basic acceptance of the person as he is, then the client will drop his natural defensiveness and use the situation. (p. 419)

Since Rogers’ work with client-centered therapy, other researchers have continued the empirical investigation of the importance of a strong relationship between therapist and client to therapy outcome. Horvath and Symonds (1991) reported that the strength of the working alliance between therapist and client is reliably correlated with treatment outcome. In a meta-analysis, these researchers reviewed 20 different data sets that included ratings of the relationship between therapist and client. The data sets included in the study came from therapists from various theoretical orientations, including psychodynamic, eclectic, cognitive-behavioral, and gestalt. The effect of therapeutic alliance was not dependent on the therapeutic approach taken by the therapist. Using the product-moment correlation coefficient $r$ as an estimate of effect size (ES), Horvath and Symonds report that the combined ES for the studies included in the meta-analysis was .26, which may be considered a moderate, consistent effect, suggesting that therapeutic
alliance is an important predictor of treatment outcome regardless of type of therapy. Interestingly, they found that the impressions of observers of therapy sessions (regarding strength of therapeutic alliance) were not as predictive of treatment outcome as the reports of therapists or clients.

Almost a decade after the study by Horvath and Symonds (1991), Martin, Garske, and Davis (2000) conducted a larger, updated meta-analysis, again examining the role of therapeutic alliance on treatment outcome. These researchers examined 79 studies that looked at the relationship between alliance and treatment outcome. They found an overall weighted alliance-outcome correlation of 0.22. Martin and colleagues argue that this correlation may be somewhat conservative, because when calculating the correlation, effect sizes reported as nonsignificant or simply not provided were coded as zero. Similar to the results of Horvath and Symonds (1991), Martin and colleagues concluded that alliance has a moderate and reliable effect on treatment outcome.

The results of the meta-analysis by Horvath and Symonds (1991) suggested that strength of therapeutic alliance has a reliable effect on treatment outcome regardless of psychotherapy type. Krupnick and colleagues (1996) added to this finding by demonstrating that therapeutic alliance is also important to pharmacotherapy outcome. These researchers evaluated the effect of therapeutic alliance for depressed patients receiving four different treatment types: interpersonal therapy, cognitive-behavioral therapy, imipramine with clinical management, and placebo with clinical management. These authors state that in the past, some researchers have argued that, although therapeutic alliance may play a role in patient compliance with doctors’ recommendations regarding pharmacotherapy, it is the medication itself that is responsible for observed
change. However, other researchers have suggested that the doctor-patient bond itself does account for some observed change in pharmacotherapy, both when an active drug is used as well as in the case of the use of a placebo (Downing & Rickels, 1978). The study sought to further investigate the degree to which therapeutic alliance was a factor in pharmacotherapy. Additionally, Krupnick et al. pointed out that because many of the therapists in the Horvath and Symonds meta-analysis identified the type of therapy they employed as “eclectic,” the results of the analysis did not fully address the question of whether different therapy types tend to more reliably produce strong therapeutic alliance than other treatment modalities. Krupnick and colleagues pointed out that, although there had been a number of studies at that time that investigated the role of alliance in psychodynamically-oriented therapies, far fewer studies had investigated the role of therapist-client bond in other orientations of treatment. Part of the goal of their study was to provide data regarding this issue.

Level of depression at end of treatment was evaluated by clinical raters using the Hamilton Rating Scale for Depression (HRSD) and by patients using the self-report Beck Depression Inventory (BDI). Both HRSD and BDI scores were correlated with ratings of alliance in all treatment types. Alliance was assessed using a modified version the Vanderbilt Therapeutic Alliance Scale (VTAS), which was completed by a trained observer watching the session via videotape. Raters scored level of alliance as between 0 (not at all) to 5 (a great deal). Results indicated that there were high mean levels of alliance across all conditions, and that alliance was related to treatment outcome in all conditions. Mean level of alliance was similar across groups, ranging from 3.6 (placebo group) to 3.9 (interpersonal therapy). These data suggest that treatment type does not
predict strength of therapeutic alliance. Using regression analysis, Krupnick et al. (1996) concluded that, across groups, alliance accounted for 19% of the variance observed in post-treatment HRSD scores and 18% of the variance in post-treatment BDI scores ($p < 0.001$), indicating a significant and robust effect of therapeutic alliance on subsequent ratings of depression. Interestingly, results of this study seem to contradict the findings of Horvath and Symonds (1991) meta-analysis, which concluded that observers’ impressions of therapeutic alliance were not especially predictive of treatment outcome. Recently, a study examining predictors of therapeutic alliance in group therapy for individuals experiencing auditory hallucinations investigated whether there was a difference between Cognitive-Behavioral Therapy (CBT) and Supportive Therapy (ST) groups in terms of strength of alliance (Johnson, Penn, Bauer, Meyer, & Evans, 2008). Similar to the results of Krupnick et al., Johnson and colleagues found that there was no main effect between therapy type and alliance ($t(8) = 0.15, p = .89$).

Some researchers take a more cautious approach in their commentary regarding therapeutic alliance. Siev, Huppert, and Chambless (2009) caution against concluding that, because therapeutic alliance has a moderate and reliable effect across treatment techniques and disorders, all treatment techniques are equally efficacious to each other. These authors argue that treatment technique is at least as significant as therapeutic alliance in accounting for change that occurs over the course of therapy, and outline several studies that indicate the appropriate treatment technique can make a significant difference in the degree to which a patient improves. They point out that improvement in therapeutic alliance in CBT protocols treating depression often occurs after the patient has already experienced significant improvement in symptoms. Tang and DeRubeis
(1999) also note that improvement in therapeutic alliance occurs early on in CBT for depression, specifically after what they term "sudden gains" that occur early in treatment. In their model, Tang and DeRubeis hypothesize that the improved alliance that occurs in turn leads to improved treatment efficacy, in essence arguing that the therapeutic alliance that emerges after the early sudden gain acts as a mediator for future continued improvement. Despite their caution that data indicating that therapeutic alliance contributes to treatment outcome does not mean that treatment technique is irrelevant, Siev et al. also acknowledge that alliance can impact treatment outcome. They argue that it is most likely to contribute to treatment outcome if the therapist conducting treatment makes the development of therapeutic alliance a specific treatment focus.

Despite the fact that a number of studies have suggested that therapeutic alliance is predictive of therapy outcome, some remained unconvinced that alliance was responsible for observed change. Klein et al. (2003) and Crits-Cristoph, Connolly Gibbons, and Hearon (2006) pointed out that few previous studies investigating the effect of therapeutic alliance had addressed the possibility that change that appeared to be attributable to alliance may in fact be explained by an extraneous or alternative (i.e., third) variable. These authors pointed out that variables such as prior improvement of symptoms or particular patient characteristics often occur along with alliance, making it difficult to determine which factor is accounting for observed variance in treatment outcome. Because most studies examining the effect of therapeutic alliance on treatment outcome assess alliance sometime after therapy has already begun, it is possible that some change that appears to result from alliance is in fact a spurious result of change that has already begun due to some other factor.
In order to address this issue, Klein and colleagues (2003) assessed the results of a study that treated 455 participants diagnosed with Major Depressive Disorder. Participants were assigned to one of three groups: psychotherapy, medication (Nefazodone), or psychotherapy and medication combined. The psychotherapy used in the study was cognitive-behavioral analysis system of psychotherapy (CBASP), which Klein et al. (2003) identified as a structured, short-term therapy specifically designed for the treatment of depression. The primary outcome measure was the Hamilton Rating Scale for Depression (HRSD). The primary process measure was an abbreviated version of the Working Alliance Inventory (WAI). In an effort to control for extraneous variables such as patient characteristics and change on the primary outcome variable starting before the development of the therapeutic alliance, the researchers employed mixed effects-growth-curve models to examine the relationships between the therapeutic alliance and change in depressive symptoms. Using this method, the authors concluded that alliance did have a causal effect on treatment outcome after controlling for potential extraneous variables. This finding held across all groups within the study, although patients in the combined treatment group tended to report the strongest level of alliance with their therapist.

As illustrated above, there is substantial evidence that therapeutic alliance is reliably correlated to treatment outcome. With that being the case, it would be useful to have a clear definition of what is meant by the term, which has had a variety of summary descriptions over time (Andrusyna, Tang, DeRubeis, & Lubersky, 2001). As Martin et al. (2000) also point out, what is now commonly referred to as a singular construct has actually emerged from multiple understandings of the underpinnings of the relationship
between patient and therapist. These authors identify a sense of a collaborative partnership within the therapy, the emotional connection between therapist and patient, and mutual goals for treatment as core characteristics of the construct of therapeutic alliance. It has also been defined as “a summary term referring to a number of interpersonal processes at play in psychological treatment which can generally be considered to act in parallel to (and theoretically independently of) specific manualized treatment techniques” (Elvins & Green, 2008, p. 1168). Callaghan, Naugle, and Follette (1996) offer a radically behavioral definition of the therapeutic relationship and its role in eliciting client change. These authors suggest that, from the perspective of a Functional Analytic Psychotherapist (FAP), the relationship between the therapist and client should be characterized by intimacy and caring, but not unconditional positive regard. Rather, they suggest the role of the therapist is to differentially reinforce behavior that occurs within the therapy room that is consistent by the type of behavior change the patient wishes to make. In this paradigm, the authors appear to suggest that the relationship serves as a mediator for change, by creating an environment in which the patient is willing to emit clinically relevant behaviors and establishing the responses of the therapist as reinforcing or punishing of clinically-relevant client behaviors. Bordin (1979) provided a description of therapeutic alliance that appears to have demonstrated significant influence on subsequent study of alliance. This author proposed that alliance is the combination of the agreement between therapist and patient on therapeutic goals, the agreement on therapeutic tasks to reach these goals, and the interpersonal bond between the therapist and patient.
A 2007 study (Iacoviello et al.) provides evidence that seems to suggest that agreement between patient and treatment provider on treatment tasks may influence alliance. Participants diagnosed with Major Depressive Disorder were randomized into one of three experimental conditions: psychotherapy (supportive-expressive therapy), pharmacotherapy (sertraline), or pharmacotherapy with pill placebo. Before being randomized into a group, participants were asked whether they would prefer treatment via talk therapy or medication. Iacoviello and colleagues found that patients who preferred and received talk therapy indicated a significant improvement in level of therapeutic alliance (as assessed by patient report on the California Psychotherapy Alliance Scale [CALPAS]), with $t(2) = 2.11, p < .04$. Patients who indicated that they preferred talk therapy but were randomized to pharmacotherapy with either sertraline or placebo indicated decreases in alliance over the 9-week treatment period. The decrease in rating of alliance was not significant for those receiving sertraline ($t(2) = 1.46, p < .15$), but was for those receiving placebo ($t(2) = 3.16, p < .002$). There were no significant differences in change in alliance over the course of treatment between treatment groups for participants who initially indicated a preference for pharmacotherapy therapy. These results suggest that, at least under some circumstances, match between a client’s preference for mode of treatment and the type of treatment provided (i.e., agreement on treatment “tasks”) can influence the development of therapeutic alliance.

While each of the descriptions above provides acceptable, working descriptive summary labels for the concept of alliance, it is evident that uncertainty remains regarding a truly unified definition of alliance. One piece of data that supports this claim is the relatively large number of scales currently in use that are designed to measure some
aspect or form of working alliance (Elvins & Green, 2008). These authors identified no less than 32 measures that are designed to assess the strength of some aspect or conceptualization of therapeutic alliance. This high number of scales, and the lack of agreement across scales regarding dimensions of alliance to be measured, suggests that despite the strong basis of empirical support for the importance of therapeutic alliance to treatment outcome, there is still work to do in developing a consensus on the definition of the construct.

The lack of a clear consensus regarding the precise definition of Therapeutic Alliance invites theoretical speculation regarding the concept and its mechanism of action. Elvins and Green (2008) postulate that alliance can have effect on treatment outcome by working in parallel to a particular treatment technique. Others have expanded on this idea and speculated that therapeutic alliance may set the foundation for effective therapy, or perhaps even be the very essence of therapy that leads to effective outcome (Priebe & McCabe, 2008). This suggestion is unique from and perhaps more bold than that of Elvins and Green, in that it suggests that a strong therapeutic alliance is necessary for effective therapy, and acts as a mediator for desired outcome. This is distinct from the assertion of Elvins and Green, whose definition seems to imply an additive effect of alliance to treatment, but does not go so far as to identify alliance as essential for desired treatment outcome.

Despite the lack of agreement among scholars on a tight operational definition of the concept of therapeutic alliance, available evidence strongly indicate that a successful relationship between therapist and client will, at the very least, increase the chances that intervention is successful. Conversely, a poor therapeutic alliance would seem likely to
impede the progress of therapy. In fact, if alliance is particularly bad, one might expect that a patient might choose to avoid or stop therapy altogether. According to one study, psychiatric patients rated their subjective satisfaction of their relationship with their clinician as the most important factor in determining whether they were receiving good services (Johansson & Eklund, 2003). The study included both inpatient and outpatient participants with a variety of diagnoses, including mood disorders, stress-related disorders, and personality disorders. All had been treated prior to the study, and had received treatment type of their clinicians’ discretion. Therapies used included pharmacological, supportive therapy, and social skills training. Seven former outpatients and nine former inpatients participated in the study. The authors concluded that participants from both groups rated the relationship between care provider and patient as the most important variable in treatment.

There are some significant limitations to the study by Johansson and Eklund (2003). All participants from both groups were no longer in treatment. Those who had re-entered treatment were excluded. Thus, there is strong probability of a selection bias. It is not surprising that former patients whose symptoms had significantly improved would retroactively rate their relationship with their doctor or therapist as positive. However, asking these questions after the treatment had successfully ended means that it is difficult to determine whether the patients’ feelings of a strong relationship with their therapist led to improvement in their symptoms, or vice versa. Additionally, the authors selected a purely qualitative design, evaluating participants’ impressions of their treatment experience via 2-hour, open-ended interviews. While such a method allows for in-depth
Despite the significant limitations of the Johansson and Eklund (2003) study, their results do suggest that, from a patient’s perspective, the relationship between the individual receiving treatment and his or her therapist is very important. While the data from their study do not allow for strong conclusions regarding the direct effect of the therapeutic relationship on treatment outcome, their conclusions do suggest that patients are much more satisfied with treatment from a provider with whom they feel a strong bond. Regardless of the direct effect that this satisfaction has on treatment outcome, it does seem likely to have an effect on treatment participation. As Priebe and McCabe (2008) point out, if the connection between therapist and client is sufficiently unsatisfactory, the patient may simply choose to avoid treatment altogether. Thus, regardless of the direct influence of the therapeutic relationship on outcome, if it predicts treatment participation, it is certainly an important variable to investigate.

The current study is designed to empirically investigate whether poor therapeutic alliance predicts early, unplanned, and abrupt cessation of treatment initiated by the patient. Unplanned, abrupt cessation of psychotherapy initiated unilaterally by the patient has long been recognized as a significant problem. There is room for an updated meta-analysis of the issue, but judging from the most recent published, peer-reviewed meta-analytic data available, it is estimated that approximately 47% of psychotherapy patients drop out of treatment (Wierzbicki & Pekarik, 1993). This number was based on the authors’ review of 125 studies conducted across a wide range of settings, with a variety of populations and using a number of different treatment techniques. Based on a review of
the literature, it appears that the relationship between therapeutic alliance and treatment dropout is an under-studied aspect of the role of alliance in treatment. Only one published study (Robbins, Turner, Alexander, & Perez, 2003) directly addressing the influence of treatment alliance on treatment discontinuation was discovered. This study investigated whether alliance was predictive of dropout in family therapy for adolescents with externalizing behavior problems. The authors reported that, perhaps surprisingly, individual levels of alliance between parents and therapist and between adolescent and therapist (as estimated by raters via videotape) were not predictive of dropouts. However, unbalanced alliances (in which parents’ alliance with the therapist is rated as significantly higher than the adolescent’s) was predictive of dropouts. As Robbins et al. point out, this was one of few studies to examine the role of therapeutic alliance in family therapy, and it is unclear to what extent findings regarding the role of alliance in family therapy is generalizable to its role in individual therapy, and vice versa. Based on the fact that Robbins et al. demonstrated that alliance as measured at the systemic level that was the focus of treatment (i.e., the family), and the wealth of literature demonstrating the importance of therapeutic alliance on treatment outcome in individual therapy, it seems quite plausible that therapeutic alliance will be predictive of treatment retention in individual therapy.

**Statement of Purpose**

The current study was designed to empirically investigate whether poor therapeutic alliance predicts early, unplanned, and abrupt cessation of treatment initiated by the patient. Specifically, we investigated whether there are significantly different levels of therapeutic alliance between participants who drop out of therapy and those who
complete therapy (i.e., treatment status effect). Secondarily, the study investigated whether stronger therapeutic alliance is associated with a particular treatment type. In addition to examining whether or not there were differences in ending levels of alliance, the study also assessed whether there were differences between the change in level of therapeutic alliance across time spent in the study, based on either treatment status or treatment type. Finally, the study examined whether a particular treatment type moderates the degree to which therapeutic alliance predicts dropping out of treatment, resulting in a treatment by status interaction effect. The study tested the following hypotheses:

1. Participants who are rated as having poor therapeutic alliance with their therapists will be more likely to drop out of treatment abruptly and before planned than participants rated as having strong therapeutic alliance with their therapist. It is also predicted that sessions preceding discontinuation of treatment will be rated as less deep and less smooth than comparable sessions from treatment completers.

2. Supportive therapy and cognitive modification will lead to equal levels of therapeutic alliance.

Therapeutic alliance scores will be equally predictive of dropouts regardless of treatment type, meaning treatment type will not act as a moderating variable

METHODS

Participants

Participants for the current study were from a previous study investigating the efficacy of two different types of treatment for individuals with low self-esteem (Clore & Gaynor, 2009). In the Clore and Gaynor study, 57 university students with pre-treatment
self-esteem and general distress scores at least one standard deviation from the normative mean were randomized to six sessions of either cognitive modification (CM) or supportive therapy (ST). The current study included 17 participants who dropped out of the Clore and Gaynor study, as well as 14 who completed that study. As illustrated in Table 1, dropouts were similar to completers on demographic variables.

Table 1

Demographic Information

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dropouts ($n = 17$)</th>
<th>Completers ($n = 14$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>%</td>
</tr>
<tr>
<td>Gender (Female)</td>
<td>12</td>
<td>71</td>
</tr>
<tr>
<td>Caucasian</td>
<td>11</td>
<td>65</td>
</tr>
<tr>
<td>Average age</td>
<td>21.65</td>
<td>n/a</td>
</tr>
<tr>
<td>History of mental health treatment</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>Currently taking psychotropic medication</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>History of medication</td>
<td>6</td>
<td>35</td>
</tr>
</tbody>
</table>

Participants and completers were similar on pre-treatment Rosenberg Self-Esteem Scale (RSES) scores, $F(1, 55) = .30, p = .59$, and on Brief Symptom Inventory (BSI) scores, $F(1, 55) = .04, p = .84$. There was a mean group difference between dropouts and completers on Beck Depression Inventory (BDI) scores that approached statistical significance, $F(1, 55) = 3.00, p = .09$, with dropouts demonstrating slightly higher and less variable scores ($M = 26.65, SD = 7.02$) than completers ($M = 21.93, SD = 10.25$) on this measure. Participants for the current study did not attend any new sessions, as the
strength of the therapeutic alliance between participant and therapist was evaluated via observation of videotapes of past sessions conducted in the randomized clinical trial.

**Design**

First, videotapes from appropriate sessions from the Clore and Gaynor (2009) study described above were identified. That study used five different therapists, and two different treatment conditions (CM and ST). Four of the five therapists (therapists 2–5) from the Clore and Gaynor study had dropouts, and those therapists' participants were reviewed for potential inclusion in the present study. Selected videotaped sessions were matched based on therapist and, when possible, on treatment type as well. Starting with therapist 2, participants with this therapist who dropped out of the study were identified. We then determined how many completers this therapist had. This process was repeated for therapists 3–5. Next, as many of the dropouts as possible were matched with a comparable completer. All completers that are included in the study were matched with a dropout based on therapist and the session number that was viewed and rated. All included completers were also matched to a dropout based on treatment type, with one exception. One included completer was matched to a dropout who had a different treatment type because there were no available completers with the same treatment type for that therapist. When possible, participants were also matched based on sex and age. Figure 1 illustrates the process of selecting and pairing participants.

After identifying all participants who dropped out and selecting appropriate, matched completers, the primary investigator identified each dropout’s final session and selected the videotape of that session for review. The videotapes of the corresponding sessions for matched completers were also selected. For example, if a participant who
Figure 1. Participant Selection and Matching Process

dropped out of ST following his third session was identified, we identified a ST, completer treated by the same therapist, and evaluated alliance based on session three. By the end of this process, all dropout participants for all four therapists were identified and, when possible, each was matched with a similar completer as described above. A total of 17 total dropouts were selected, along with 14 completers that were appropriate for comparison. Table 2 depicts the participants selected for the present study.

In addition to identifying final sessions before dropout and corresponding sessions for matched completers as described above, we also watched and rated the first session
Table 2

*Selected Participants*

<table>
<thead>
<tr>
<th>Therapist</th>
<th>Tx Type</th>
<th>Sex</th>
<th>Age</th>
<th>Therapist</th>
<th>Tx Type</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>CM</td>
<td>Male</td>
<td>21</td>
<td>2</td>
<td>CM</td>
<td>Male</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>ST</td>
<td>Female</td>
<td>21</td>
<td>2</td>
<td>ST</td>
<td>Female</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>CM</td>
<td>Female</td>
<td>21</td>
<td>2</td>
<td>CM</td>
<td>Female</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>CM</td>
<td>Female</td>
<td>20</td>
<td>3</td>
<td>CM</td>
<td>Female</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>CM</td>
<td>Female</td>
<td>20</td>
<td>3</td>
<td>CM</td>
<td>Female</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>ST</td>
<td>Female</td>
<td>18</td>
<td>3</td>
<td>ST</td>
<td>Female</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>CM</td>
<td>Male</td>
<td>23</td>
<td>3</td>
<td>CM</td>
<td>Male</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>CM</td>
<td>Female</td>
<td>22</td>
<td>3</td>
<td>CM</td>
<td>Female</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>ST</td>
<td>Male</td>
<td>27</td>
<td>3</td>
<td>ST</td>
<td>Female</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>ST</td>
<td>Male</td>
<td>26</td>
<td>3</td>
<td>ST</td>
<td>Male</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>ST</td>
<td>Female</td>
<td>20</td>
<td>5</td>
<td>CM</td>
<td>Female</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>ST</td>
<td>Female</td>
<td>22</td>
<td>4</td>
<td>ST</td>
<td>Male</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>CM</td>
<td>Female</td>
<td>20</td>
<td>4</td>
<td>CM</td>
<td>Female</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>ST</td>
<td>Female</td>
<td>21</td>
<td>4</td>
<td>ST</td>
<td>Female</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>CM</td>
<td>Female</td>
<td>18</td>
<td>4</td>
<td>CM</td>
<td>Female</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>CM</td>
<td>Male</td>
<td>26</td>
<td>4</td>
<td>CM</td>
<td>Female</td>
<td>19</td>
</tr>
</tbody>
</table>

*Totals*:

CM=9  M=5  AVG=21.65  
ST=8  F=12  

*Totals*:

CM=9  M=4  AVG=21.79  
ST=5  F=10  

for each participant selected for the study. For those participants for whom their first
session was not also their last (or matched to a dropout who had quit after one session, in
the case of the completers), watching the first session in addition to a subsequent session
allowed us to assess whether or not there were significant differences in change in
alliance over the course on treatment based on either treatment status or treatment type.

Measures

Working Alliance Inventory-Observer Form

The Working Alliance Inventory-Observer Form (WAI-O; Horvath & Greenberg,
1989) is an assessment designed to measure the therapeutic alliance between a therapist
and client. It contains 36-items asking raters to evaluate three aspects of alliance: tasks
(efficacy of work within session); goals (relevance and importance of the aims of
treatment; and bond (mutual connection between patient and therapist, including trust,
acceptance, and confidence). The authors originally developed both client and clinician
versions of the form. It has since been modified to produce an observer form (WAI-O;
Tichenor & Hill, 1989). The WAI has good internal consistency, with alphas of 0.87,
0.82, and 0.68 for the goals, tasks, and bonds subscales, respectively (Horvath &
Greenberg, 1989). The WAI-O has demonstrated good construct validity, given that it
correlates strongly with the Vanderbilt Therapeutic Alliance Scale (VTAS; 0.84),
California Psychotherapy Alliance Scales (CALPAS; 0.82), and the Penn Helping
Alliance Rating Scale (Penn; 0.71; Tichenor & Hill, 1989). Each of these measures has
been independently evaluated and shown good reliability. Among 32 evaluated measures
of alliance, the WAI has been identified as having strong and thoroughly evaluated
construct validity (Elvins & Green, 2008). Following a large, meta-analytic review,
Martin et al. (2000) recommend the WAI as the preferred alliance measure for research protocols. These authors state that it is appropriate for use with all types of therapy. The WAI-O was completed following observation of each session. Each item is rated between 1 and 7. When scored, the WAI-O produces a total score as well as scores on all three subscales.

**Session Evaluation Questionnaire**

The Session Evaluation Questionnaire (SEQ; Stiles, 1980) allows raters to evaluate the depth (i.e., worth and strength) and smoothness (i.e., comfort) of a therapy session, as well as impressions of participants’ level of arousal (i.e., alertness and energy), and mood. (Stiles, Gordon, & Lani, 2002). It has been revised several times since its initial publication in order to increase internal consistency and ensure robust, internally consistent sets of items. The most recent iteration of the measure is SEQ-Form 5 (Stiles, Gordon, & Lani, 2002). Raters score each item between 1 and 7. The SEQ is designed to be completed by either a therapist or a client. For the current study, observers completed the measure based on the perspective of the participant that they observed. The wording of the introduction to the second portion of the measure, which contains items that load onto the Positive Mood and Arousal subscales, was altered to read “right now the participant feels.” The measure has demonstrated good-to-excellent internal consistency for both the smoothness, depth, participant mood, and participant arousal dimensions, with alpha coefficients of 0.93, 0.90, 0.90, and 0.81, respectively (Reynolds et al., 1996). Raters completed this measure following the observation of each session.
Likability Ratings

In addition to completing the formal measures described above, the raters made a global assessment of likability of both the therapist and participant following each session viewed. They answered the questions “Did the client like the therapist?” and “Did the therapist like the client?” for each session viewed. Ratings were based on a 4-point scale, with a “1” corresponding to a response of “not at all,” and a “4” corresponding to a response of “very much.”

General Procedure

Setting. All sessions took place in therapy rooms at Western Michigan University (WMU). These rooms are small, painted rooms with a table, two chairs, and minimal decorations. The raters evaluated videotapes by watching them on a VCR in the research laboratory of the primary investigator. This is a quiet space that allowed for minimal distraction for the raters. All videotapes were recorded from cameras mounted on the wall of the therapy rooms. The perspective on the tapes showed both the participant and the therapist in profile, from a slightly elevated angle. Two different rooms were used for the sessions, so in some sessions the camera was mounted on the west side of the room aimed at a dyad sitting on the east side of the room, while the orientation was vice versa for other dyads.

Raters. Two undergraduate research assistants (RAs) were selected to watch selected videotapes and rate the therapeutic alliance between therapist and participant using both the WAI-O and the SEQ.

Training. Before evaluating the videotaped sessions, the two raters underwent a training procedure to help them get familiar with the two rating measures and how they
were to be used to evaluate sessions. To begin, they watched two videotapes of master therapists (Carl Rogers and Aaron Beck) conducting therapy sessions. These sessions were selected because of the widely recognized expertise of the two therapists, and also because the type of therapy depicted in the tapes approximates the types of therapy on the tapes the raters were to evaluate (Rogers approximates supportive therapy; Beck approximates cognitive modification). The raters watched these tapes with the student investigator, taking time as needed to discuss what was occurring in the session on tape and beginning to become familiar with signs of therapeutic alliance in therapy. In doing so, the student investigator emphasized the differences in the approaches of Rogers and Beck, and how the two of them may be seen as attempting to create therapeutic alliance in different ways.

Next, the raters watched tapes from therapist 1 from the Clore and Gaynor (2009) study. This therapist had 17 completers and was used as a model for therapy as would be seen on the tapes included in the study. This therapist was chosen as a training model for two main reasons. First, none of her participants dropped out, so there were no dropouts to match with her completers using the procedure described above. Second, because she had a relatively large number of participants who completed, there was a large pool of available potential training sessions for the raters to view and on which they could practice. To start, the raters watched four tapes, and then completed the WAI-O, the SEQ, and likability ratings based on those sessions. The student investigator then reviewed the responses of both raters in order to determine the level of agreement on the WAI-O and SEQ. Consistent with procedures reported by Tichenor and Hill (1989), the aim was for the responses of the raters to show agreement of $r \geq .70$, using a Pearson’s product-
moment correlation. Following the initial review of the four tapes, agreement for between raters was .88 and .64 using Pearson’s $r$, and .87 and .61 using an intraclass correlation coefficient, on the WAI-O and SEQ, respectively.

Because agreement on the SEQ was short of the desired .70 level, a second training day was conducted. This session started by reviewing two additional sessions conducted by therapist 1 from the Clore and Gaynor (2009) with both raters and the student investigator present. Raters were asked to make ratings independently. Following this, the student investigator reviewed each of the raters’ ratings before continuing. When differences of 2 points or more on the SEQ were observed, the investigator brought such instances to the raters’ attention, and asked each of them to explain the rationale for the decision he or she made. The goal was to help raters get a general idea of how the other was looking at occurrences on the tapes, and to attempt to focus them on the same things in making scoring decisions. This procedure is similar to that used by Tichenor and Hill (1989) when they were doing the training for their study of alliance (V. Tichenor, personal communication, June 16, 2009). After viewing two sessions in this way, the raters viewed two more tapes, this time making their ratings independently and without discussion after their decisions had been made. Average agreement on the SEQ for these two sessions was .71 using Pearson’s $r$, meeting the a priori standard established.

However, a great deal of variability across the two sessions was observed, with agreement on one participant equal to .93 and agreement on the other equal to .48. For both sessions, agreement on the first half of the SEQ was over .90. In fact, agreement on the first half of the SEQ ranged between .65–.98 for all six of the sessions rated independently and without subsequent discussion and comparison across the two training sessions, with only
one of the six below .70. All the items that make up the Depth and Smoothness subscales are included in the first half of the SEQ. Because at the time these data were collected, it was on these subscales that we were expecting to see differences, this degree of consistent good agreement on the first half of the measure was considered sufficient to move on to rating the sessions to be included in the study.

**Evaluation of Therapeutic Alliance.** Following the training, the raters viewed the videotapes of selected sessions and rated them using the WAI-O and the SEQ. The raters were blind as to whether the participant they were viewing dropped out of therapy or completed therapy. Participants were not asked to rate more than five tapes in one day in an effort to avoid fatigue and maintain the fidelity of ratings. After the first week of ratings (19 sessions), agreement to that point was assessed. It was observed that agreement on the WAI-O remained very good (i.e., over .85). It was also observed that agreement on the SEQ was inconsistent. While it was over .70 for most participants, it was also below .50 for six, even when only considering agreement on the first half of the measure. As planned, a mid-treatment re-training session was conducted. In this meeting, two additional sessions by therapist 1 from the Clore and Gaynor (2009) were viewed, rated, and discussed, using the same training procedure described above. Additionally, following the training, one of the sessions that had been rated as part of the study was re-watched and rated. The remaining five sessions on which the raters had poor agreement were put back into the remaining tapes to be rated in random order. For balance, six other sessions that either had not yet been rated or had been rated with good agreement were also selected for re-rating. The raters were told that some sessions had been selected for re-rating, but were not told why or on what basis they were selected.
After all selected sessions were rated, we evaluated whether there was a difference in level of therapeutic alliance on both the WAI-O and the SEQ between participants who dropped out and those who completed treatment. We also examined whether there was a difference of the level of therapeutic alliance between participants who received cognitive modification therapy compared to those who received supportive therapy. Finally, we examined whether treatment type moderated the effect of therapeutic alliance. Level of agreement between raters on both measures was assessed, again using both Pearson’s $r$ and intraclass correlation coefficient. Other data were analyzed using a $2 \times 2$ ANOVA. Fixed factors were participant status (completer, dropout) and treatment type (cognitive modification, supportive therapy). Dependent variables were WAI-O and SEQ total scores, as well as the subscale scores on the respective measures, and the half-scale total scores (i.e., first half total and second half total) on the SEQ. Additionally, likability ratings of both therapist and client were included in ANOVAs.

After noting some interesting results (see below) on the status factor across all dependent variables, we conducted additional binomial analyses in an attempt to make sense of the data. We conducted both between-individual and between-group binomial analyses. For the between-individual analysis, we compared each dropout’s score on a dependent variable (e.g., WAI-O total score) with the completer with whom he or she had been matched. For each comparison, we made a binomial comparison of whether the dropout or completer had a higher alliance score. We completed such analyses for each matched pair on every included dependent variable. We made similar comparisons at the group level, making a rating on each included dependent variable of whether the dropout group or the completer group had a higher average alliance score.
There were a large number of potential dependent variables for analysis upon completion of data collection. Counting the total and subscale scores on the WAI-O and the SEQ as well as the likability scales, there were 11 dependent variables. Additionally, we had individual ratings from each rater and composite ratings based on the average of the two raters' ratings, bringing the potential DVs for analysis from 11 to 33. Further, because of the re-rating of select tapes described above, we had the option of using original or best-agreement ratings for each of these methods of looking at the data, doubling the number of potential DVs for analysis from 33 to 66. With all of these potential data sets, we felt it necessary to choose one of these for primary analyses. We settled on the composite scores (i.e., the average of both raters' data) from the ratings that yielded best agreement on the more established tools (i.e., the WAI-O and the SEQ). These scores had equal input from each rater, and because they were from the sessions on which our raters demonstrated the best agreement, they provided the least amount of difference between the raters' scores available. This decision left nine primary dependent variables for analysis. To account for the large number of dependent variables, statistical significance was considered both at the traditional, non-adjusted $p \leq .05$, and a Bonferroni-corrected level of $.006 (.05/9)$.

RESULTS

Rater Agreement

Rater agreement was calculated for the WAI-O and the SEQ. Additionally, agreement was calculated on the raters' likability ratings. In assessing agreement, when there was a session that was rated twice, the rating that produced the best agreement for that session was used in the calculation of overall agreement. Table 3 below summarizes
overall agreement for all sessions used in the study, using both Pearson’s $r$ and intraclass correlation coefficient (ICC). All statistics reported below were significant at the $p = .000$ level.

Table 3

*Rater Agreement*

<table>
<thead>
<tr>
<th></th>
<th>WAI-O</th>
<th>SEQ Total</th>
<th>SEQ 1st Half</th>
<th>SEQ 2nd Half</th>
<th>Likability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s $r$</td>
<td>.90</td>
<td>.59</td>
<td>.61</td>
<td>.58</td>
<td>.64</td>
</tr>
<tr>
<td>ICC</td>
<td>.90</td>
<td>.58</td>
<td>.61</td>
<td>.57</td>
<td>.64</td>
</tr>
</tbody>
</table>

*Note.* 1st half SEQ contains all items that make up Depth and Smoothness subscales; 2nd half SEQ contains all items that make up participant Mood and Arousal subscales.

After looking at the level of rater agreement, we assessed whether there were differences in levels of therapeutic alliance between treatment dropouts and completers (participant status), as well as between participants who received cognitive modification treatment and those who received supportive therapy (treatment type). We also examined whether there was an interaction effect between treatment status and treatment type. These queries were assessed via a series of $2 \times 2$ ANOVAs, in which treatment status and treatment type were always used as the two fixed factors, and the various measures of alliance (WAI-O total and subscores, SEQ total and subscores, etc.) were used as the dependent variable.
Therapeutic Alliance and Treatment Status

The results described below are based on analyses on the ratings on which rater 1 and rater 2 demonstrated the best agreement, unless otherwise noted. While completers consistently had slightly higher scores on total score and subscale scores, there were no significant differences in level of therapeutic alliance as measured on the WAI-O. When looking at overall WAI-O scores for the last session rated, the $2 \times 2$ ANOVA using the average ratings of raters 1 and 2 (completer $M = 206.39$, $SD = 14.28$; dropout $M = 199.26$, $SD = 24.96$) produced a treatment status $F = .828$ ($p = .37$). When individual ratings of either rater 1 or 2 were used, similar results were obtained, with neither indicating a significant status effect on WAI-O total scores but both slightly favoring strength of therapeutic alliance in the completer group over the dropout group. The composite ratings looking at the Task, Goals, and Bond subscales of the WAI-O also indicated nonsignificant differences between completers and dropouts (Task subscale $F = .608$, $p = .44$; Goals subscale $F = .671$, $p = .42$; Bond subscale $F = 1.177$, $p = .29$). As with the overall WAI-O score, individual ratings of both rater 1 and rater 2 also indicated insignificant differences between dropout and completer scores on all three of the WAI-O subscales, though all ratings indicated slightly higher scores for completers than dropouts.

Results on the SEQ regarding differences in therapeutic alliance between completers and dropouts more strongly suggest a status effect than those on the WAI-O. While still not significant, the difference between the composites of the raters total SEQ scores was closer to reaching the traditionally recognized $p = .05$ level of significance than any of the comparisons using the WAI-O total score and subscale scores. Mean total SEQ scores were higher for participants who completed treatment than for those who
dropped out (completer SEQ total score $M = 93.11$, $SD = 11.53$; dropout SEQ total score $M = 85.88$, $SD = 9.06$; $F = 2.54$, $p = .12$). Rater 1’s individual ratings actually indicated a significant difference between completers and dropouts on final SEQ total score ($F = 4.575$, $p = .04$), with completers rated as having a higher average SEQ total ($M = 93.50$, $SD = 10.95$) than dropouts ($M = 85.71$, $SD = 14.05$). Rater 2 also rated completers as having a higher average SEQ total score than dropouts on the last session rated for each ($M = 92.71$, $SD = 14.57$; $M = 86.06$, $SD = 19.39$, respectively). This difference is less than that observed between rater 1’s ratings, and with more variance as indicated by somewhat larger standard deviations. As a result, differences on total SEQ score as rated by rater 2 were nonsignificant ($F = 1.203$, $p = .28$).

Closer analysis of differences in SEQ scores between dropouts and completers revealed that the trends toward group differences on the total score were largely driven by ratings on the second half of the measure. Interestingly, the second half of the SEQ contains no items that load onto either the Depth or Smoothness SEQ subscales. Instead, they load onto subscales rating patient Mood and Arousal (Stiles et al., 1994; described in more detail below). The difference between the composite ratings of rater 1 and rater 2 for the first half of the SEQ for completers ($M = 50.36$, $SD = 5.70$) and dropouts ($M = 47.56$) was not significant ($F = 1.293$, $p = .266$). While both individual raters rated completers as having on average slightly higher scores on the first half of the SEQ than dropouts, these differences did not approach significance. On the Depth (completer $M = 4.30$, $SD = .76$; dropout $M = 4.22$, $SD = .90$), and Smoothness (completer $M = 4.86$, $SD = .75$; dropouts $M = 4.47$, $SD = 1.09$) subscales of the SEQ, differences in average ratings were also nonsignificant ($F = .277$, $p = .60$; $F = 1.27$, $p = .27$, respectively). The notably
smaller means represented in the subscale means compared to the first half means is a result of the method of calculating the scores. While ratings for individual items in the first half of the SEQ were simply summed to obtain the first half total, the Depth and Smoothness subscales are obtained by averaging the scores on the items that make up the scales (Stiles, Gordon, & Lani, 2002). Neither individual rater identified significant differences between dropouts and completers on the first half of the SEQ or either of the two subscales contained in the first half of the SEQ.

On the second half of the SEQ, the average of rater 1 and rater 2's ratings indicated a significant difference between completers ($M = 42.75, SD = 6.42$) and dropouts ($M = 37.71, SD = 7.40; F = 4.71, p = .039$). Items on the second half of this measure are designed to measure participants’ affect (Stiles et al., 1994). Specifically, the second half of the SEQ is designed to assess how positive and aroused the participant is feeling at the end of the therapy session. Both individual raters identified completers (rater 1, $M = 41.86, SD = 5.93$; rater 2, $M = 43.64, SD = 8.83$) as having a higher score than dropouts (rater 1, $M = 36.94, SD = 5.44$; rater 2, $M = 38.47, SD = 9.84$). The difference identified by rater 1 was highly statistically significant ($F = 10.21, p = .004$), while the difference identified by rater 2 was not ($F = 2.08, p = .16$).

Ratings on the second half of the SEQ can be further broken down into subscales that assess positive mood and arousal. Higher scores indicate more positive mood on the Mood subscale and increased arousal on the Arousal subscale. Examining results in this way indicated that differences on the second half of the SEQ were driven by differences on the Arousal subscale. Both Rater 1 and rater 2 rated completers as being more aroused than dropouts in their individual ratings. Rater 1 rated completers as having an average
Arousal score of $M = 3.97$ ($SD = 1.07$), and dropouts as having an average Arousal score of $M = 3.16$ ($SD = 0.77$). This difference was statistically significant ($F = 13.74, p = .001$). Rater 2 rated completers as having an average arousal score of $M = 4.34$ ($SD = 0.99$), and dropouts as having an average arousal score of $M = 3.64$ ($SD = 1.13$). This difference was also statistically significant ($F = 4.66, p = 0.04$). The composite average rating of rater 1 and rater 2 also rated completers ($M = 4.16, SD = 0.93$) as being more aroused on average than dropouts ($M = 3.40, SD = 0.91$), a difference that was also statistically significant ($F = 9.58, p = 0.005$). Both raters also rated completers as slightly higher than dropouts on the Mood subscale, but neither of them indicated statistically significant differences on this scale in their individual ratings nor the composite average rating (completers ($M = 4.39, SD = 0.79$, dropouts $M = 4.24, SD = 0.84$; $F = 0.10, p = 0.75$).

On the summary likability ratings, differences between completers and dropouts fell short of statistical significance. On the therapist likability item (a rating of 1–4 on the question “Does the client like the therapist?” with a 4 indicating that the participant liked the therapist “very much,” and a 1 indicated “not at all”), the average of rater 1 and rater 2’s ratings had completers ($M = 3.32, SD = .67$) as liking their therapist more on average than did dropouts ($M = 2.97, SD = .78$). However, this difference fell short of statistical significance ($F = 2.70, p = .11$). This same trend was observed when looking at rater 1 and rater 2’s ratings individually, in that both rated the completers as liking their therapist more on average than did dropouts, but with differences short of the $p = .05$ level of statistical significance. Average rater 1 and rater 2 ratings of how much the therapist liked the client (using the same scale described above) very slightly favored completers ($M =$
3.39, \(SD = .45\) over dropouts \((M = 3.29, SD = .56)\), but at a level that was nowhere close to statistically significant \((F = .56, p = .461)\).

**Therapeutic Alliance and Treatment Type**

There were no statistically significant differences in therapeutic alliance as measured by the WAI-O based on treatment type. Average scores from the raters yielded alliance means that were numerically slightly higher for participants who received ST \((M = 204.88, SD = 21.43)\) than for those who received CM \((M = 200.75, SD = 20.82)\). However, this difference did not approach statistical significance \((F = .347, p = .56)\).

Findings were similar on the Task \((F = .551, p = .46)\), Goals \((F = .061, p = .81)\), and Bond \((F = .520, p = .48)\) subscales of the WAI-O, with average scores slightly numerically higher for participants in the ST group in each instance.

Some clearly significant differences based on treatment type were detected on the SEQ. Average combined ratings of rater 1 and rater 2 indicated that participants in the ST group had a higher SEQ average total score \((M = 95.19, SD = 14.18)\) than those in the CM group \((M = 84.78, SD = 13.71)\), which was a statistically significant difference \((F = 4.80, p = .04)\). Both raters' individual ratings produced a higher average SEQ total score for participants in the ST group than for those in the CM group. Rater 1’s ST SEQ total score \((M = 95.54, SD = 12.15)\) was higher than his CM SEQ total score \((M = 84.67, SD = 12.18)\). This difference was statistically significant \((F = 7.74, p = .01)\). Rater 2’s ST SEQ total score was also higher \((M = 94.85, SD = 18.18)\) than his CM SEQ total score \((M = 84.89, SD = 16.09)\). Although the difference between rater 2’s ST and CM means is very similar to the difference between rater 1’s means, it fell short of statistical significance
(F = 2.63, p = .12), due in large part to the much larger degree of variability in the scores produced by rater 2.

There was a numerical difference between participants in the ST group (M = 51.96, SD = 7.95) and those in the CM group (M = 46.56) according to the combined average ratings of rater 1 and rater 2 on the first half SEQ score. This difference fell just short of the traditionally accepted .05 level of statistical significance (F = 3.91, p = .058). As noted above, all items for the Depth and Smoothness subscales are contained within the first half of this measure. Each individual rater’s average ratings for the first half SEQ score produced similar results. Both rated participants in the ST group as having higher average scores than those in the CM group, but with a difference that trended toward but did not reach statistical significance (rater 1, F = 3.723, p = .06; rater 2, F = 3.07, p = .09).

Most of the difference between the ST and CM groups observed on the first half of the SEQ was a result of difference on the Depth subscale. Average combined ratings of rater 1 and rater 2 on this subscale produced higher average scores for the ST group (M = 4.80, SD = .75) than for the CM group (M = 3.87, SD = .65). This difference was highly statistically significant (F = 12.41, p = .002). Both raters rated the ST and CM groups as different on their individual ratings (rater 1, ST M = 4.59, SD = .75; rater 1, CM M = 3.83, SD = .62; rater 2, ST M = 4.90, SD = .94; rater 2, CM M = 3.94, SD = .85). The differences between groups on the SEQ Depth subscale as rated by the individual raters were statistically significant in both cases (rater 1, F = 9.93, p = .004; rater 2, F = 7.64, p = .01). There was also a small difference between the groups on the smoothness subscale. The combined average ratings of rater 1 and rater 2 rated the ST group (M =
4.71, $SD = .88$) as having a slightly higher score than the CM group ($M = 4.60, SD = 1.03$) on this subscale, but the difference was nowhere near statistically significant ($F = .19, p = .67$).

A difference between the ST and CM groups was also detected on the second half of the SEQ. Combined average ratings of rater 1 and rater 2 rated participants who received ST ($M = 42.42, SD = 7.54$) as having numerically higher average second half SEQ scores than those who received CM ($M = 38.22, SD = 6.83$). This difference was trending toward statistical significance ($F = 3.57, p = .07$). Ratings on the second half of the SEQ were notably discrepant between the two raters. Rater 1 rated the ST group ($M = 42.23, SD = 6.44$) as having a higher average second half SEQ score than the CM group ($M = 36.94, SD = 4.89$), which were statistically significant ($F = 11.20, p = .002$). While rater 2 also rated the ST group ($M = 42.62, SD = 10.32$) as having a higher average second half SEQ score than the CM group ($M = 39.50, SD = 9.13$), the difference was smaller and there was more variability across his ratings, resulting in a nonsignificant difference ($F = .921, p = .35$) between the groups. (The difference between rater 1 and rater 2’s ratings on the second half of the SEQ remained consistent regardless of whether best agreement ratings or original ratings were compared. It was also consistent when individual sessions on which raters’ agreement was below .5 were thrown out.)

As described above, the second half of the SEQ is divided into subscales that assess level of positive mood and level of arousal. Looking at these subscales, it is clear that differences between participants in the ST group and those in the CM group on the second half of the SEQ were driven by differences on the Arousal subscale. Both rater 1 and rater 2 rated participants in the ST group as significantly more aroused on average.
Rater 1 produced a mean for the ST of $M = 4.12$ ($SD = 1.00$) compared to $M = 3.00$ ($SD = 0.74$) for the CM group. This difference was statistically significant ($F = 19.21, p = .00$).

Rater 2’s average Arousal scale rating was $M = 4.48$ ($SD = 1.14$) for the ST group, compared to $M = 3.58$ ($SD = 0.95$) for the CM group, a difference that was also statistically significant ($F = 4.66, p = 0.04$). Expectedly, based on prior results, the composite rankings of the ST group ($M = 4.30, SD = 0.94$) suggested more arousal than those in the CM group ($M = 3.34, SD = 0.82$), which was also significant ($F = 13.97, p = .005$). Both raters rated the ST and CM groups as nearly identical in score on the Positive Mood subscale. Rater 1’s ratings numerically slightly favored the ST group ($M = 3.34, SD = 0.61$) over the CM group ($M = 4.29, SD = 0.51$), while rater 2’s ratings slightly favored the CM group ($M = 4.37, SD = 1.10$) over the ST group ($M = 4.23, SD = 1.19$). Neither of these differences was anywhere close to statistically significant. The composite ratings also very slightly numerically favored the CM group on average Positive Mood scale, but was nonsignificant ($F = 0.04, p = 0.84$).

Combined average of rater 1 and rater 2’s ratings of therapist likability indicated that participants in the ST group ($M = 3.46, SD = .69$) liked their therapist more than did participants in the CM group ($M = 2.89, SD = .70$), a difference that was statistically significant ($F = 5.94, p = .02$). Both rater 1 (ST $M = 3.54, SD = .66$; CM $M = 2.78, SD = .73$) and rater 2 (ST $M = 3.38, SD = .77$; CM $M = 3.00, SD = .77$) rated therapists as more likeable in the ST group than the CM group. However, the difference between the groups was significant according to rater 1’s ratings ($F = 10.03, p = .004$), while the difference between rater 2’s ratings ($F = 2.26, p = .16$) was nonsignificant. Combined average ratings of rater 1 and rater 2 also rated the therapists as liking their clients more in the ST
group \((M = 3.62, SD = .36)\) than in the CM group \((M = 3.14, SD = .51)\), and the difference was statistically significant \((F = 7.86, p = .009)\). Both raters individually rated clients as better liked on average in the ST group than the CM group. Rater 1’s ratings indicated a significant difference between the groups \((F = 8.14, p = .008)\), while the difference between groups according to rater 2’s trended toward statistical significance \((F = 3.78, p = .06)\).

**Treatment by Status Interactions**

The series of 2 x 2 ANOVAs allowed for examination for the possible presence of a possible treatment by status interaction effect. No significant interactions were present on the WAI-O, the SEQ, or any subscales of either measure.

**Change in Alliance Between First and Last Sessions Rated**

As an additional set of analyses, we conducted a series of 2 x 2 ANOVAs to examine whether there were any significant differences based on status or treatment type in the amount of change in alliance between the first session and the last session that were rated. Obviously, only participants who had two sessions that were rated could be included for this supplementary analysis. Seven completers and eight dropouts were available, and seven participants who received ST and eight who received CM. No statistically significant differences existed on either the WAI-O or the SEQ.

Table 4 below summarizes the differences based on both status and treatment type that were statistically significant at least the .05 level. As mentioned above, the composite scores were deemed the primary evaluations of levels of alliance, and the WAI-O and SEQ total and subscales were all deemed primary dependent variables. Because this led to
nine dependent variables, a Bonferroni correction (.05/9) set the significance level at \( p \leq .006 \). Composite score differences whose \( F \)-score also surpassed the Bonferroni-corrected significance level are denoted with an asterisk (*).

Table 4

*Statistically Significant Status and Treatment-Based Differences*

<table>
<thead>
<tr>
<th>Scale/Subscale</th>
<th>Type of Rating</th>
<th>Status or Tx Type Difference</th>
<th>Group with Higher Avg. Alliance Score</th>
<th>( F )-Score</th>
<th>( p )-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ total score</td>
<td>Rater 1</td>
<td>Status</td>
<td>Completers</td>
<td>4.58</td>
<td>.042</td>
</tr>
<tr>
<td>2\textsuperscript{nd} half SEQ score</td>
<td>Composite</td>
<td>Status</td>
<td>Completers</td>
<td>4.71</td>
<td>.039</td>
</tr>
<tr>
<td>2\textsuperscript{nd} half SEQ score</td>
<td>Rater 1</td>
<td>Status</td>
<td>Completers</td>
<td>10.21</td>
<td>.004</td>
</tr>
<tr>
<td>SEQ Arousal subscale</td>
<td>Rater 1</td>
<td>Status</td>
<td>Completers</td>
<td>19.21</td>
<td>.000</td>
</tr>
<tr>
<td>SEQ Arousal subscale</td>
<td>Rater 2</td>
<td>Status</td>
<td>Completers</td>
<td>4.66</td>
<td>.04</td>
</tr>
<tr>
<td>SEQ Arousal subscale</td>
<td>Composite</td>
<td>Status</td>
<td>Completers</td>
<td>9.58</td>
<td>.005*</td>
</tr>
<tr>
<td>SEQ total score</td>
<td>Composite</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>4.80</td>
<td>.037</td>
</tr>
<tr>
<td>SEQ total score</td>
<td>Rater 1</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>7.74</td>
<td>.01</td>
</tr>
<tr>
<td>SEQ Depth subscale</td>
<td>Composite</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>12.41</td>
<td>.002*</td>
</tr>
<tr>
<td>SEQ Depth subscale</td>
<td>Rater 1</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>9.93</td>
<td>.004</td>
</tr>
<tr>
<td>SEQ Depth subscale</td>
<td>Rater 2</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>7.64</td>
<td>.01</td>
</tr>
<tr>
<td>2\textsuperscript{nd} half SEQ score</td>
<td>Rater 1</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>11.20</td>
<td>.002</td>
</tr>
<tr>
<td>SEQ Arousal subscale</td>
<td>Composite</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>13.97</td>
<td>.005*</td>
</tr>
</tbody>
</table>
Table 4—Continued

<table>
<thead>
<tr>
<th>Scale/ Subscale</th>
<th>Type of Rating</th>
<th>Status or Tx Group with Higher F-Score</th>
<th>Group with Higher Avg. Alliance Score</th>
<th>$F$-Score</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ Arousal subscale</td>
<td>Rater 1</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>19.21</td>
<td>.000</td>
</tr>
<tr>
<td>SEQ Arousal subscale</td>
<td>Rater 2</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>4.66</td>
<td>.04</td>
</tr>
<tr>
<td>Therapist likability</td>
<td>Composite</td>
<td>Treatment</td>
<td>Supportive therapy</td>
<td>5.94</td>
<td>.02</td>
</tr>
<tr>
<td>Therapist likability</td>
<td>Rater 1</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>10.03</td>
<td>.004</td>
</tr>
<tr>
<td>Client likability</td>
<td>Composite</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>7.86</td>
<td>.009</td>
</tr>
<tr>
<td>Client likability</td>
<td>Rater 1</td>
<td>Treatment</td>
<td>Supportive Therapy</td>
<td>8.14</td>
<td>.008</td>
</tr>
</tbody>
</table>

**Binomial Analyses of Differences in Alliance by Status**

Examination of the means produced by the $2 \times 2$ ANOVAs consistently indicated higher numerical alliance scores for completers than dropouts, even though many differences fell short of the statistical significance. In order to investigate the reliability of this directional effect for status, a series of binomial analyses on status based differences were conducted. Such analyses were conducted both between groups on all measures and between paired individuals on all measures.

For the between-group binomial analyses, we compared the number of mean scores that were higher for completers than dropouts for all outcome variables. Higher mean were coded as “1s” and lower means “0s” (without regard for the degree of difference between the means), and the binomial analysis tested whether the distribution of 1s and 0s was significantly significant from what would be expected if the probability
of a 1 or a 0 on any variable was equal to .50. Using the composite rating, there were 13
outcome variables to compare. Completers were rated as having higher scores than
dropouts on all 13 ($p = .0001$). Table 5 below summarizes the results.

Table 5

*Between-Group Binomial Analyses*

<table>
<thead>
<tr>
<th>Scale or Subscale</th>
<th>Group with Higher Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAI-O total score</td>
<td>Completers</td>
</tr>
<tr>
<td>WAI-O Task subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>WAI-O Goal subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>WAI-O Bond subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>SEQ total scale</td>
<td>Completers</td>
</tr>
<tr>
<td>1st half SEQ</td>
<td>Completers</td>
</tr>
<tr>
<td>SEQ Depth subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>SEQ Smoothness subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>2nd half SEQ</td>
<td>Completers</td>
</tr>
<tr>
<td>SEQ Arousal subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>SEQ Positive Mood subscale</td>
<td>Completers</td>
</tr>
<tr>
<td>Therapist likability</td>
<td>Completers</td>
</tr>
<tr>
<td>Client likability</td>
<td>Completers</td>
</tr>
</tbody>
</table>

For the between-individual analyses, we compared paired completer and dropout
individual scores on each of the outcome variables described above. Although there were
17 dropouts included in the study, only those who had been paired with a completer were used in the between-individuals analysis. There were 13 dropouts to whom a completer was available for a direct paired comparison. Two dropouts were each paired with two completers. In each of these cases, separate comparisons were conducted between the dropout and each of the completers to whom that dropout had been matched. (Note: Upon watching the videotapes for the study, it became clear that one participant quit treatment in a planned way with her therapist because of impending surgery. She informed her therapist that she was going to have to terminate treatment before it was complete, and continued to attend treatment until the time of her surgery. Because this circumstance was deemed qualitatively different than the other dropouts, this pair was not included in these one-to-one comparisons.)

As with the between group analyses, this analysis was based on who had the higher score, not the degree of difference between the scores. Using composite rankings, individuals who completed treatment had a higher score than their matched dropout individual in 87 of 169 comparisons (51%). Individuals who dropped out of treatment had a higher score than their matched completer in 66 of 169 comparisons (40%). Fourteen of 143 comparisons (9%) indicated identical scores. Significant differences were detected on the SEQ Depth subscale, SEQ Arousal subscale, and on the second half of the SEQ. Table 6 below depicts the between-individual comparisons in more detail.

DISCUSSION

The primary research question that this study sought to address was whether therapeutic alliance, as evaluated by trained raters, was predictive of propensity to dropout of treatment unexpectedly. The results suggest that aspects of the construct of
Table 6

*Between-Individual Binomial Analyses*

<table>
<thead>
<tr>
<th>Scale or Subscale</th>
<th>Comparison Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAI-O total score</td>
<td>7 completers higher, 6 dropouts higher</td>
</tr>
<tr>
<td>WAI-O Task subscale</td>
<td>6 completers higher, 6 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>WAI-O Goal subscale</td>
<td>7 completers higher 5 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>WAI-O Bond subscale</td>
<td>5 completers higher, 7 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>SEQ total scale</td>
<td>8 completers higher, 5 dropouts higher</td>
</tr>
<tr>
<td>1st half SEQ</td>
<td>5 completers higher, 7 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>SEQ Depth subscale</td>
<td>5 completers higher, 8 dropouts higher</td>
</tr>
<tr>
<td>*SEQ Smoothness subscale</td>
<td>*9 completers higher, 4 dropouts higher</td>
</tr>
<tr>
<td>*2nd half SEQ</td>
<td>*9 completers higher, 3 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>*SEQ Arousal subscale</td>
<td>*10 completers higher, 2 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>SEQ Positive Mood subscale</td>
<td>6 completers higher, 6 dropouts higher, 1 tie</td>
</tr>
<tr>
<td>Therapist likability</td>
<td>5 completers higher, 3 dropouts higher, 5 ties</td>
</tr>
<tr>
<td>Client likability</td>
<td>5 completers higher, 4 dropouts higher, 4 ties</td>
</tr>
</tbody>
</table>

*Rows denoted with an asterisk (*) indicates a significant difference between dropouts and completers in binomial analysis. In order to be statistically significant at the p ≤ .05, at least 9 comparisons had to favor individuals from one group over the other.*

**Totals:**

- Completers higher in 87/169 comparisons (51%)
- Dropouts higher in 66 of 169 comparisons (39%)
- Completers and dropout scores identical in 16/169 comparisons (9%)
therapeutic alliance do appear to be related to dropouts, but the picture the data paint is not entirely clear. Neither the WAI-O total score nor any of the subscale scores on the WAI-O (Task, Goals, and Bond) indicated any significant differences between completers and dropouts. However, there were some statistically significant differences indicated on the SEQ. The difference between completers and dropouts as scored by rater 1 on the SEQ total score tested as significant, favoring completers. Still, there are a couple of reasons to be cautious in interpreting this difference. First, although rater 2’s ratings also numerically favored completers, the difference did not test as significant. The same was true of the composite of the two raters’ ratings. The lack of agreement between the raters on the statistical significance of the difference on the SEQ total score, as well as the fact that the difference between completers and dropouts detected by the composite ratings was not significant, cautions against drawing conclusions based on total SEQ score.

The second reason for caution in interpreting the total SEQ score results is that closer analysis of the SEQ data indicates that the differences between completers and dropouts on the SEQ were largely driven by items on the second half of the measure, specifically by the items that compose the SEQ’s Arousal subscale. Differences between completers and dropouts were more robust on this variable than on any other employed in the study. This finding suggests that in investigating SEQ differences, it makes sense to focus most closely on this subscale.

While level of client arousal may not be as associated as closely with the construct of therapeutic alliance as other aspects considered in this study (bond, depth, etc.), it has previously been found to be predictive of improvements in self-esteem and self-reported symptoms of depression (Missirlian, Toukmanian, Warwar, & Greenberg, 2005). In this
study, participants rated as being relatively aroused at mid-treatment tended to show improved levels of self-esteem on a 10-item version the Rosenberg Self Esteem Scale (RES; Bachman & O’Malley, 1977) and on the Beck Depression Inventory (BDI; Beck, 1972) in comparison to those who were rated as being less aroused. Ratings of arousal were made by trained raters using the Client Emotional Arousal Scale-III (CEAS-III; (Warwar & Greenberg, 1999). The raters watched videotapes of therapy sessions of individuals being treated for depression and made ratings on the CEAS-III while blind to the treatment outcome of the participant who they were rating. Missirlian and colleagues used hierarchical regression analysis to evaluate the relative explanatory power of a variety of predictor variables, and found that arousal accounted for approximately 30% of the variance detected on the BDI. Additionally, the authors reported that arousal as measure by the CEAS-III at mid-treatment was significantly predictive of improved self-esteem post-treatment.

The finding in the current study that level arousal as measured by trained raters on the SEQ is predictive of treatment retention seems to add to the literature suggesting that arousal is important to treatment outcome. Missirlian et al.’s (2005) study looked only at participants treated with process-experiential psychotherapy. The authors of that study described experiential therapies as having the goal of helping clients become aware of emotional experience within the context of a therapeutic bond; all [experiential therapies] subscribe to the importance of engaging clients in an in-depth exploration of their feelings, perceptions, and thoughts associated with problematic life events; and all believe that for therapeutic change to occur, therapy must help clients to access different facets of their experience of such events so that they can develop a fuller understanding of the meaning that the experience has for them. (p. 861)
The results from the current study suggest that participant arousal may be important in therapy using other techniques as well, as participants who were rated as having higher levels of arousal were more likely to complete treatment regardless of whether they were in the CM group or the ST group. However, it seems likely that the relationship between arousal and the benefits of therapy is non-linear. It should be noted that the average composite Arousal score for completers in the current study was 4.16, compared with 3.30 for dropouts. The Arousal score is based on a 7-point scale, so the mean score for completers is very close to the mid-point of that scale. Missirlian and colleagues (2005) assert that a score of 4 (which is also the mid-point on the CEAS-III) is ideal for participants receiving process-experiential therapy, because arousal at that level best facilitates processing. The general relationship between level of experienced stress (or arousal) and performance has long been a topic of interest within psychology. As early as 1908, psychologists have demonstrated that increased arousal improves performance, but only to a certain point, after which continued increase in arousal leads to a decline in performance (Yerkes & Dodson, 1908). It seems plausible that what is known as the “Yerkes-Dodson Law” may apply to psychotherapy, and the results of the current study appear to indicate that the proper degree of arousal increases the probability that a therapy participant continues to come to therapy. More research is necessary to establish the optimal level for keeping therapy patients in treatment, and to more finely investigate whether or not the “Yerkes-Dodson Law” applies to alliance and therapy retention. The mechanism by which arousal may encourage treatment retention is unclear. Perhaps participants who appear more aroused in treatment are more engaged in the therapeutic process, and thus are more committed to doing the work entailed by psychotherapy.
Future research would be necessary to better understand the mechanism by which arousal may related to propensity for treatment dropout. Relatedly, future studies may seek to investigate more precisely the relationship between arousal during therapy and treatment outcome.

Decades before the study by Missirlian and colleagues (2005), B. F. Skinner (1953) offered theoretically-driven speculation of a possible mechanism of action of psychotherapy that seems to indirectly suggest arousal as a key to therapeutic success. Based on his experimental work with operant conditioning, Skinner hypothesized that psychotherapy was most likely to be successful if the therapist is able to establish therapy as a non-punishing environment. He conceived many problems that are targets of psychotherapy as problems of avoidance, whereas the client acts in an attempt to avoid emotions that are perceived as aversive that tend to follow certain behaviors. Through a therapist establishing herself as a non-punishing audience, Skinner argued that extinction should occur, in the sense that behaviors emitted by the client that may have previously been punished would not be, and thus the link between such behaviors and aversive emotions (e.g., anxiety) would be weakened or broken. This conceptualization of a potential mechanism of psychotherapy may be viewed as an argument for the necessity for arousal to occur in therapy in order to lead to patient improvement. In theory, the non-punishing environment leads the patient to emit a wider range of behaviors and allows the expression of a wider range of previously punished emotions. The result would likely be a more active, engaged, and expressive patient—one who is more aroused. In this conceptualization, arousal may be conceived more as an effect of the non-punishing environment that eventually leads to patient improvement (rather than the direct cause of
improvement itself), but it would nonetheless account for an association between arousal and treatment outcome.

Data from the current study most strongly suggest arousal as an aspect of therapeutic alliance that predicts whether or not a participant drops out of treatment. It was the only scale or subscale on which the difference in average score between dropouts and completers was statistically significant based on both raters’ individual ratings as well as based on the composite of their ratings. However, when the cumulative results of the binomial analyses described in the preceding section are examined, it appears that other aspects of therapeutic alliance may also relate to the probability of whether or not a participant stays in treatment until its conclusion. Both raters individually rated completers as having numerically higher alliance scores than dropouts on every scale or subscale assessed. Even though in many cases the difference was too small and/or had enough variation around the means so that the difference tested as non-significant, the fact that the results reliably favored the therapeutic alliance between completers and their therapists over the alliance between dropouts and their therapists suggested that there may be a numerically small but clinically meaningful difference present. The between-individual comparisons allowed us to more closely examine these differences.

Looking at the results of the between-individual comparisons, it appears that some of the differences between completers and dropouts favoring alliance of completers that were evident based on the between-group binomial analyses were driven by a couple of individuals rather than by a group trend. For example, even though the group mean for completers was higher on the SEQ Depth subscale than it was for dropouts on that subscale, at the individual level, 9 of the 13 comparisons between paired individual
participants favored the dropout on this subscale. This finding suggests that a couple of individual participants who completed treatment had notably high scores on this subscale, that a couple of individual dropouts had especially low scores on this subscale, or some combination of these two circumstances. Regardless, in circumstances in which the between-group comparisons favor the completers and a slight majority of individual comparisons favor the dropouts, it seems clear that there is no meaningful difference between the groups in alliance.

On the other hand, the between-individual comparisons for both the Smoothness and Arousal SEQ subscales significantly favored the completers over the dropouts. On the Arousal subscale, this finding simply provides additional evidence that there is a meaningful difference between groups that is not an artifact of a few individuals with extreme scores. Given that the ANOVAs clearly indicated significant differences on this subscale, this finding is not surprising. However, the finding that a significant number of individual completers had a higher Smoothness score than their matched dropout may be interesting. It suggests that there is a trend that those who completed treatment consistently had slightly smoother sessions than their counterparts who dropped out of treatment. This is a different route to the small difference in the group means than what occurred on the Depth subscale, when the difference was apparently the result of some extreme scores for a few individuals. It is plausible that, with a larger sample, the difference on the Smoothness scale would test as statistically significant. Whether or not the difference would be clinically significant is unclear.

With respect to treatment type, the results seem to indicate a difference in therapeutic alliance. Both raters seemed to have a clear preference for ST. Individual rater
scores and composite ratings were numerically higher for the ST group on every scale and subscale evaluated, with the exception of rater 2’s and the composite rating on the Mood subscale of the SEQ, which very slightly favored the CM group. Among the differences favoring the ST group, 13 were statistically significant, including composite ratings on the SEQ total score, the SEQ Depth subscale, the SEQ Arousal subscale, therapist likability, and client likability.

It does not appear that the raters’ clear preference for the ST mode of therapy was predictive of reductions in the symptoms targeted by Clore and Gaynor (2009). Outcome data from that study indicated that participants in the CM group improved more on average than did those in the ST group. Statistically significant time*treatment interactions suggest that participants receiving CM reported less internalizing distress ($F = 4.07, p = .02$) and more positive thinking ($F = 5.98, p = .004$) than those receiving ST. Internalizing distress and positive thinking were composite variables that emerged as the result of a principal component factor analysis, following the strategy outlined by Gotlib (1984). These data contradict the idea that the stronger working alliance reliably detected by the raters for the current study between participants and therapists in the ST group in comparison to the alliance between participants and therapists in the CM group might correlate with what would likely be considered meaningful outcome variables.

Although not as definitively as on the Arousal subscale, the differences between the ST and CM groups on ratings for SEQ total score, therapist likability, and client likability all tested as significant based on the composite ratings. The reason for the assertion that the differences are not as definitive as that detected on the Arousal subscale is that while differences between groups according to rater 1 and the composite ratings
were statistically significant, the differences according to rater 2 ratings, while favoring the ST group in each case, did not test as significant. Because Arousal was rated as a significant factor based on both treatment type and status, it appears to be the most interesting variable on which to focus. These findings suggest a treatment*status interaction effect on the Arousal subscale, but no statistically significant interaction was found. This lack of a significant interaction effect appears to signal that it was a different set of individuals in the ST group that was accounting for the relatively higher Arousal score than the individuals accounting for the relatively higher Arousal score for completers. Alternatively, it could simply be an issue of a lack of sufficient statistical power to detect a numerically small, but potentially significant interaction.

Like the comparisons based on status, the results comparing participants based on treatment type that most strongly and definitively favor ST are the Arousal subscale scores on SEQ. Both individual raters and the composite ratings significantly favored the ST group on this subscale. Because arousal seems to have been a significant predictor of propensity to drop out of treatment, this finding is worthy of additional attention. This finding would seem to suggest that since people with sufficient arousal during sessions are more likely to stay in treatment, and because those who received ST tended to be closer to the proposed optimal levels of arousal, ST should be strongly considered over CM. The data from Missirlian et al. (2005) also suggest that treatment strategies that lead to optimal levels of arousal may lead to better outcome. However, there is a problem with this conclusion: The results from Clore and Gaynor (2009) indicate no group differences in dropout rate (approximately 30% of participants from the intent-to-treat sample dropped out from both the ST and CM groups), and that CM produced superior results on
the primary outcome variables of interest. The Clore and Gaynor study was designed to test whether ST or CM was significantly better at increasing self-esteem and reducing symptoms of depression. As described above, their results indicated that CM was better in this regard. Thus, concluding that best practice would be to select ST over CM because it tends to lead to the level of arousal that makes patients more likely to complete treatment may lead the therapist to sacrifice some potency of treatment to address primary variables of interest, in spite of the results of Missirlian et al. when testing a different mode of therapy. Nonetheless, if it is the case that ST leads to the optimal level of arousal that makes patients more likely to stay in treatment (as data from the current study seem to suggest), disregarding this finding might cost therapists the opportunity to make a difference for more patients than they might otherwise.

Obviously, therapy cannot make a difference for a person if he or she does not attend sessions, so aspects of therapy that lead patients to stay in therapy are important to consider, even if the same aspects do not directly lead to improvement in symptoms. Of course, the data from this study are far too preliminary to draw a strong conclusion that ST leads to arousal, which serves as a mediating variable to increased treatment attendance. However, if future research were to provide stronger evidence that this is indeed the case, therapists and researchers would no doubt be interested in what aspects of ST were responsible for this increase in arousal. Because it seems clear that CM is more effective than ST in alleviating the primary symptoms of interest, such a finding would be unlikely to lead to the conclusion that therapists treating depression or low-self esteem should select a ST protocol over a CBT-based protocol. At the same time, it would suggest isolating the factors of ST that are leading to arousal and trying to
incorporate them into other protocol types that are more potent in treating symptoms, in an effort to keep as many patients as possible in treatment to its end. In making such a move, one key question may be whether altering treatment with demonstrated efficacy to increase its propensity to keep patients in therapy would reduce the treatment’s potency and efficacy. This is an empirical question, and beyond the scope of the current study to answer.

However, other researchers have started to investigate the possibility of altering established treatment strategies in order to use the power of the relationship between therapist and patient to increase the efficacy of treatment. Kohlenberg, Kanter, Bolling, Parker, and Tsai (2002) blended Functional Analytic Psychotherapy (FAP) with Beck’s well-established cognitive therapy (CT) in order to take advantage of the power of immediately addressing clinically-relevant client behaviors that emerge in the context of the session. The result of the blend of these techniques is a mode of therapy that the authors call FAP-enhanced CT (FECT). Unlike many others interested in the importance of the therapeutic relationship from a more psychodynamic perspective, FAP emerged from a soundly behavior analytic tradition. The changes to traditional CT that occur in FECT include an expanded rationale given to the client about the relationship between the environment, the client’s thoughts, the client’s feelings, and the client’s behavior. Whereas traditional cognitive therapists assert that thoughts have a direct causal relationship to feelings or overt behavior, FECT therapists view cognition as covert behavior, but behavior nonetheless. Because cognition is not given special status, the relationship between thoughts and behavior is explained as a behavior-behavior relationship. From this perspective, while cognition may still act as a moderating or
mediating variable affecting overt behavior, it broadens the possibilities for the causal mechanisms that lead to the emergence of emotions or behavior.

More relevant to the current discussion of therapeutic alliance, the second modification to CT that occurs in FECT includes explicit use of the relationship between therapist and client to elicit change. FECT therapists believe that the therapeutic relationship is important for promoting client change. However, the proposed mechanism by which this relationship influences change is quite different than the model of therapeutic alliance that is purported by followers of Borden or Rogers. In FAP, and by extension, in FECT, therapists look for examples of behavior that occur in-session that are relevant to problems for which the patient is seeking therapy. Called Clinically-Relevant Behaviors (CRBs), FECT therapists look for these behaviors to occur in-session. The well-established principle of reinforcement dictates that consequences most strongly influence behavior when they occur in close temporal proximity to the behavior. Following from this principle, FECT therapists theorize that clinical interventions will be most powerful if the therapist is able to provide consequences for behavior as it occurs in therapy, rather than simply talking about behavior and potential consequences that occur outside the session. Kohlenberg and colleagues (2002) call this process “in vivo” therapeutic work, and assert that it is more powerful that trying to work on imaginary or described situations that occur outside of the therapy room. They tested the relative efficacy of FECT compared to traditional CT. Four experienced CT therapists were assigned depressed clients in waves over a 6-month period. In the 7th month, the therapists underwent a 6-hour workshop for FECT. After the workshop, they were assigned similarly depressed patients in waves to be treated with FECT. Relevant
symptoms were measured by the Beck’s Depression Inventory (BDI), Hamilton Rating Scale for Depression (HRSD) Symptom Checklist-90 (SCL-90), and the Global Assessment of Functioning scale (GAF).

Although results from this study must be interpreted cautiously because it was not a randomized clinical trial, data suggest superior results for participants treated with FECT. FECT produced more treatment “responders” (i.e., those who demonstrated at least a 50% reduction of depressive symptoms from pretreatment) than did CT on all dependent variables. Averaging the BDI and the HRSD, 79% of FECT participants responded to treatment, compared to 60% in CT. On the SCL-90, 67% of FECT participants responded, compared to 45% in CT. FECT also produced fewer treatment “failures” than did CT on all outcome variables. Kohlenberg et al. (2002) defined treatment failures as participants who demonstrated less than a 25% reduction in symptoms. Although further study is needed, the evidence from the study by Kohlenberg et al. seems to suggest one way of modifying CBT-based therapies to make more use of the therapeutic relationship. Interestingly, the authors note in the discussion that the apparently superior efficacy of FECT was observed despite the fact that the therapists in the study showed very little increase in the amount of in-vivo cognitive therapy. The areas in which they did show an increase was the amount of focus on intimate and avoiding styles of relating between therapist and patient. While the study by Kohlenberg and colleagues does not address the impact of the increased focus on the therapeutic relationship on treatment retention, it does suggest that such focus may improve impact of CT on outcome. In fact, given the observation by Kohlenberg et al. that the nature of the relationship and discussion of that relationship changed more than the use of in-vivo
cognitive therapy, it suggests that such modifications in the relationship may have contributed to a significant amount of the variability in the amount of improvement observed for participants receiving FECT compared to those who received traditional CT.

It is not possible to empirically explain the differences in therapeutic alliance between the CM and ST groups based on available data. However, some significant differences in the way the sessions were run in the CM protocol compared to in the ST protocol invites speculation. The procedure of the sessions in CM was much more structured than in ST. Most of the sessions in CM involved either formally challenging identified negative thoughts or using flash card practice in an attempt to increase the fluency of identified positive thoughts. While the data from Clore and Gaynor (2009) suggest that these procedures were effective in terms of increasing self-esteem, it is possible that they were not as stimulating to the observers, especially when watching them repeatedly on videotape. It is plausible that the highly-structured format of the CM sessions limited the extent to which the client and therapist could connect. In the ST protocol, the sessions were much less structured and more free-flowing. One of the main goals in this protocol was for the therapist to understand the participant’s experiential world. Such a strategy may lead to a stronger connection between therapist and participant, particularly as perceived by the raters. That said, it must be noted that although alliance was rated as stronger in the ST group than in the CM group, it was rated as being fairly strong for both. The overall WAI-O scores were 201 and 205 for the CM and ST group, respectively. Both of these averages represent at least 80% of the maximum total score possible (252) on the WAI-O. While no benchmarks exist for this
measure, the fact that scores were well above the mean for both the CM and ST groups suggests that alliance was good for both.

It is possible that alliance in CT and CBT takes a different form or forms from different paths than in other modes of therapy. Andrusyna et al. (2001) found that when using a shortened version of the WAI-O (WAI-O-S), a two factor model of alliance was apparent. These authors suggest that alliance in CBT may be conceptualized as two factors: "Agreement/Confidence" and "Relationship." The "Agreement/Confidence" factor contained items from all three subscales (Goals, Task, and Bond) contained in the WAI-O-S, while the "Relationship" factor contained only items from the bond subscale. As the authors acknowledge, the Andrusyna et al. study had some limitations that suggest caution in extrapolating its results, such as the fact that ratings were based on audio rather than video tapes and they used only the short version of the WAI-O. Not being able to see the session may have prevented raters from picking up on non-verbal signs of alliance, and using the long version of the scale may have allowed for a more precise construct of alliance in CBT. Also, the authors in this study based all ratings on observations of the second session of a CBT protocol. It is possible that alliance may be variable in its development and presentation at different times in treatment. Future studies investigating the role of therapeutic alliance in CBT protocols may wish to further investigate the form of such alliance in CBT. Such studies may simultaneously investigate whether altering CBT-based protocols to increase therapeutic alliance adversely affects the capacity of such protocols to treat depressive symptoms and increase self-esteem.

The current study was intended to examine the influence of therapeutic alliance on treatment retention. According to available data, the factor that seems to be most clearly
associated with treatment retention is level of participant arousal as measured by the SEQ. As described in the introduction section, there is no clear consensus on a definition of the construct of therapeutic alliance. Arousal does not seem to neatly map on to the construct of therapeutic alliance as a synthesis of agreement on therapeutic goals and tasks in addition to bond between therapist and patient, as proposed by the enduring definition of Bordin (1979) and assessed by the most widely used assessment of Therapeutic Alliance (the Working Alliance Inventory; Horvath & Greenberg, 1989). As noted above, the SEQ in this study was modified, so that the introduction to the items on the second half of the tool (which contains all items that make up the Arousal subscale) reads “right now the participant feels . . .” rather than “right now I feel . . . .” The original wording introducing the section is clearly asking the patient completing the form to rate their immediate post-session emotions, as opposed to directly assessing their impressions of their alliance with their therapist. That said, there are existing definitions of therapeutic alliance that are broader than that proposed by Bordin (1979). Elvins and Green (2008) conceive alliance as the result of interpersonal processes between patient and therapist. While not explicitly asking about alliance, the questions on the arousal subscale may be an indirect assessment of some such processes. In the end, the data from this study suggest that participant arousal is predictive of treatment retention regardless of whether or not one conceptualizes arousal as defined in this study as part of the construct of therapeutic alliance. Its place within or outside that construct is a theoretical issue, while the relationship between arousal and treatment attendance is empirical.

One limitation of the current study was the disappointing level of agreement between raters on the SEQ. The hope was that raters would demonstrate agreement at the
level of at least $r \geq .70$. While agreement easily surpassed that level on the WAI-O ($r = .90$), it fell noticeably short of .70 on the SEQ (.59). This level of agreement can be classified as fair-to-good (Chicchetti & Sparrow, 1981). Although this level of agreement easily surpassed statistical significance ($p = .00$), stronger of agreement would have made for easier interpretation of results. Unfortunately, all significant between-group differences were detected on the SEQ, on which the raters’ agreement was below what would be optimal. We had no way of determining whether one rater’s evaluations should be considered more valid than the others, as there was no real anchor benchmark against which to compare their evaluations. It was for this reason that we elected to use the composite ratings as the primary dependent variable, as opposed to using one of the rater’s individual ratings.

There were factors that indicate that the ratings are valid despite the less-than-optimal level of agreement. Perhaps most significantly, the raters agreed on which group mean was higher (i.e., completers or dropouts; ST or CM) in every comparison with the exception of the Mood subscale of the SEQ in comparing participants based on treatment type. However, on this subscale, both raters rated the ST group and the CM group as very similar, so the fact that they each slightly a different group is not especially notable. In all cases that the composite scores indicated a significant difference between the groups being compared, the raters agreed on which group had a higher mean score, although it was sometimes the case that one rater’s difference tested as significant and the other’s did not. In considering the above factors, it appears that although a higher level of agreement between raters would have been desirable, they agreed sufficiently on a molecular level so that the ratings should be considered valid.
Another limitation of the current study was sample size. Because the project was conducted using archival data, we were limited by the videotapes that were available from the Clore and Gaynor (2009) study. Within this sample, there were only 17 dropouts available. Because the completers included in the sample for the current sample each needed to be matched to a dropout who had the same therapist, there were only 14 available completers who met criteria for inclusion. A larger sample size would allow for increased power to detect differences that were potentially meaningful, yet small-to-medium in magnitude, and a more definitive answer as to whether some of the differences observed based on both status and treatment type were significant.

A third limitation of the study is the fact that its most interesting results were found on a dependent variable that was not part of a priori predictions. As described above, the primary aim of this study was to examine whether or not level of therapeutic alliance was predictive of treatment retention. It was predicted that alliance would be stronger for completers than dropouts as measured by the WAI-O total and subscale scores, as well as by the SEQ total score and Smoothness and Depth subscale scores. No predictions were made regarding the relationships between the subscales contained in the second half of the SEQ and treatment retention. All items that compose the Arousal subscale of the SEQ are contained in the second half of the instrument. Because all analyses in the current study were conducted using statistics based on null hypothesis testing, results regarding the Arousal subscale would have been stronger if an a priori prediction regarding Arousal had been made.

This study has indicated some avenues of potential future research for investigators interested in the impact that therapeutic alliance has on treatment. Evidence
from the current project suggests that supportive therapy may lead to stronger therapeutic alliance, and, particularly, levels of arousal that may be optimal for encouraging participants to remain in treatment until its end. Future research projects may wish to dismantle a supportive therapy protocol in an effort to determine which aspect(s) of such therapy are most responsible for stronger alliance. Answering this question may be important, because it may help keep participants in treatment. Past research has identified therapist flexibility, honesty, respectfulness, trustworthiness, confidence, warmth, interest in the patient, and openness all contribute to the development of good alliance (Ackerman & Hilsenroth, 2003). If it is true that therapist variables can influence the development of alliance, it begs the question of whether therapists can be instructed in a way that helps them increase their ability to form strong alliance with their patients.

The data from the Kohlenberg et al. (2002) suggest one method of altering CBT-based protocols to increase focus on the therapeutic relationship, and possibly increase efficacy. Additionally, Crits-Cristoph, Connolly Gibbons, Crits-Cristoph, et al. (2006) conducted a pilot study in an attempt to begin to investigate whether therapists can be explicitly trained in a way to help them improve their capacity to form strong therapeutic alliance with their patients. These authors trained five relatively novice (1 to 3 years of postdoctoral therapy experience) therapists in a 16-session therapy for depression called "alliance-fostering therapy." Alliance-fostering therapy is a treatment that combines aspects of psychodynamic-interpersonal therapy with strategies for building therapeutic alliance based on the Bordin (1979) model of alliance. Strategies for building alliance drew on CBT, motivational enhancement therapy, client-centered therapy, and supportive therapy.
Results from the study were mixed. Overall, Crits-Cristoph, Connolly Gibbons, Crits-Cristoph, et al. (2006) found moderate-to-large increases in therapeutic alliance from pre- to post-treatment, although these changes were not statistically significant. There were significant differences in the amount of change in alliance across therapists included in the study. Crits-Cristoph, Connolly Gibbons, Crits-Cristoph, et al. found that there were significant differences in the amount of improvement on one subscale of the CALPAS (the Working Capacity scale), but not other assessments of alliance. They hypothesize that this may be attributable to some aspects of the construct of therapeutic alliance being easier to learn than others. Also, despite the fact that therapists in the study were relatively in inexperienced, the authors indicated that they believed that after intensive grad school training, a clinical internship, and some postdoctoral clinical experience, the therapists may have already developed their own style of building alliance and be quite skilled in so doing, which would create a problem of a ceiling effect. The counterpoint of this argument may be that if clinicians are already developing skills for building alliance through standard training procedures by the time they are 1–3 years post-training, an intensive training procedure to help them learn to build alliance may be unnecessary. Despite its limitations, this study indicates that it may be possible to explicitly teach therapists to become better at building therapeutic alliance with their patients.

Future research may also investigate whether CBT-based protocols could be altered in order to increase therapeutic alliance. CBT is one of the most widely used forms of treatment for depression, and many studies have indicated its efficacy. However, data from the current study indicate that therapeutic alliance in CBT may not be as strong
as in supportive therapy protocols. As many researchers have demonstrated (e.g., Klein et al., 2003; Martin et al., 2000), alliance is a reliable predictor of outcome. It is plausible that increasing the therapeutic alliance in CBT protocols may boost their efficacy even further. Kohlenberg and colleagues (2002) describe one method for doing so, and provide preliminary evidence that increased focus on therapeutic alliance in therapy can enhance treatment effects.

Despite its limitations, the results of the current study do suggest that aspects of the therapy process that can be empirically evaluated do impact whether or not a therapy participant completes a treatment regimen as planned or drops out unexpectedly. Supportive therapy may lead to stronger therapeutic alliance than CBT-based therapies. Data from this study strongly suggest that level of participant arousal is predictive of propensity to dropout of treatment, and that the ST group in this study had more optimal levels of arousal as measured by the SEQ than did those in the CM group. On the other hand, outcome data from Clore and Gaynor (2009) indicate that the CM protocol produced better outcomes on assessments of depressive symptomology and self-esteem. If future research continues to indicate that arousal is key to keeping patients in treatment, and something about supportive therapy contributes to such arousal, it would be useful to isolate the aspects of such therapy that lead to optimal arousal, and investigate whether these aspects can be integrated into more potent therapy modalities without compromising their efficacy.
REFERENCES


