Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: http://www.tandfonline.com/loi/tpsr20

Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome

Timothy Anderson, Mary Ellen J. Crowley, Lina Himawan, Jennifer K. Holmberg & Brian D. Uhlin

Psychotherapy Research, 2015
http://dx.doi.org/10.1080/10503307.2015.1049671

Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome

TIMOTHY ANDERSON[1], MARY ELLEN J. CROWLEY[2], LINA HIMAWAN[1], JENNIFER K. HOLMBERG[1], & BRIAN D. UHLIN[3]

[1]*Department of Psychology, Ohio University, Athens, OH, USA;* [2]*Department of Psychiatry, Harvard Medical School, Boston, MA, USA &* [3]*Department of Psychiatry, Kaiser Permanente, Pleasanton, CA, USA*

(Received 4 July 2012; revised 21 April 2015; accepted 29 April 2015)

Abstract

Objectives: Therapist effects, independent of the treatment provided, have emerged as a contributor to psychotherapy outcomes. However, past research largely has not identified which therapist factors might be contributing to these effects, though research on psychotherapy implicates relational characteristics. The present Randomized Clinical Trial tested the efficacy of therapists who were selected by their facilitative interpersonal skills (FIS) and training status. **Method**: Sixty-five clients were selected from 2713 undergraduates using a screening and clinical interview procedure. Twenty-three therapists met with 2 clients for 7 sessions and 20 participants served in a no-treatment control group. **Results**: Outcome and alliance differences for Training Status were negligible. High FIS therapists had greater pre–post client outcome, and higher rates of change across sessions, than low FIS therapists. All clients treated by therapists improved more than the silent control, but effects were greater with high FIS than low FIS therapists. From the first session, high FIS therapists also had higher alliances than low FIS therapists as well as significant improvements on client-rated alliance. **Conclusions**: Results were consistent with the hypothesis that therapists' common relational skills are independent contributors to therapeutic alliance and outcome.

Keywords: therapist characteristics; common factors; outcome research; process research; interpersonal skills

> Whether they openly acknowledge it or not, individuals approaching a psychotherapist are seeking a "good" human relationship and a satisfying relatedness. If their history, which is embodied in their current functioning, allows them to enter such a relationship, and if the psychotherapist can provide the kind of relationship that allows them to form a satisfying relatedness, the basic conditions for a successful therapeutic result are met. (Strupp, 1980, p. 603)

Recent research findings have demonstrated an association between therapist characteristics and therapy outcomes (e.g., Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009; Dinger, Strack, Leichsenring, Wilmers, & Schauenbirg, 2008; Okiishi, Lambert, Nielsen, & Ogles, 2003; Wampold & Bolt, 2006). These findings further supported Wampold's (2001) landmark meta-analysis (see also Wampld & Imel, 2015), indicating that therapist effects accounted for a meaningful portion of outcome variance (between 6% and 9%). When there is a sufficient sample size of clients per therapist, studies have found that the individual therapist, treated as an independent variable, significantly predicted client outcomes (Okiishi et al., 2003; Wampold & Bolt, 2007), though there have also have been findings to the contrary (Elkin, Falconnier, Martinovich, & Mahoney, 2006).

Prior research identifying therapist effects and characteristics has been limited by the fact that (a) therapist effects are identified without planned measurement of the individual differences, traits, skills and other characteristics that might account for the differences, (b) findings are the result of secondary analyses of archival data in which primary aims of the original study were to *reduce or eliminate* these therapist factors in favor of testing the effects

Correspondence concerning this article should be addressed to Timothy Anderson, Department of Psychology, Ohio University, Athens, OH 45701, USA. Email: andersot@ohio.edu

of therapeutic procedures, and (c) the study of therapist effects is a relatively new area of research in contemporary psychotherapy research. As Garfield (1997) summarized, the therapist is a "neglected" variable in psychotherapy research.

What therapist characteristics might contribute to effective therapy? Even while there has been a shortage of empirical studies on therapist characteristics, there has been an abundance of clinical research on therapy process and relational variables that, by association, implicate the therapist. An APA Division 29 task force on Empirically Supported Relationships (ESRs; see Norcross, 2011) identified "The Person of the Therapist" as one of three major sources that likely account for effects of relational process variables. Thus, one plausible clue might be that the most significant ESR process variables imply analogous therapist characteristics. For example, the fact that empathy is among the top process-outcome effects might imply that the therapist's *skillfulness* in being empathic with others might be *one* of the *indicants* of positive therapist effects. In fact, most of the ESRs could seamlessly be defined as a therapist skill that is used to promote facilitative conditions, including empathy (Elliott, Bohart, Watson, & Greenberg, 2011), therapist-offered alliance (e.g., Crits-Christoph et al., 2006; Horvath, Del Re, Flückiger, & Symonds, 2011), positive regard, warmth and/or genuineness (Farber & Doolin, 2011). Not only have these variables received strong empirical support, but recommendations from the task force were directed toward individual therapists, including how individual therapists might effectively develop skills that would optimize the strength of those ESRs. In addition, there are a number of common therapist skills that have been grounded in clinical theory and research, including therapist persuasiveness (Frank & Frank, 1993), ability to repair alliance ruptures (Safran & Muran, 2000); verbal fluency and expressiveness (e.g., Greenberg & Paivio, 1997; Rice & Kerr, 1986), and the ability to enhance expectations and hopefulness (Constantino, Glass, Arnkoff, Ametrano, & Smith, 2011).

The therapist contributions to these relationship-based variables have been difficult to parse out from other process variables for several reasons. First, therapeutic relationship variables tend to be highly inter-correlated and it is not clear whether (as a skill variable) there are numerous independent relational therapist skills or whether therapist relational skills are more generic and form a general interpersonal relatedness factor. Second, the unique therapist contribution for each of these relationship variables has been difficult to measure because any therapist's actions are reciprocally influenced by each client's characteristics and other therapeutic context variables. Thus, measuring therapist skills within therapy sessions will always be influenced and potentially confounded by each individual client. Third, treatments being studied within a research setting may alter these relational variables in numerous ways because of expectations by clients about scientific treatments (Wampold, 2007), therapist allegiances about the treatment (Hollon, 1999), and demand characteristics (Anderson & Strupp, 1996).

There are serious obstacles to attaining experimental control of therapist interpersonal contributions. Most randomized clinical trials (RCTs) in contemporary psychotherapy research have the opposite aim of systematically removing therapist effects as part of a primary objective of experimental specification of techniques or treatments. Many RCTs have carefully attended to equalizing all therapist interpersonal variance by including therapists who demonstrate fidelity to the treatment manual being tested, and more recently attending to maintaining equally high alliance-building, empathy, warmth, and other EST-type skills. This laudable goal has aided the understanding of treatment packages. However, attentiveness in RCTs to selecting therapists who display good clinical skills is no replacement for understanding what therapist characteristics and skills are optimal in the first place or identifying their independent effects.

While thousands of RCTs have effectively controlled for technical and treatment variables, studies that have effectively controlled for relational variables are sparse to non-existent. Treatments as independent variables share common characteristics with relational variables (Hatcher & Barends, 2006; Wampold, 2007), including their continuous nature when measured with adherence and competence measures. Although the manipulation of both treatment and relational variables presents different experimental challenges, there is no reason that many relational questions in psychotherapy cannot also be advanced through experimental methods. For example, therapist pre-treatment characteristics have been successfully, though less commonly, identified in research (e.g., Connolly, Crits-Christoph, Barber, & Luborsky, 2000). It stands to reason that experimental control of these therapist characteristics and skills is necessary to better determine if it is the identified therapist-specific skills that are accounting for changes in process and outcome.

Therapist Characteristics and Experimental Control

What experimental strategies might allow for control of therapist characteristics? In a classic study that

inspired the design of the current study, Strupp and Hadley (1979) reasoned that it would be possible to separate specific factors of psychotherapy from common factors through pre-treatment selection of therapists. The treatment group was defined at the person-level, professionally trained therapists (assumed to have both specific / technical factors and common factors) who were compared to a person-level control group of college professors, untrained in the techniques of psychotherapy (and hence, assumed to have common/relational factors only). There were no significant differences between the groups. Unfortunately, the equivalent results garnered considerable attention to implications of training, which was not a purpose of the study. Training status as a variable was selected purely as a design choice for separating technical factors from common relational factors. Thus, the design of the study specifically was *not* a test of the practical effects of training. As Strupp (1998) reflected, the selection of therapists for the study, "should have been equated on the nonspecific (warmth and friendliness) dimension *in advance*, not after the fact" (p. 20).

As Lambert (2013) noted, the myriad of different common factors make it difficult to specify and separate any one common factor. Common factors most often are given primary attention as effective process and mediator variables, but are then demoted to control conditions when forming groups within experimental designs. As controls, these groups are often labeled as common factor control conditions (Baskin, Tierney, Minami, & Wampold, 2003). Wampold (2007) elegantly demonstrated that such control groups become increasingly effective as the number of structurally equivalent, but non-specific, treatment characteristics are added to the treatment. Hence, a threshold is reached in which common factor control groups attain equivalent outcome effects to specific treatment groups (Wampold, & Imel, 2015). Therefore, it stands to reason that experimental separation of any common factor, without the support of accompanying common factors (i.e., a full treatment), would generate an essentially inert condition. The effectiveness of Strupp and Hadley's (1979) untrained college professors, then, plausibly could be explained as being due to the sum effect of all other common factors being present in both conditions, which after all, was an explicit aim of that study. The aim of the present study was to adapt the Strupp and Hadley (1979) design, but instead of identifying technical training as the experimental group of interest, we focused instead on specifying the *interpersonal skills* of therapists as the experimental manipulation.

The Present Study

The rationale for the present study was based on the accumulation of findings that (a) individual therapist effects exist, (b) common relational processes repeatedly correlate with therapy outcomes, (c) attempts to isolate therapist techniques from common relational factors have not been successful, highlighting instead (d) the unexplained variation of those relational variables among therapists within those studies, and (e) individual case observations of even highly trained therapists show wide variations in basic interpersonal skills. Thus, we reasoned that specifying therapist interpersonal skills might be identified prospectively and without influence of client interactions. We refer to these as therapist facilitative interpersonal skills (FIS), which include the ability to effectively understand and send interpersonal messages as well as the ability communicate a rationale for another's problems and to propose new and effective solutions. For this study, we aimed to separate therapists into two basic independent groups based on their *FIS*. Therapists were selected prospectively, based on their level of social skills and performance on a therapy simulation task. In addition, therapists in the present study were also selected based on *Training Status*, which like Strupp and Hadley (1979), was included in the design for further specifying the common relational skills involved within a treatment setting. However, the primary focus of the study was on the effects of *FIS* and not a generic assessment of training. Training Status as a variable allowed us to assume that the therapists' interpersonal skills were *not* necessarily the result of training experiences.

The study was designed to be analogous to typical RCTs in psychotherapy, except that the critical independent variable was not a specific treatment or therapy manual, but instead was the therapists' common interpersonal skills (i.e., FIS). Instead of removing variation of individual therapists' relational styles, we encouraged and controlled this variation through pre-treatment selection. The opposite was true of training and technical abilities. We attempted to dampen the effects of prior training by selecting both trained and untrained therapist and encouraging all therapists to rely on whatever strategies they thought would be most helpful with their assigned clients.

We hypothesized that clients treated by high FIS therapists would have more improved outcomes compared to those treated by low FIS therapists. In addition to this main effect with FIS, we predicted that there would be an FIS × Training interaction whereby high FIS therapists with Training would have more improved outcomes compared to the

remaining 3 cells in the design (all low FIS and high FIS without training). We made no predictions in regard to the trained versus untrained therapists because of previous equivocal finding on Training Status.

Method

Participants

Therapists. Fifty-six doctoral students applied to participate as therapists and were selected based on their scores on measures of social skills and a performance analyses (see procedures). Of these, 23 (8 males; 15 females) were selected as therapists in the study. Eleven therapists were in a clinical psychology training program at a mid-western university and had completed at least two years of training. The remaining 12 therapists had no clinical or psychotherapeutic training, but had completed at least two years in doctoral programs in various other disciplines (3 from Biology, 1 from Chemistry, 2 from experimental psychology, 1 from History, 1 from Comparative Arts, 2 from Communication, and 1 from Human Sciences). Therapists ages ranged from 23 to 53 years (mean = 30.61 years; *SD* = 9.32). Therapists self-identified as 83% Caucasian, 13% Asian, and 4% Hispanic.

Clients. Clients in this study were selected from persons with significant distress and who qualified for a DSM IV psychological disorder. However, participants who served as "clients" differed from most psychotherapy clients in one important respect. The participants in the present study were selected from a general university sample and none were actively seeking clinical services when they were recruited into the study. Thus, the word "client" will be used to describe these individuals, although they were not actively seeking therapy services. Because of ethical considerations about asking persons with significant psychological problems meeting untrained "therapists," this study invoked protections of the participants that were similar to other psychotherapy studies on pseudotherapists (e.g., Strupp & Hadley, 1979; see procedures).

As seen in the CONSORT diagram (see Figure 1), 2,713 undergraduate students were screened with the Symptom Checklist—90—Revised (SCL-90-R; Derogatis, 1983) from a psychology participant pool at a moderate-sized university located in the Midwestern region of the USA. From those, 231 scored 2 standard deviations above the mean Global Severity Index (GSI) of the non-clinical standardized sample as well as above the mean for the outpatient sample. These participants were asked to return approximately one week later for a second administration of the SCL-90-R and, if symptoms remained above the 2 *SD* threshold, a diagnostic assessment interview was then conducted by one of 9 masters-level clinicians who had been trained in diagnostic interviewing using the DSM-IV. The clinical interviewers determined if the remaining participants' problems met criteria for at least one DSM Axis I or II disorder. The diagnostic interview excluded 92 (35.4%), who were judged to have sub-clinical or non-diagnosable problems. An additional 11 participants were also excluded for severe substance dependence, suicidal risks, and severe personality disorders. A small number of participants were excluded for various other reasons (see Figure 1).

A total of 82 (31.6%) met criteria and were randomized into the study of which 65 completed the study as clients. Forty-five of these clients were assigned to one of the therapist conditions and 20 clients were assigned to a wait list control condition. Demographic information about the clients is provided in Table I.

Measures

Outcome questionnaire—45 (OQ-45). The OQ-45 (Lambert et al. 2004), is a 45-item general symptom measure. The OQ-45 has sub-scale measures of subjective discomfort, interpersonal relationships, and social role performance. The measure has reasonably good internal consistency (alphas have ranged from .70 to .93; Ogles, 1996), and has the advantage of both brevity and for its treatment outcome focus. The OQ-45 was the only symptom measure administered at both evaluation and therapy sessions: Pre-treatment, sessions 1, 3, 5, 7, termination, and 3-month follow-up. Among the various time points in the present study, the OQ-45 was internally consistency (α = .72 to .79).

SCL-90-R (Derogatis, 1983). The SCL-90-R, containing 90 5-point items, is a multidimensional self-report symptom inventory. Each item is rated on a 5-point scale of distress, ranging from 0 ("not at all") to 4 ("extremely"). The GSI is the mean item score and reflects a patient's overall level of symptomatic distress. In this study, the GSI had excellent internal consistency (α = .94 and .97 at pre-treatment and termination; α = .96 at 3-month follow-up).

Inventory of interpersonal problems (IIP-64). The IIP-64 (Horowitz, Alden, Wiggins, & Pincus, 2000; Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988) is a measure of interpersonal distress and commonly used for measuring treatment changes in

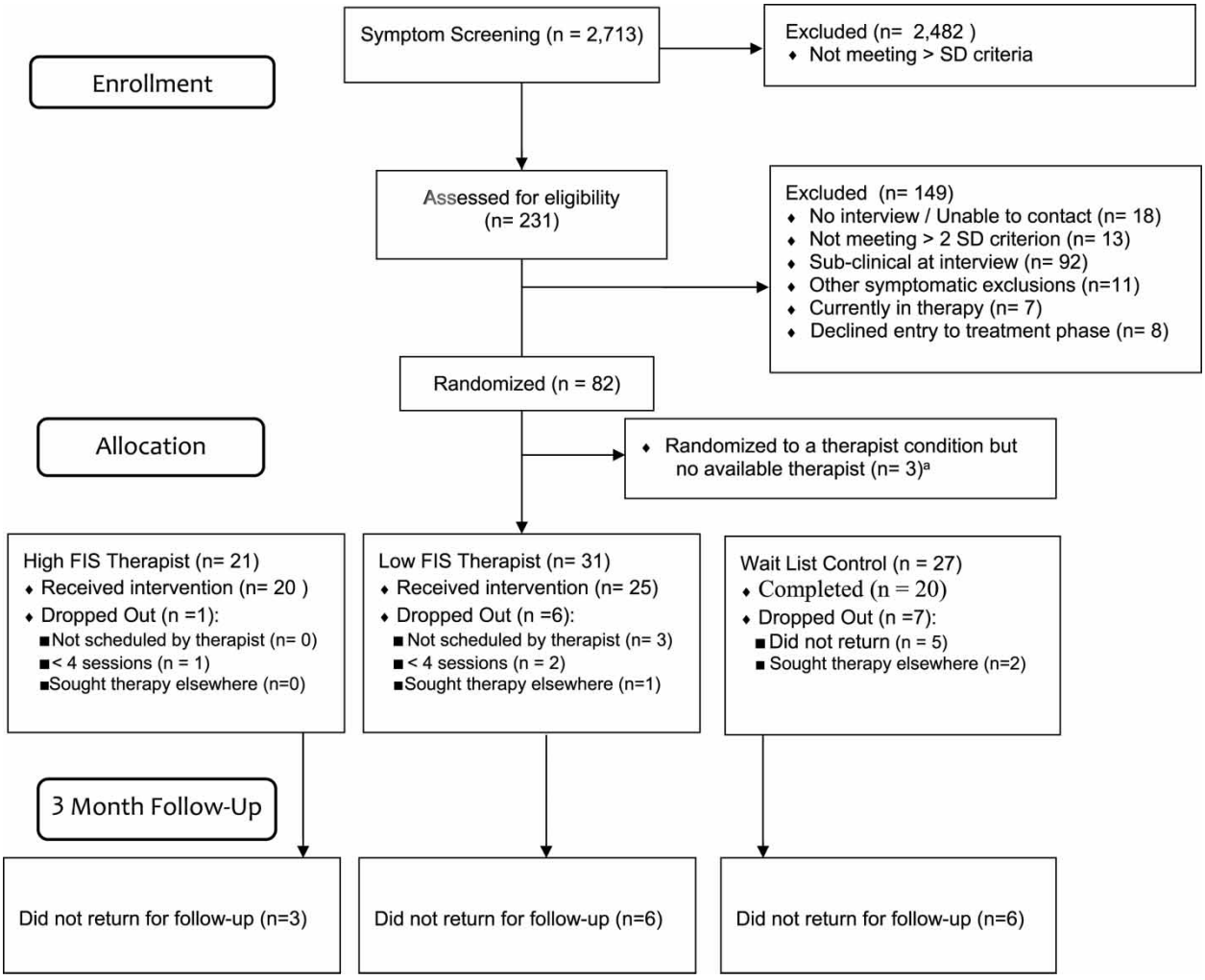

Figure 1. Phases of client enrollment, intervention allocation, and follow-up in the randomized trial.
Note: These three participants were assigned to one of the therapist conditions, but a specific therapist could not be selected or identified because there were no therapists available to receive these cases. These participants were thus given referrals and debriefed.

the interpersonal domain. The degree of distress associated with each item is rated on a 5-point scale, ranging from 1 (not at all) to 5 (extremely). The brief version of the 127 item IIP was used in the present study, the IIP-64 (IIP-CX; Alden, Wiggins, & Pincus, 1990), which balances item selection across 8 octants of the interpersonal circumplex (8 items each). Test–retest reliability for the IIP for a 10-week period has been reported at .98 for the overall inventory and internal consistency ranging from .82 to .93 (Horowitz et al., 1988). The IIP was administered at pre-treatment, session 3, termination, and 3-month follow-up. The present sample had good internal consistency as well ($\alpha = .93$ at pre-treatment and termination and $\alpha = .95$ at 3-month follow-up).

Global assessment scale (GAS). The GAS (Endicott, Spitzer, Fleiss, & Cohen, 1976), or otherwise referred to as the Global Assessment of Functioning scale or Axis V within DSM version III and IV, is commonly applied in diagnostic interviewing. The GAS is a single rating of overall psychological functioning, that ranges 0–100, where higher scores represent higher functioning. GAS ratings were made by the assessing clinician at pre-therapy and again at termination.

Target complaints (Battle et al., 1966). Target complaints were administered by the assessing clinician during the assessment interview. The advantage of this traditional measure of outcome was that it assessed specific client difficulties using vocabulary and concepts most similar to the client's form of thinking about their problems. Clients identified their three most significant presenting problems and assessing clinicians recorded these. Each of the problems was rated using a 5-point Likert scaling for

Table I. Client demographic and diagnostic characteristics of FIS, training and control groups.

	All clients meeting with a therapist/helper ($n = 45$)									
	FIS				Training status					
	High ($n = 20$)		Low ($n = 25$)		Training ($n = 21$)		No training ($n = 24$)		No treatment ($n = 20$)	
	n	Pct.	*n*	Pct.	*n*	Pct.	*n*	Pct.	*n*	Pct.
Sex										
Female	12	26.7%	8	17.8%	9	20.0%	17	37.8%	9	45.0%
Male	8	17.8%	10	22.2%	11	24.4%	7	15.6%	11	55.0%
Race										
African-American	0	0.0%	1	2.2%	0	0.0%	1	2.2%	0	0.0%
Asian/Pacific Islander	0	0.0%	1	2.2%	0	0.0%	1	2.2%	0	0.0%
Caucasian	20	44.4%	20	44.4%	20	44.4%	20	44.4%	20	100%
Hispanic	0	0.0%	2	4.4%	0	0.0%	2	4.4%	0	0.0%
Age	18.9 ($SD = 1.2$)		19.3 ($SD = 1.2$)		19.2 ($SD = 1.4$)		19.0 ($SD = 0.93$)		19.1 ($SD = 1.1$)	
Diagnostic grouping										
Adjustment disorder	4	8.9%	3	6.7%	3	6.7%	4	8.9%	5	25.0%
Major depression	4	8.9%	3	6.7%	2	4.4%	5	11.1%	3	15.0%
Dysthymia	5	11.1%	6	13.3%	6	13.3%	5	11.1%	4	25.0%
Generalized anxiety	2	4.4%	8	17.8%	5	11.1%	5	11.1%	4	20.0%
Misc. (phobia, panic, eating)	2	4.4%	4	8.9%	3	6.7%	3	6.7%	2	10.0%
Personality	3	6.7%	1	2.2%	2	4.4%	2	4.4%	1	5.0%

severity. The mean score for the three Target Complaints was used as an indicator of overall targeted distress. Target complaints were administered during the pre-treatment and termination clinical interviews.

Global outcome rating (GOR). The Global Outcome Rating provided a single estimate of the overall progress during the therapy (Bein et al., 2000; Strupp & Hadley, 1979). The GOR provided a broad and overall judgment of change itself rather a single time period assessment of subjective distress or functioning. GOR was rated on an 11-point scale ranging from −5 ("Very much worse") to +5 ("Very greatly improved"). Both the assessing clinician and the clients made independent GOR ratings at the termination assessment interview.

Working alliance inventory (WAI-C and WAI-T). The WAI (Horvath, 1981; Horvath & Greenberg, 1986), perhaps the most widely used and cited of alliance scales, contains subscales for measuring agreement on tasks, goals, and the existence of a therapeutic bond. Each sub-scale contains 12 items, which the participant rated on a 1 to 7 scale. Both therapist (WAI-T) and client (WAI-C) forms of the WAI were administered at sessions 1, 3, 5, and 7; the WAI-C also was administered at termination and at three-month follow-up. Across all sessions, the WAI-C had good internal consistency with alpha ranging from a low of .79 to a high of .90, and the WAI-T ranging from $\alpha = .80$ to .81.

Social skills inventory (Riggio, 1986). The SSI is a 90-item self-report questionnaire that assessed self-reported social skills. Items were scored using 5-point Likert scaling, from 1 = "not at all like me" to 5 = "exactly like me." The SSI measures skills in expressivity, sensitivity, and control in verbal (social) and non-verbal (emotional) domains. The total of the items provided an overall indicator of social skills, which was used in this study. The scale has high internal consistency and factor analytic studies have supported the multidimensional structure of the scale. Coefficient alphas range from .75 to .88. Test–retest correlations range from .81 to .96 for a two-week interval, and alpha coefficients range from .62 and .87 (Riggio, 1989). Convergent validity and discriminant validity for the SSI were supported in a series of studies conducted by Riggio (1986). In the present study, the SSI was completed by both therapists (as a selection variable) and clients in the study (at pre-treatment). The SSI had good internal consistency in the present sample ($\alpha = .88$).

Empathy and sociability (Gough, 1987). The Empathy and Sociability subscales were drawn from

the California psychological inventory (CPI) as additional measures of relational dispositions that might facilitate relatedness in therapy. Items on these sub-scales are endorsed as "True" or "False" and scoring is the sum of endorsed items (with some items reverse-scored).

The Empathy scale consisted of 38 items designed to measure dispositions, both overtly and implicitly, to adapt one's behavior to their needs and wishes. Hogan (1969) described persons with high Empathy as sociable, talkative, and outgoing, as well as having the ability to place themselves into another person's perspective to sense how they feel and think. Test–retest reliability over a period of two months has found the scale to be stable ($r = .84$) and to have acceptable Internal consistency (alpha = .69). Empathy has well-established validity, with Empathy significantly correlating with clinician-rated social acuity, various measures of social and interpersonal adequacy, and socio-political intelligence. Kendall and Wilcox (1980) found that Empathy was correlated with therapist effectiveness in treating hyperactive children.

The Sociability scale consisted of 36 items that are designed to identify persons who are outgoing and involved in social relationships. Test–retest reliability for the Sy scale has been shown to be between $r = .68$ and .71 and has acceptable Internal consistency (Gough, 1987). Sociability has been shown to correlate with related measures such as ego functioning, social desirability, sociability, and dominance. Therapist in the present study completed the Empathy and Sociability measures prior to beginning any therapeutic work with their clients.

Procedures

Therapist selection: Initial screening. Therapists were selected from 56 graduate students in various doctoral-level Ph.D. programs at a mid-western university. All prospective therapists responded to newspaper and/or direct e-mail advertisement. Screening was designed to select relatively equal blocks of participants in high versus low levels of FIS and Training Status (trained versus untrained). All 56 applicants were first screened with the Social Skills Inventory (SSI; Riggio, 1989). Those scoring in the upper or lower quartile of their gender-specific mean on the SSI (cutting out the middle 50% of the distribution) were asked to return for a performance-based screening of their interpersonal skills.

Performance task. The performance task was designed as a final threshold that therapists needed to reach in order to be included in the high FIS group. The rationale for including the performance task was that some prospective therapists might overrate their self-reported interpersonal skills. A slightly revised version of this performance task was described by Anderson et al. (2009), which required therapists to respond to 8 brief simulated therapy situations. These situations were drawn from a review of all third-session recordings from the Vanderbilt II psychotherapy project (Strupp, 1993). Segments were selected for the difficulty of the client–therapist interpersonal transactions. Segments were selected to represent various alliance rupture scenarios, including (a) a confrontational and angry client ("You can't help me"), (b) a passive, silent and withdrawn client ("I don't know what to talk about"), (c) a confused, yielding, but passively controlling client ("I have a tendency to put everything up so high, or not at all … "), and (d) an actively controlling and blaming client ("You should make it a point to be where you're supposed to be when you're supposed to be there!"). Actors were hired to re-enact the eight scenarios (two for each client), which were video recorded. Actors memorized the transcripts from the segments and were coached by the research team on how to capture the interpersonal style of the patients they were enacting. Clients were filmed by a camera that was directly facing the actors. Each recorded segment was approximately one-minute long.

Prospective therapists were presented with the 8 brief videos and were prompted to respond at predefined moments "as if" they were the therapist in the situation. The verbal responses from the prospective therapists were audio-recorded. Two licensed Ph.D. research clinicians rated each of the 8 recorded responses for each prospective therapist. Raters were provided with 10 items (also developed for the present study) for rating each of the 8 recorded responses. Item content was selected from the clinical and research literature on common therapist interpersonal skills and facilitative conditions. Specifically, items included ratings of warmth/acceptance, empathic accuracy, collaboration (alliance-bond capacity), problem focus, interpersonal responsiveness, verbal fluency, emotional expression, persuasiveness, hopefulness, and avoiding problematic complementarity. Performance analysis ratings were expressed as the mean item. Inter-rater reliability for the FIS total scores was acceptable (Intra-Class Correlation = .80).

The time that it took to process ratings of the performance analysis was longer than desirable for the logistics of scheduling therapists and clients in the study. Hence, we made initial therapist assignments based on the scores of the SSI and several weeks later completed the performance analysis. Once completed, the performance analysis ratings were used as

a final threshold for inclusion in the high FIS group. Thus, in order to be selected in the high FIS group, it was necessary to score in the upper quartile of the SSI and to score above the mean on the ratings of the performance analysis. Two prospective therapists had high SSI scores, but scored below the mean on the performance analysis. As these therapists had already begun working with clients for the study, they were retained, but bumped into the low FIS group. Finally, therapists completed the Empathy and Sociability subscales of the CPI (Gough & Bradley, 2002) for exploratory analysis. These measures were also completed prior to any assignment of clients.

Client and Therapist Preparation. Both client and therapist participants in the study were only partially fulfilling the role definitions for these terms. That is, "clients" were recruited into the study and were not seeking treatment or services from a professional therapist. During the initial assessment, it was clear to both clients and clinical interviewers that there was a significant problem and/or distress, even though clients had not sought professional help for those problems. Similarly, the untrained "therapists" in the sample had not been in the role of therapist before this study. Clients were informed that the purpose of the study was to understand the manner in which relationships between clients and therapists develop and facilitate change. Furthermore, clients were informed that in order to study these relationships, the research protocol was designed to minimize interference with the development of relationships with their therapist. All clients were paid for their participation in the study in order to further underscore that their role was defined as a research participant.

Therapists were prepared for their role through a brief (approximately 45 min) orientation to guidelines and procedures for the university psychology clinic, where all sessions took place. Therapists with training were also told to enter into a helper role with participants who had not initiated a request for services but agreed to discuss their problems for seven sessions. We use the terms client and therapist for convenience and because both groups met the majority of these role requirements. Thus, even untrained therapists worked out of our psychology clinic, scheduled clients, met for sessions with clients who were coming to sessions for help with psychological problems that met diagnostic criteria. Clients were informed of their confidentiality, the taping procedures, and other treatment options prior to the beginning of therapy. The university institutional review board approved the protocol for this study.

Several safeguards were put into place in order to protect participants from potential adverse consequences of discussing their problems with untrained or low skill therapists. (1) As stated in the consent form, all clients were informed that their relationship, while having numerous similarities to psychotherapy (e.g., meeting in a clinic), explicitly did not "involve 'psychotherapy' as it is legally defined or any medical treatment." (2) Clients received a list of genuine psychotherapy referrals, including the possibility of seeing a fully supervised therapist in the training clinic where the study took place. While few accepted referrals at the outset, several clients accepted referrals to meet within the clinic after they had completed the study. (3) All sessions were audio and/or video recorded and each session was reviewed by the principal investigator in order to assure that appropriate boundaries were maintained and to identify potential issues of client self/other harm. (4) Clients were told that they should not disclose anything unless they were fully comfortable doing so. (5) They were also informed that they could discontinue their sessions at any time, receive full payment, and still receive a referral.

Additional safeguards were set in place that involved the therapists. (6) Therapists were instructed about expectations for ethical conduct, which included following American Psychological Association ethical guidelines for therapist–patient relationships (e.g., confidentiality). (7) Therapists were also informed about the general parameters of typical therapy sessions and practice (e.g., 45–50-min sessions, note taking). Therapists were required to sign an agreement that they would follow these procedures.

Following these basic guidelines, therapists were asked to help their client in any way that they believed would be useful to alleviate the client's problems. However, no advice was provided in regard to specific techniques and treatment strategies. Each therapist was asked to directly contact the principal investigator in order to discuss any situations that they felt unable to address. Questions rarely arose, though there were two situations which required intervention with untrained therapists: (1) One of the therapists had begun having sessions with the door open and (2) one had wanted to take a copy of his notes home for further review.

Client assignment. During the initial interview, clinical interviewers were blind to client assignment of the independent groups (FIS, Training Status, or Control). Researchers attempted to keep those

conducting the clinical assessment interviewers blind to group assignments at the termination interviews. It appeared that they remained blind to FIS assignment; however, it became clear that clinical interviewers were aware when clients had not had treatment. Training Status assignment also appeared transparent to some of the clinical interviewers, especially when the therapist had no clinical training (i.e., based on the clients' reported experience). As client participants completed their clinical assessment interview, they were assigned through block randomization to the FIS × Training Status conditions. When therapists within a condition were fully assigned, the block was closed and the remaining open conditions were randomly assigned until all conditions were closed. Clients who were excluded from the study, regardless of the reason, were given referrals for psychological services. Whenever a participant had significant distress, they were encouraged to seek therapy and in such cases, we were more proactive in linking the excluded participant with an appropriate referral (e.g., helping the person make an initial appointment).

Therapy and assessment sessions. Approximately one week after the evaluation session, clients were assigned a therapist and began seven weekly meetings with their therapist. After sessions 1, 3, 5, and 7, clients were asked to complete measures evaluating their progress on the OQ-45 and the therapeutic relationship. Ninety-six percent of clients completed all seven of the offered therapy sessions. Both therapy and wait-list control participants completed measures at pre-treatment, termination, and three-month post-treatment follow-up. The termination evaluation was planned to occur 8 weeks after the baseline assessment, which translated to approximately one week after the last therapy session for those in the treatment conditions. Termination evaluations actually occurred between 0 and 11 days after the completion of the therapy (mean = 4.8 days; $SD = 2.9$) and also included a clinical interview. At the termination interview, all clients again were presented with referral options. Three-months post-treatment, a follow-up evaluation was conducted for therapy clients. The three-month post-treatment follow-up occurred approximately 20 weeks after the baseline assessment.

Statistical Analyses

Most of the outcome variables used in the study (i.e., GSI, IIP-64, GAS and target complaints) were collected only at the beginning and end points of treatment (i.e., pre-treatment, termination, follow-up). Thus, we conducted MANOVA inferential tests for these outcome measures in order to test the combined effect of multiple, similar measures taken at only a few time points. If there was a significant omnibus result, we conducted follow-up analyses utilizing repeated-measures ANOVA. For measures that were not collected at pre-treatment, but only at termination and 3-month follow-up (i.e., GOR), we conducted a univariate ANOVA. A power analysis for the repeated-measures MANOVA and within–between interaction indicated that the sample size for the therapist independent variables and therapy cases ($n = 45$) was sufficient for the present study.

For measures in which therapy session data were collected (i.e., OQ-45, WAI), we used hierarchical linear modeling (HLM) to model the change over time or the growth curve (Bryk & Raudenbush, 1992; Singer, 1998). These analyses examined the effect of the independent variables of (a) training status and (b) FIS on the intercept and slopes of the OQ-45 and WAI variables.

Finally, exploratory analyses utilizing HLM were used to examine the influence of various therapist measures, many of which were used to screen and assign therapists into the high versus low FIS categories, on OQ-45, WAI-C and WAI-T.

Results

Means and Standard Deviations for outcome measures at pre-treatment, termination and 3-month follow-up are provided in Table I. Descriptives and zero-order correlations for therapist measures used in this study are provided in Table II. All of these social and interpersonal measures were correlated at a moderate level and all but one of the relationships was at a high moderate range ($r = .59$ to $r = .68$).

There were no significant pre-treatment differences among any levels of both FIS and Training on any of the dependent measures. Therapist and client age, sex, and race were included as variables in the analyses of the outcome and alliance dependent variables and none emerged as significant.

A total of 14 clients discontinued before completing the required 4 session criterion to be included in the study. Interestingly, only one of these (7.1%) were assigned to the high FIS group, 6 (42.9%) were in the low FIS group, and 7 (50%) were in the wait-list control group. Three of the dropouts (2 low FIS; 1 control) discontinued because they had found a therapist elsewhere and an additional 3 (all low FIS) discontinued after speaking to their assigned therapist on the phone to schedule an initial

Table II. Descriptive statistics and zero-order correlations for therapist variables.

	Descriptive statistics						
Variable	*M*	*SD*	Zero-order correlations				
SSI	294.17	34.46	1.00				
FIS	2.99	0.52	.60**	1.00			
Emp	23.96	4.87	.43*	.68**	1.00		
Sy	24.00	5.20	.61*	.62*	.59*	1.00	
Age	30.61	9.32	.24	.17	.24	−.05	1.00
			SSI	FIS	Emp	Sy	Age

Notes: SSI, social skills inventory; FIS, facilitative interpersonal skills. Performance ratings; Emp, empathy; Sy, sociability.
*$p < .05$.
**$p < .001$.

appointment. There were no statistical differences between those who discontinued versus those who completed on either the 2 symptom screenings or the clinician-rated GAS. In terms of Training Status assignment, 2 of the dropouts had been assigned to therapists with clinical training, while 5 had been assigned to therapists with no clinical training.

Outcome

Pre- and post-treatment comparisons: FIS, training status, and control comparisons. Because numerous variables were available for the pre- and post-treatment evaluation sessions, we attempted to reduce Type I error, by conducting a doubly multivariate repeated-measures MANOVA analysis using Training Status and FIS as between-subject measures on the 5 outcome measures (GSI, OQ-45, IIP-64, GAS, & Targets) with Time as a within-subjects factor (pre-treatment and termination). Therapist training was not a factor in outcomes since there was no Training Status × Time interaction, $F(1, 41) = 0.45$, ns, nor was there a Training Status × FIS × Time interaction when including all outcome measures, $F(1, 41) = 1.54$, ns. Since Training Status was not significant in the omnibus analysis, it was excluded from additional analysis.

For all measures, there was a significant omnibus FIS × Time interaction, $F(1, 41) = 6.75$, $p = .01$ ($\eta^2 = .14$), indicating that across all outcome measures, therapist FIS was a significant factor in client improvements. Follow-up analyses, using repeated pre–post ANOVAs, were conducted in order to better understand the effect of FIS on GSI, OQ-45, IIP-64, GAS and Targets. In addition to that, univariate ANOVAs were conducted to investigate the effect of FIS on GOR (both client-rated and clinician-rated). As seen in Table III, at post-treatment, each outcome measure was significantly different among the high FIS, low FIS, and no-treatment groups.

Analysis comparing high FIS, low FIS, and control groups are presented in Table IV. In terms of comparisons to the no-treatment control group, those in the high FIS group had substantially improved outcomes at the post-treatment evaluation session for all outcome variables, except for the IIP-64 in which there was a non-significant trend ($p = .07$). Those treated by low FIS therapists had significantly improved outcomes on the GSI and client-rated GOR, but were no different than the no-treatment control for the remaining 5 outcome indicators (OQ-45, GAS, clinician-rated GOR, Targets, and the IIP-64).

Pre- and post-treatment comparisons: High and low FIS. The more interesting comparisons of the study were between the high versus low FIS groups (presented in the left column of Table IV). There were significant differences in the changes made by those participants seen by high FIS therapist relative to the changes made by participants seen by low FIS therapists on the OQ-45, GAS, and the IIP-64. Those in the high FIS group also trended toward greater improvements than their low FIS counterparts on Target Complaints and the Client-rated.

Outcome at 3 months post-treatment. Results from the doubly multivariate repeated-measures MANOVA using pre-treatment, termination and 3-month follow-up periods for Therapist FIS on the GSI, OQ-45, and IIP-64 (GAS and Targets were not collected at follow-up) found a Time × Therapist FIS interaction, $F(2, 32) = 3.54$, $p = .04$, a moderate effect ($\eta^2 = .18$). (Footnote: A similar level of significance and effect is found when this analysis is conducted using only pre-treatment and 3-month follow-up periods (excluding the termination period, $F(1, 33) = 6.61$, $p = .02$, $\eta^2 = .17$).

Similar to the pre-treatment to termination results, repeated-measure ANOVAs on each of the outcome measures (GSI, OQ-45 and IIP-64)

Table III. Means, standard deviations and effects for FIS and no-treatment groups on outcome measures.

								MANOVA tests							
		Pre-treatment		Post-treatment		3 Months post-treatment		Post-treatment				3 Months post-treatment			
Measure	Treatment	*M*	*SD*	*M*	*SD*	*M*	*SD*	df	*F*	η^2	ES	df	*F*	η^2	ES
GSI	High FIS	1.83	0.57	1.14	0.67	1.09	0.58	2,61	3.49*	.10	0.96	2,44	0.56	.02	1.09
	Low FIS	1.87	0.56	1.15	0.45	1.11	0.54				1.23				1.17
	Control	1.56	0.42	1.31	0.42	1.09	0.53				0.52				0.45
OQ-45	High FIS	87.75	18.53	60.75	27.22	59.35	25.56	2,62	3.90*	.11	1.02	2,47	3.34*	.13	1.11
	Low FIS	81.48	10.39	67.88	17.20	66.42	14.78				0.85				1.03
	Control	80.80	18.98	71.30	16.91	71.42	21.08				0.46				0.41
IIP-64	High FIS	113.35	28.63	88.35	41.92	85.59	38.20	2,62	4.71**	.13	0.89	2,47	3.49*	.13	0.92
	Low FIS	100.80	22.08	96.08	27.07	96.05	32.07				0.24				0.60
	Control	110.65	27.28	98.25	27.67	100.85	38.45				0.56				0.33
GOR-Client	High FIS			3.70	1.30	3.06	1.52	2,62	7.33**	.19		2,40	1.31	.06	
	Low FIS			2.90	1.41	2.53	1.77								
	Control			2.05	1.36	1.78	2.77								
GAS	High FIS	59.05	4.42	71.00	8.15			2,62	6.99**	.18	1.77				
	Low FIS	60.72	4.77	68.24	7.76						1.11				
	Control	62.50	6.17	66.75	6.13						0.64				
GOR-Clinician	High FIS			3.30	1.45			2,62	4.58**	.13					
	Low FIS			2.24	1.30										
	Control			1.90	1.86										
Targets	High FIS	3.75	0.56	2.47	0.80			2,62	3.65*	.11	1.71				
	Low FIS	3.75	0.42	2.88	0.82						1.27				
	Control	3.75	0.44	3.12	0.86						0.88				

Note: ES = Cohen's *d* statistic with correction for correlation of each within-subject measures (Morris & DeShon, 2002)
*$p < .05$.
**$p < .01$.

Table IV. Comparisons of high FIS, low FIS and no treatment on outcome measures.

		Contrast								
		High FIS versus Low FIS			High FIS versus no treatment			Low FIS versus no treatment		
Measure	Within group	df	*F*	η^2	df	*F*	η^2	df	*F*	η^2
GSI	Pre–post	1, 43	0.97	.02	1, 36	5.70*	.14	1, 41	3.96*	.09
	Pre-3 mos.	1, 34	0.07	.00	1, 26	1.02	.04	1, 28	0.73	.02
OQ-45	Pre–post	1, 43	4.53*	.10	1, 38	5.00*	.12	1, 43	0.68	.01
	Pre-3 mos.	1, 34	4.63*	.12	1, 29	4.53*	.14	1, 31	0.29	.01
GAS	Pre–post	1, 43	4.92*	.10	1, 38	13.23***	.26	1, 43	3.02	.07
GOR-Clinician	Post-treatment	1, 43	6.65**	.13	1, 38	7.03**	.16	1, 43	0.52	.01
GOR-Client	Post-treatment	1, 43	3.81†	.08	1, 38	15.41***	.29	1, 43	4.16*	.09
	3 mos.	1, 32	0.87	.03	1, 24	2.53	.09	1, 24	0.71	.03
Targets	Pre–post	1, 43	3.43††	.07	1, 38	7.00**	.16	1, 43	0.96	.02
	Pre-3 mos.									
IIP-64	Pre–post	1, 43	8.24**	.16	1, 38	3.52††	.09	1, 43	0.85	.02
	Pre-3 mos.	1, 34	6.06*	.15	1, 29	3.68††	.11	1, 31	0.11	.00

Note: Comparisons are repeated-measures ANOVA on each measure except for GOR, which is a univariate ANOVA.
†$p = .06$.
††$p = .07$.
*$p < .05$.
**$p < .01$.
***$p < .001$.

found that Therapist FIS was significantly different for the OQ-45 and the IIP, but not on the GSI (see Table IV). A univariate ANOVA conducted on client-rated GOR (clinician-rated GOR was not collected at 3-month follow-up) showed a non-significant result. For comparisons involving the no-treatment control group, those in the high FIS group had greater improvements on the OQ-45 and a non-significant trend on the IIP-64. In contrast, the participants seen by low FIS therapists were not any more improved than those who were in the no-treatment group. For High versus low FIS group comparisons, Clients in the high FIS group had significantly improved outcomes on the OQ-45 and the IIP-64 (see Table IV).

Session-series analyses. An HLM with linear growth curve was fitted on OQ-45 data, where level 1 included these sessions: Pre-treatment, sessions 1, 3, 5, 7, termination, and 3-month follow-up, and level 3 included therapist variables of Training status and FIS classification. We started with a model that included Training Status × FIS and Training Status × FIS × Session interactions terms. The result showed that there were no significant Training status × FIS classification and Training status × FIS classification × Sessions interactions, and thus both terms were dropped from subsequent analyses. Results indicated that the overall intercept and the overall slope or rate of change were significant, F (1, 42) = 1298.88, $p < .0001$ and F (1, 42) = 47.76, $p < .0001$, respectively. Neither Training Status nor the FIS classification significantly affected the intercept, F (1, 213) = 0.28, *ns* and F (1, 213) = 0.42, ns, respectively.

Training status did not affect the slope or the rate of change of the OQ-45 (Training Status × Sessions was not significant), F (1, 213) = 0.14, ns.

FIS classification significantly differentiated symptom changes across the sessions, F (1, 213) = 7.09, $p = 0.01$. Figure 2 displays these differences on FIS. Descriptively, clients' began treatment with OQ scores above the standard cut-off of clinical significance of OQ > 63 with an intercept of OQ = 77.1 (intercept was centered at the pre-treatment session). Across all clients, the rate of change was 1.00 OQ points per session (SE = 0.14). Clients who were seen by therapists with low FIS improved by decreasing 0.62 OQ points per session (SE = 0.20), whereas clients seen by high FIS therapists dropped 1.39 OQ points per session (SE = 0.21).

Therapeutic Alliance

Client-rated alliance. Because session data were available for the WAI-C, an HLM with a linear growth curve was conducted where level 1 included these sessions: Sessions 1, 3, 5, 7, termination, and 3-month follow-up, and level 3 included therapist variables of training status and FIS classification. As in the analysis for OQ data, we started with a model that included Training status × FIS classification

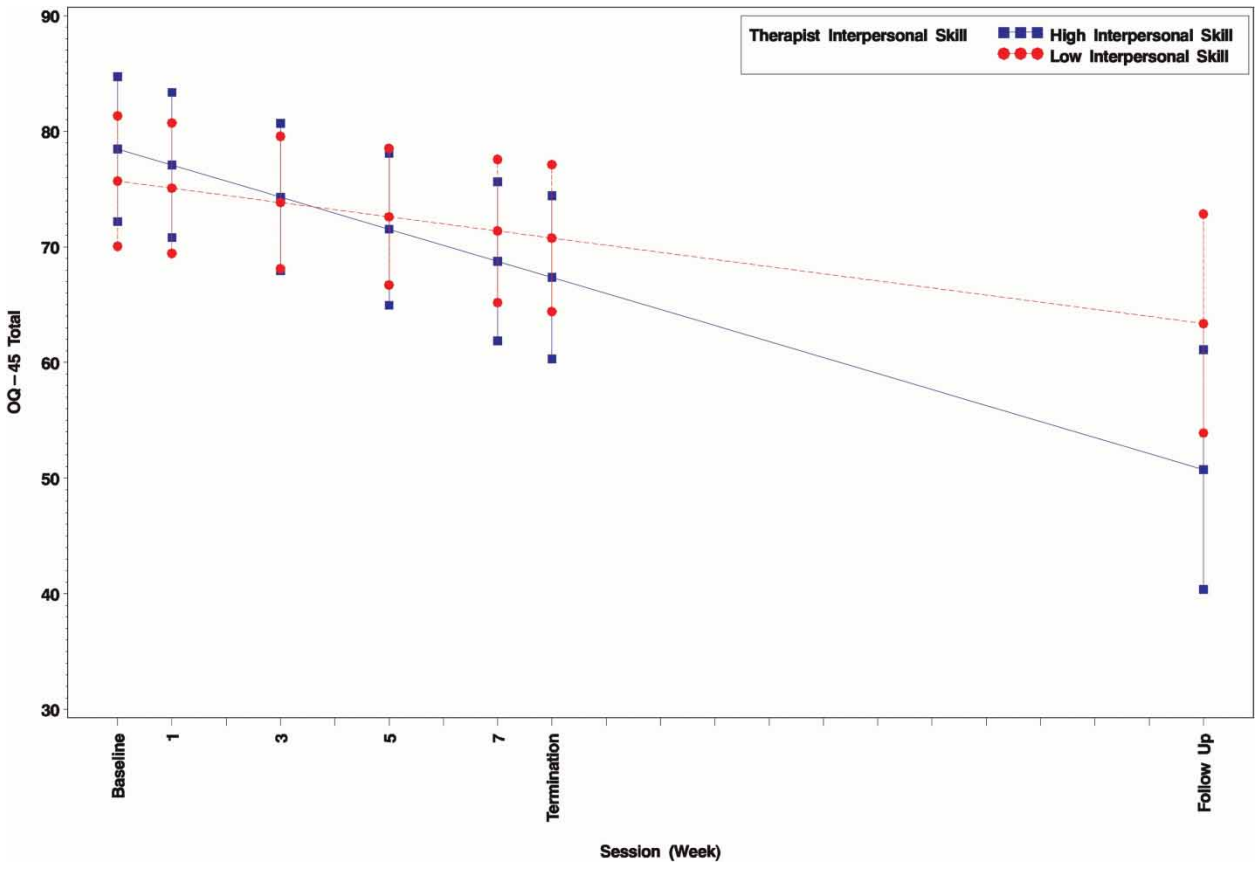

Figure 2. Therapist FIS on OQ-45 across sessions.
Note: Figure displays slopes by grand mean and standard error at each assessment and therapy session. Therapy sessions are numbered 1, 3, 5, and 7. Pre-Treatment was one week prior to the first therapy session, termination was one week after the last session, and follow-up was 12 weeks after the termination session.

and Training status × FIS classification × Sessions interactions terms. The result showed that there were no significant Training status × FIS classification and Training status × FIS classification × Sessions interactions, and thus both terms were dropped from subsequent analyses. Results indicated that the overall intercept and the overall slope (rate of alliance change) were significant, F (1, 41) = 2668.76, $p < .0001$ and F (1, 41) = 6.70, $p < .05$, respectively.

Training Status did not significantly affect the intercept, F (1, 162) = 3.14, ns; however, the FIS classification significantly affected the intercept, F (1, 162) = 3.92, $p = 0.05$. In terms of the slope or rate of change, the Training status did not account for the slope on the WAI-C (time × training), F (1, 162) = 0.02, ns.

The intercept (WAI-C score at session 1) for the high FIS therapist was 212.55 (SE = 5.97) and was approximately 15 scores lower for the low FIS therapist at 196.81 (SE = 5.23). However, as seen in Figure 3, there were differences in the alliance rate of change on the FIS classification, F (1, 162) = 5.02, $p = .03$, with high FIS clients gaining 1.28 per session (SE = 0.40) on the WAI-C compared to a gain of 0.09 per session (SE = 0.35) for the clients of low FIS therapists.

Therapist-rated alliance. Because the WAI-T only included sessions 1, 3, 5, 7 (since therapists only rated clients after they met with the client), an HLM with a linear growth curve was fitted with level 1 included in these sessions: Sessions 1, 3, 5 and 7, and level 3 included therapist variables of training status and FIS classification. As in the analysis for OQ and WAI-C data, we started with a model that included Training status × FIS classification and Training status × FIS classification × Sessions interactions terms. The result showed that there were no significant Training status × FIS classification and Training status × FIS classification × Sessions interactions, and thus both terms were dropped from subsequent analyses. Results indicated that the overall intercept and the overall slope or rate of change were significant, F (1, 42) = 3086.91, $p < .0001$ and F (1, 42) = 28.34, $p < .0001$, respectively.

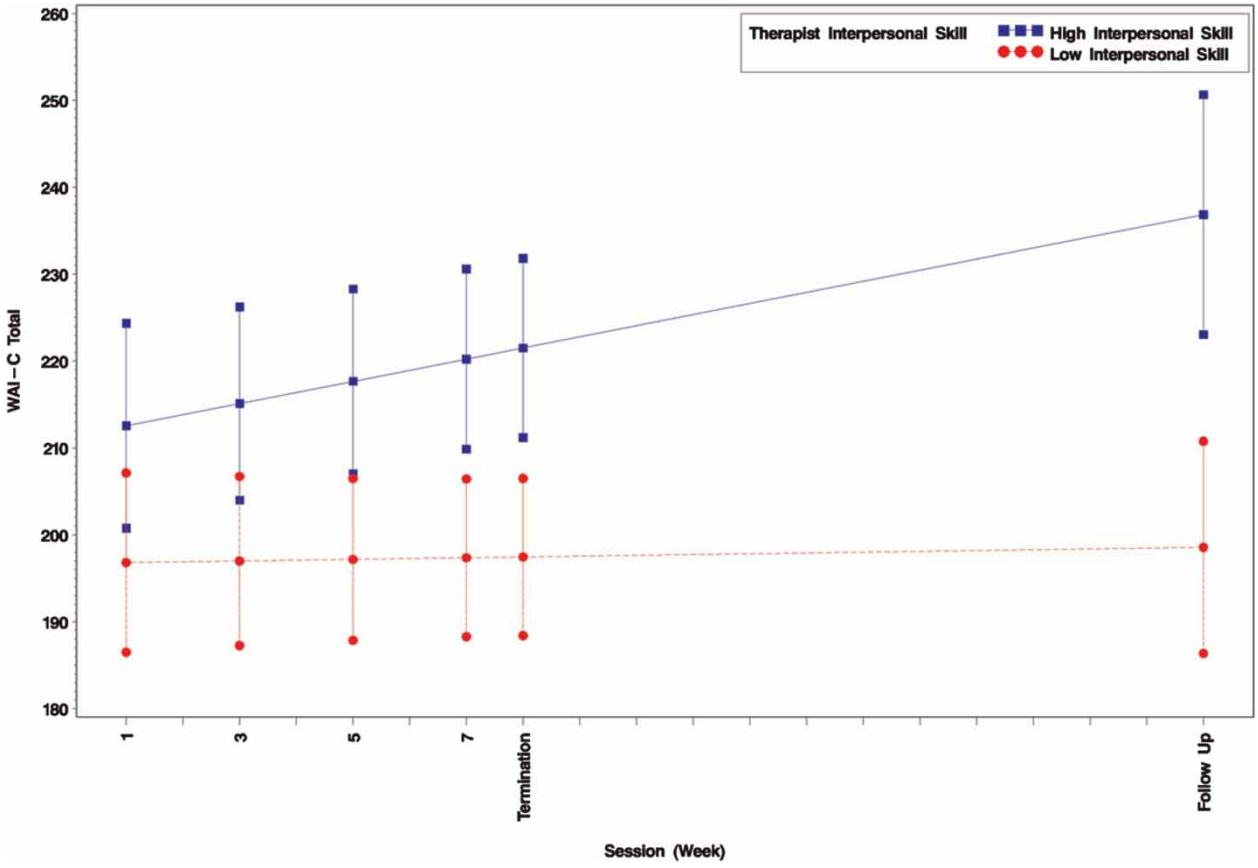

Figure 3. WAI-C across sessions by therapist FIS.
Note: Figure displays slopes by grand mean and standard error at each rating period. Sessions measured on the WAI-C include therapy sessions 1, 3, 5, and 7 as well as the termination (1 week post-treatment) and Follow-up evaluation (12 weeks post-termination) periods.

In terms of the intercept, there was no significant effect of Training status on intercept, F (1, 86) = 1.24, ns. However, in terms of rates of alliance change, there was a significant Training status × Sessions effect, F (1, 86) = 5.24, p = .02, with therapists with clinical training rating the alliance increase by 3.85 WAI-T points per sessions (SE = 0.74) compared to an increase of 1.54 WAI-T points per session (SE = 0.69) for those with no clinical training.

There was a significant effect of FIS classification on intercept, F (1, 86) = 11.26, p < .01. The intercept level (i.e., session 1) on the WAI-T for high FIS therapists was 190.62 (SE = 4.81) and approximately 20 points lower for low FIS therapists at 168.90 (SE = 4.33). However, FIS classification did not influence the rate of change on WAI-T, F (1, 86) = 2.35, ns.

Exploratory Analyses with Therapist Variables

Exploratory analyses of the therapist screening measures were conducted in additional mixed-model analyses of the OQ-45, WAI-C, and WAI-T. For these analyses, therapist variables were entered separately as Level 3 predictors in replacement of the Training Status and FIS variables. There were two sets of therapist variables tested in the exploratory analyses; one set consists of Social Skills (SSI) and FIS Performance Ratings and the other set consists of the Empathy and Sociability scales from the CPI. Two separate HLMs analyses were conducted, one for each set of therapist variables with non-significant variables dropped from the final model.

Therapist relational skill variables. Therapist SSI was a significant predictor of decreases in symptoms of the OQ-45 across sessions, F (1, 213) = 4.15, p = .04, as well as increases in WAI-T, F (1, 86) = 5.99, p < .02, but did not predict the rate of change in WAI-C.

However, the FIS performance ratings did not predict changes on the OQ-45 and alliance scores.

Therapist dispositional variables. Therapist Empathy predicted rate of change on the OQ-45, where higher empathy scores marginally, but

significantly, predicted decreased symptoms, F (1, 213) = 4.05, p = .05. Therapist Empathy also accounted for increased slopes on the WAI-C scores across all sessions, F (1, 162) = 5.69, p = .02. Therapist Sociability did not predict rates of change in OQ-45 or WAI-C, but was predictive of a linear increase in WAI-T scores across sessions, F (1, 86) = 26.38, p < .0001.

Discussion

Therapists high in FIS treated clients who improved more in their therapy outcomes than clients treated by low FIS therapists. These changes were observed by most of the outcome measures, including client ratings of subjective distress (OQ-45, IIP-64, and GOR) as well as independent clinician assessments (GAS, GOR, and Target Complaints). However, differences between the low and high FIS groups did not emerge on the GSI. In addition, the client-rated GOR and Target Complaints were only marginally different between the low and high FIS therapists. In terms of rate of change, the only outcome measure that was administered across both therapy and assessment sessions (the OQ-45) showed a significantly greater rate of change for clients treated by the high FIS therapists compared to those treated by low FIS therapists.

Results from the full omnibus analysis allowed for some anchored comparisons to a "silent" no-treatment control group. The primary differences among these groups were due to the high rate of change among those seen by high FIS therapists. In fact, compared to the no-treatment control group, those seen by high FIS therapists significantly improved at termination on all of the outcome measures, whereas those seen by the low FIS therapists displayed significant differences from the no-treatment control group on only two of the seven outcome measures. For several of the outcome measures, those seen by low FIS therapists had nearly the same outcomes as those who were in the no-treatment control group.

Similarly, measures of the alliance differed by FIS condition. High FIS therapists had higher client-rated alliances from the first session of therapy. Furthermore, the alliances for clients with high FIS therapists continued to increase throughout the course of their treatment, whereas clients with low FIS therapists did not significantly increase over the course of their therapies. High FIS therapists also rated their alliances as higher than low FIS therapists from the first session of treatment, but both high and low FIS therapist ratings remained flat throughout the remainder of therapy. Overall, the findings provide evidence of a therapist effect on outcomes and the therapeutic alliance, and this effect is specific to therapist pre-existing relational skills.

These findings converge with findings on both therapist effects on outcome (e.g., Okiishi et al., 2003; Wampold & Bolt, 2007) as well as relational influences on outcome in process research (e.g., Norcross, 2011). Because numerous findings from process research show that the optimal conditions within therapy sessions, which most correlate to outcome, include the presence of empathy, alliance, warmth and positive regard (see Norcross, 2011), it was natural to assume that therapist interpersonal abilities that would reflect and include these general interpersonal abilities would be related to the creation of those positive processes and outcomes within therapy. The current findings lend support to these long-held assumptions in research as well as time-honored theories about what therapist characteristics are most likely to enhance these optimal facilitative conditions within therapy sessions and lead to improved outcomes.

These differences in therapist FIS were unrelated to Training Status. In fact, the absolute mean differences between the training levels on outcome measures were exceedingly small. The absolute value of the clients' ratings of the alliance, though not statistically significant, were actually higher for the *untrained* therapists (the greatest difference being at session 5, where the difference was at p = .07, but again, the omnibus test across sessions was non-significant). Interestingly, the only Training Status difference that we found was that therapists with clinical training rated a faster rate of change across sessions in their therapeutic alliances compared to therapist without clinical training. One possible explanation for the significant increases of the trained therapist alliance ratings is that those receiving training had heightened awareness to the alliance construct from their professional training and hence were more attentive to taking actions that would improve the alliance. These trained therapists may have been more sensitive to the nuances of alliance ruptures within the relationship too, and hence may have more quickly perceived and acted on these relational tensions in ways that untrained therapists might not have perceived. The fact that their clients were not reflecting those increases in their ratings of the alliance leads us to speculate whether the trained therapists were (a) more aware of the alliance and taking some corrective action to improve the alliance, though perhaps not in ways recognizable to their clients, or (b) taking actions that they believed would improve the alliance, but in doing so, they may not have been appropriately responsive to their clients' needs (e.g., through poor timing of interventions).

The lack of Training Status effects on outcomes and alliance is consistent with other findings in the literature (e.g., Christensen & Jacobson, 1994). While no conclusions are drawn about training from this study, there are several possible research issues to consider for this *null* finding. First, the current study was not intended to provide a general assessment of training since our focus was on therapist relational abilities, a point which Strupp (1998) made about the "Vanderbilt I" findings. Second, our focus was on therapist FIS and while therapists had some training, even the "trained" therapists had not completed professional training when they participated in the study. Third, the "Training Status" variable could not capture the effects of technique because there was no specific technical intervention that the therapists were being asked to perform. In fact, our aim was to equalize technical factors in an analogous way in which most RCTs remove therapist effects, and our instructions to the therapists were generic. As a preliminary study on the effects of therapist skill, our goal was to minimize the effects of the technical operations and therefore it should not be surprising that there was a lack of a training effect. We remain optimistic that more controlled studies with much larger samples of therapist effects, including careful measurement of therapists' pre-existing interpersonal aptitude and skill, will ultimately lead to identification of significant training effects in future psychotherapy studies. For example, untangling what specific circumstances in which pre-existing, untrained interpersonal skills may be sufficient for *some* clients and problems may contribute to our understanding of how more specific techniques and advanced training may be required for *other* clients and problems. In conclusion, our finding regarding Training Status seems to underscore the explanation provided by Strupp and Hadley (1979), who concluded that client improvements, regardless of whether treated by trained professional or untrained therapists, were

> generally attributable to the healing effects of a benign human relationship. More specifically, therapeutic change seemed to occur when there was a conjunction with a patient who was capable of taking advantage of such a relationship (i.e., not too resistant and highly motivated) and a therapist whose interventions were experienced by the patient as expressions of caring and genuine interest. (p. 1135)

Limitations

The interpersonal skills used to identify therapists were also broad and lacking specificity. Furthermore, this study did not attempt to identify the specific in-session interpersonal behaviors that might have been present with these therapists. Thus, we cannot identify what specific interpersonal behaviors, skills or traits might have influenced the outcomes and alliances of these cases. For example, the moderately high correlations among the therapist relationship measures in this study might imply that these interpersonal characteristics may be challenging for more exact specification. A similar issue of overlap exists among relational process variables, such as empathy and the alliance. However, the broad-based definition of therapist interpersonal skills was intentional. By analogy, it is increasingly common for RCTs of specific treatments to examine a myriad of techniques and modules, such that it is rarely possible to know the specific components that were effective. Perhaps future study on relationship factors might benefit by using a similar strategy of aggregating several relational characteristics. Hence, we know that the best process predictors of outcome (alliance and empathy) are also broadly defined and are interdependent (Norcross, 2011). It will be important for future research to not only replicate these findings, but to identify more specific therapist characteristics as well.

Similarly, the study did not provide observational measurement of what interpersonal skills were used within therapy sessions and thus we cannot assume that therapist interpersonal skills were what led to the client outcomes. Future study should focus on identifying both therapist skills before treatment (independent of their client's behavior) as well as therapist skills within sessions that include responsiveness (and potential interdependence) to their client's behavior.

A related limitation is that we used a broad measure of social skills as our primary selection of therapists as well as an additional threshold of a performance task for prospective therapists entering the high FIS group. The advantage of using broad self-report measures of social skills was that the measures likely captured a wide range of relational skills. The performance-based task provided a check on self-report biases that might be expected to exist among prospective therapists and helpers. However, we also discovered numerous limitations in our approach. When these therapist variables were used in analyses as continuous variables in exploratory analyses, the SSI emerged as a significant predictor of changes on the OQ-45, but the ratings from performance task did not predict changes. Interestingly, a recent study with these same measures (Anderson et al., 2009) found the opposite: The performance analysis ratings

predicted changes on the OQ-45, but the SSI was not a significant predictor in that study. There are several possible reasons for the different findings in the two studies. First, analysis of the SSI and the performance analysis were post-hoc and exploratory. Second, it is also unclear how removing the middle range of the SSI distribution, leaving a bifurcated distribution, might have affected these analyses on other therapist variables. Interestingly, we discovered that the performance analysis was more normally distributed than the SSI. It seems likely that raters of the performance analysis in this study might have used other cases for comparison, which might explain why the ratings were more normally distributed. Third, the ratings for the performance analysis from the current study had not been developed into a separate measure of therapist skills for the current study (which predated Anderson et al., 2009). For the current study, the performance task was developed in order to provide an additional threshold for the prospective therapists to cross before being considered for entry into the high FIS group.

Other methodological limitations also detract from the generalizability of these findings. First, the relatively small sample size limited our ability to draw conclusions from interactions. The fact that the study selected therapists from the tails of the distribution on the SSI might also lead to speculation that these findings are limited to more extreme examples of relational skills. However, these extremes are reflective of actual therapists who had received at least two years of clinical training, including psychotherapy (and who did not differ on the relational constructs from those in the sample who were untrained). Another limitation is that the measure used to select clients for the initial screening (SCL-90-R) was not repeated after therapy sessions (as was the OQ-45). Furthermore, it is of concern that this same measure was one of the only measure for which there were no differences between high and low FIS therapists. However, clients on the other outcome measures had pre-treatment scores in the clinical range. On the OQ-45, which was used as a repeated measure across therapy sessions, the pre-treatment mean score was far above the dysfunctional cutting score for a university sample and was more similar to clients beginning treatment in a community clinic. The extremely high OQ-45 scores within a university sample also could contribute to creating other unknown systematic confounds with such an unusual, and selectively sampled, group. Of course, the same limitation would also be true of clients in treatment studies, which leads to the next important limitation.

Because "clients" in the present study were selected, they were not like genuine therapy clients, in that they were not seeking treatment at the time that they arrived for their assessment interviews. This was by design, because of ethical concerns. However, the participants in the study were similar to "real" clients, in that they were highly distressed and actually were more distressed, based on pre-treatment scores, than typical university students who seek treatment at a university clinic or counseling center. The high levels of reported distress, combined with the apparent lack of treatment-seeking behavior, made this sample somewhat unique. The participants likely were similar to those for whom outreach programs and educational interventions about psychotherapy are targeted. Thus, it is possible that therapist relational skills might be more effective for "clients" who might expectedly be more cautious about seeking psychotherapy services than those clients who are actively help seeking.

Conclusion

Nonetheless, the relational therapist effects from this study are consistent with the preponderance of studies demonstrating the predictive effects of relational characteristics within therapies. Because relational variables are commonly the strongest predictors of therapeutic outcomes, at some point it would be worth considering giving greater credence to systematically selecting therapists for professional training who are most likely to develop those facilitative conditions in their treatments. In most clinical programs, those qualities have taken a back seat to cognitive abilities and measures of traditional academic performance. The results of this study give us pause in the assumption that these relational abilities in our graduate students are "self-selected," meaning that students most interested in therapy are also most likely to display the relational characteristics that are most likely to result in these "relational pre-requisites" to skillful delivery of specific treatment models in therapy.

Future research could more accurately assess the applied usefulness of relational skills by selecting therapists on their relational abilities as part of a study that also examines a specific mode of treatment. An optimal choice for the treatment would be one that is designed to enhance the very relational skills that are selected on FIS, or the relational skills selection variable. For example, a treatment specifically designed for enhancing the therapeutic alliance (Crits-Christoph et al., 2006), Brief Relational Therapy for addressing alliance ruptures (Safran &

Muran, 2000), Time Limited Dynamic Psychotherapy (Strupp & Binder, 1984), and Emotion Focused Therapy (Greenberg, 2015) are excellent prospects as treatments in which therapist FIS would be expected to enhance the technical and training effects of these treatment manuals. Furthermore, high FIS therapists might be expected to accelerate at a more successful rate in acquiring helping skills (e.g., Hill, 2009) during structured training.

Common factors, contextual, and generic theories of psychotherapy would predict that any treatment should be effective so long as both relational and technical factors are both *interacting in harmony* (Wampold & Imel, 2015). As has been repeatedly noted, the relationship factors that are commonly assumed to be the central core of the therapeutic alliance are constantly integrated with and informed by "the treatment" (Hatcher & Barends, 2006; Horvath et al., 2011; Wampold, 2007) as well as other factors from contextual models of psychotherapy (Frank & Frank, 1993; Orlinsky & Howard, 1986). Furthermore, processes such as the therapeutic alliance are relationship based, but are strongly linked to skillful use of the techniques and principles of treatment (Hatcher & Barends, 2006; Wampold, 2007). A similarly contextual point is that this study differed from typical RCT designs in contemporary treatment research, in that the relationship variable (FIS) was defined as the independent variable instead of a "treatment" defined with specific techniques and strategies. Our design rationale had been to experimentally control a therapist relationship factor in a way analogous to how treatment is controlled in most RCTs. There is little doubt that therapists' relational abilities cannot be fully independent from the treatment context, including client characteristics, the setting, the treatment modality used, the therapeutic relationship, as well as extra-therapeutic events. Still, attempts at scientific control of therapist relational skills might provide one lens to understanding how professional (and even some nonprofessional) healing relationships can bring hope and ameliorate suffering in those who are psychologically distressed.

Acknowledgements

We are grateful to several persons who helped with this project, which we referred to as the Ohio University Helping Relationship Study (OUHRS). Special thanks are extended to Karen Deisseroth, Andrew Weiss, Marilyn McCabe and Kimi Carson. In addition, we are thankful to the several clinical interviewers and the therapists who worked on this project.

References

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, *55*, 521–536.

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology*, *65*, 755–768.

Anderson, T., & Strupp, H. H. (1996). The ecology of psychotherapy research. *Journal of Consulting and Clinical Psychology*, *64*, 776–782.

Baskin, T. W., Tierney, S., Minami, T., & Wampold, B. E. (2003). Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology*, *71*, 973–979.

Battle, C. C., Imber, S. D., Hoehn-Saric, R., Stone, A. R., Nash, E. R., Frank, J. D. (1966). Traget complaints as criteria of improvement. *American Journal of Psychotherapy*, *20*, 184–192.

Bein, E., Anderson, T., Strupp, H. H., Henry, W. P., Schacht, T. E., Binder, J. L., & Butler, S. F. (2000). The effects of training in Time-Limited Dynamic Psychotherapy: Changes in therapeutic outcome. *Psychotherapy Research*, *10*, 119–131.

Bryk, A., & Raudenbush , S. W. (1992). *Hierarchical linear models for social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage.

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

Connolly, M. B., Crits-Christoph, P., Barber, J. P., Luborsky, L. (2000). Transference patterns in the therapeutic relationship in supportive-expressive psychotherapy for depression. *Psychotherapy Research*, *10*, 356–372.

Constantino, M. J., Glass, C. R., Arnkoff, D. B., Ametrano, R. M., & Smith, J. Z. (2011). Expectations. In J. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 354–376). Oxford: Oxford University Press.

Crits-Christoph, P., Connolly Gibbons, M. B., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research*, *16*, 268–281.

Derogatis, L. R. (1983). *SCL-90: Administration, scoring, and procedures manual for the revised version*. Baltimore: Clinical Psychometric Research.

Dinger, U., Strack, M., Leichsenring, F., Wilmers, F., & Schauenbirg, H. (2008). Therapist effects on outcome and alliance in inpatient psychotherapy. *Journal of Clinical Psychology*, *64*(3), 344–354.

Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006). Therapist effects in the National Institute of Mental Health treatment of depression collaborative research program. *Psychotherapy Research*, *16*, 144–160.

Elliott, R., Bohart, A. G., Watson, J. C., & Greenberg, L. S. (2011). Empathy. In J. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 132–152). Oxford: Oxford University Press.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The global assessment scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, *33*, 776–771.

Farber, B. A., & Doolin, E. M. (2011). Positive regard and affirmation. In J. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 168–186). Oxford: Oxford University Press.

Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing: A comparative study of psychotherapy* (3rd ed.). Baltimore: The Johns Hopkins University Press.

Garfield, S. L. (1997). The therapist as a neglected variable in psychotherapy research. *Clinical Psychology: Science and Practice*, *4*, 40–43.

Gough, H. (1987). *California psychological inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G., & Bradley, P. (2002). *CPI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Greenberg, L. S. (2015). *Emotion-focused therapy: Coaching clients to work through their feelings* (2nd ed.). Washington, DC: American Psychological Association.

Greenberg, L., & Paivio, S. (1997). *Working with emotions in psychotherapy*. New York, NY: The Guilford Press.

Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. *Psychotherapy: Theory, Research, Practice, Training*, *43*, 292–299.

Hill, C. E. (2009). *Helping skills: Facilitating exploration, insight, and action*. Washington, DC: American Psychological Association.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307–316.

Hollon, S. D. (1999). Allegiance effects in treatment research: A commentary. *Clinical Psychology: Science and Practice*, *6*, 107–112.

Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *Inventory of interpersonal problems: Manual*. San Antonio, TX: The Psychological Corporation.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S., (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885–892.

Horvath, A. O. (1981). *An exploratory study of the working alliance: Its measurement and relationship to outcome* (Unpublished doctoral dissertation). Vancouver, Canada.

Horvath, A. O., & Greenberg, L. S. (1986). Development of the working alliance inventory. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 529–556). New York, NY: Guilford.

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. In J. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 25–69). Oxford: Oxford University Press.

Kendall, P. C., & Wilcox, L. E. (1980). Cognitive-behavioral treatment for impulsivity: Concrete vs. conceptual training in nonself-controlled problem children. *Journal of Consulting and Clinical Psychology*, *48*, 80–91.

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). Hoboken, NJ: John Wiley & Sons.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., …Burlingame, G. M. (2004). *Administration and scoring manual for the outcome questionnaire —45*. Orem, UT: American Professional Credentialing Services.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125.

Norcross, J. C. (2011). *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed.). Oxford: Oxford University Press.

Ogles, B. M. (1996). Assessing outcome in practice. *Journal of Mental Health*, *5*(1), 35–46.

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy*, *10*(6), 361–373.

Orlinsky, D. E., & Howard, K. I. (1986). Process and outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed., pp. 311–381). New York, NY: Wiley.

Rice, L., & Kerr, L. N. (1986). Measures of client and therapist vocal quality. In L. S. Greenberg & W. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 73–106). New York, NY: Guilford.

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, *51*(3), 649–660.

Riggio, R. E. (1989). *Manual for the social skills inventory*. Palo Alto, CA: Consulting Psychologists Press.

Safran, J. D., & Muran, J. C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide*. New York, NY: Guilford.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24, 323–355.

Strupp, H. H. (1980). Success and failure in time-limited psychotherapy: A systematic comparison of two cases: Comparison I. *Archives of General Psychiatry*, *37*, 595–603.

Strupp, H. H. (1993). The Vanderbilt psychotherapy studies: Synopsis. *Journal of Consulting and Clinical Psychology*, *61*, 431–433.

Strupp, H. H. (1998). The Vanderbilt I study revisited. *Psychotherapy Research*, *8*, 17–29.

Strupp, H. H., & Binder, J. L. (1984). *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy*. New York, NY: Basic Books.

Strupp, H. H., & Hadley, S. W. (1979). Specific versus nonspecific factors in psychotherapy: A controlled study of outcome. *Archives of General Psychiatry*, *36*, 1125–1136.

Wampold, B. E. (2001). *The great psychotherapy debate*. London: Lawrence Erlbaum Associates.

Wampold, B. E. (2007). Psychotherapy: *The* humanistic and effective treatment. *American Psychologist*, *62*, 857–873.

Wampold, B. E., & Bolt, D. M. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research*, *16*, 184–187.

Wampold, B. E., & Bolt, D. M. (2007). Appropriate estimation of therapist effects: One more time. *Psychotherapy Research*, *17*(2), 256–257.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. New York, NY: Routledge.