Psychotherapy Volume 31/Spring 1994/Number 1

DIFFERENCES IN THE EFFECTIVENESS OF PSYCHODYNAMIC THERAPISTS: A PROCESS–OUTCOME STUDY

LISA M. NAJAVITS
Harvard Medical School

HANS H. STRUPP
Vanderbilt University

Using data from the Vanderbilt II study, 16 therapists were identified as either "more-effective" or "less-effective," and specific behaviors associated with each were identified. Effectiveness of therapists was defined by patients' outcome scores and length of stay in treatment. Measures were completed by patients, therapists, independent observers, and therapists' supervisors. In-session behaviors were assessed at sessions 3, 8, 16, 22. Results indicated that more effective therapists displayed more positive behaviors (e.g., warmth, alliance), fewer negative behaviors (e.g., attacking, blaming), and more self-criticism than less effective therapists. All therapists, however, showed some negative behavior. Significant results were almost entirely relationship-oriented. Length of stay was far more often related to therapist behavior than were outcome scores. Therapists, independent observers, and patients were able to identify therapists as either more effective or less effective, while supervisors were largely unable to make the distinction. Finally, a subgroup of three "most-effective" therapists was found.

. . . for ultimately, and precisely in the deepest and most important matters, we are unspeakably alone; and many things must happen, many things must go right, a whole constellation of events must be fulfilled, for one human being to successfully advise or help another. Rainer Maria Rilke

The difficulty of achieving beneficial results in psychotherapy has long been a concern within the field. While a consensus has emerged that therapy is, on average, effective (Garfield & Bergin, 1986), there remain questions concerning negative effects, early dropout from treatment, and low success rates with resistant patients (Strupp & Binder, 1984). Thus, while therapy appears generally beneficial, there is much room for improvement.

The therapist's role remains an especially problematic question. The importance of the therapist has been emphasized in theory (Frank, 1961; Freud, 1957; Strupp, 1960) and is wisely held to be important in clinical practice (Lambert, Shapiro & Bergin, 1986). However, empirical research has not yet identified therapist factors that contribute substantially to outcome (Beutler, Crago & Arizmendi, 1986). Several assumptions in the psychotherapy literature may have hindered research on therapist factors. One assumption, the "uniformity myth of therapists" (Kiesler, 1966), is the tendency to view therapists as interchangeable in research studies. Luborsky et al. (1986) reviewed over 500 outcome studies, and found that almost all compared treatments (*e.g.*, psychodynamic versus behavioral) but not therapists. A

This article was part of the dissertation of the first author, completed at Vanderbilt University. The author wishes to acknowledge those to whom the dissertation was dedicated: Magda Najavits, Margita Kolan, and Burke Nersesian.

Portions of this article were previously presented at the annual meetings of the American Psychological Association (Boston, MA, August 1990) and the Society for Psychotherapy Research (Lyons, France, July 1991).

Correspondence regarding this article should be addressed to Lisa M. Najavits, McLean Hospital, Proctor III, 115 Mill St., Belmont, MA 02178.

second assumption might be termed the "assumption of averages": the tendency of most psychotherapy research to focus on mean scores across therapists. While this strategy illuminates characteristics of helpful therapeutic activities in general, it fails to describe extreme behavior of very expert or very poor therapists. Finally, a third assumption could be called "the assumption of therapist demographics." Therapists are typically compared on demographic data such as number of years experience or training discipline (psychology, social work, psychiatry). According to meta-analyses and reviews, however, such demographic data have not typically been found related to outcome (Berman & Norton, 1985; Hattie, Sharpley & Rogers, 1984; Lambert, Shapiro & Bergin, 1986).

Empirical Studies of Therapist Effectiveness

A few studies have addressed variability in therapist performance. Ricks (1974) coined the terms "Supershrink" and "Subshrink" in his report of two therapists who treated severely disturbed adolescent boys. Of patients who later developed schizophrenic symptoms, 27% of the caseload of the "Supershrink" fit this category, compared to 88% of the cases of "Subshrink." This study was limited, however, in its assessment of patients 20 years post-treatment and use of case notes. Orlinsky & Howard (1980) studied a sample of 23 "traditional verbal" therapists, using case notes. They found six therapists with beneficial results in at least 70% of their patients, and no negative effects. In contrast, five therapists showed beneficial results in fewer than 50% of their cases, with a 10% rate of negative effects. Miller, Taylor & West (1980) discovered success rates ranging 25–100% among nine therapists treating 31 substance abusers in three behavioral treatments. However, this study did not evaluate effectiveness independent of type of treatment. Other early studies are described in Lambert (1989) though most, like these, have major methodological shortcomings.

Lambert, Shapiro & Bergin (1986) have also summarized meta-analyses on the topic of negative effects, that is, patients who worsen in treatment. Approximately 9%–11% of patients fit this category. Perhaps the most comprehensive work on therapists comes from Luborsky, Crits-Christoph, and colleagues (Crits-Christoph et al., 1991; Crits-Christoph & Mintz, 1991; Luborsky et al., 1986). They reanalyzed data from four major outcome studies of psychodynamically based treatments. In all four studies, the therapist factor was significant, and stronger than that for treatment type. "Matching" of patient characteristics and therapists was not significant. They also discovered that therapist performance varied widely within caseloads (Luborsky et al., 1986). In a later analysis of 15 outcome studies, the therapist factor was again significant, and represented about 9% of outcome. This was termed a "medium" size effect (Crits-Christoph et al., 1991). Finally, the use of treatment manuals and more experienced therapists were associated with greater consistency in therapists' results.

In addition to outcome data, another way to assess therapist effectiveness is length of stay in treatment. Garfield & Bergin (1986) found that many patients leave treatment before therapists would expect. Approximately 50% depart before the eighth session. Sledge et al. (1990) found a 32%–67% rate of attrition for time-limited psychodynamic therapy in a community mental health setting.

What characteristics are associated with therapists who differ in effectiveness? Lafferty, Beutler & Crago (1989) studied 30 trainees treating a total of 60 outpatients, with "effectiveness" defined as patient change on the Global Severity Index (a measure of overall level of pathology). Less effective therapists showed less empathy, and valued comfort, stimulation, and intellectual goals less than did more effective therapists. They also rated their patients as more involved in treatment and themselves as more supportive than did the more effective therapists. In the study by Ricks (1974), the "Supershrink" surpassed the "Subshrink" in effort, greater support of patients' autonomy, use of resources outside of therapy, and better relationships with patients' parents. The "Subshrink" became overwhelmed and depressed when treating difficult cases. Luborsky et al. (1985) compared patient descriptors (demographics, diagnosis, illness history), therapist qualities (emotional adjustment, interest in helping patients, and therapeutic skill), therapeutic alliance, and degree to which therapists adhered to a treatment model. They found that the therapist qualities were related to the ability to establish alliance, which in turn was associated with patient improvement.

In short, these studies provide evidence for therapists' differences in performance, and suggest an initial description of more and less effec-

tive therapists. However, they are quite limited in methodology (*e.g.*, the use of case notes, self-report data, trainees, small samples, and single measures of outcome). As Luborsky et al. (1986) have remarked, the dearth of research on therapist differences is "amazing" (p. 502).

In the current study, data from the Vanderbilt II Study (Strupp, 1993) will be used to differentiate 16 experienced therapists as "more effective" and "less effective." Two indices of effectiveness will be used. First, patient outcome scores will be used since degree of patient improvement is the standard in the field for testing effects of therapy. Second, patient dropout rates will be used, since previous reports have documented the importance of dropouts both in general samples (Garfield & Bergin, 1986); and specifically in time-limited psychodynamic therapy (Sledge et al., 1990), the treatment used by all therapists in the current study. Once therapists are ranked on effectiveness, the more effective and less effective therapists will be compared to identify qualities which distinguish them, such as warmth, alliance, blaming, attack, affirmation, ignoring, and helpfulness toward patients. Several sub-questions will also be addressed: How consistent are therapists in effectiveness? Who can distinguish more effective from less effective therapists (a comparison of patients, therapists, and independent observers, and supervisors)? How does training of therapists affect their performance? What qualities are associated with the therapist–patient alliance? Do all therapists show negative behavior? Which are more crucial for distinguishing therapists: relationship factors or specific therapy techniques?

It should be noted at the outset that this study suffers several inherent flaws: the lack of a control group, heterogeneous diagnoses of the patient population, and the retrospective design of the study. However, we believe it extends beyond previous work in several important ways. Multiple measures of therapist effectiveness are used (length of stay, in addition to outcome), and the consistency of effectiveness is addressed. Also, the data set is more extensive than any previously used to address the topic of therapist effectiveness. A wide variety of measures are used (18 in all), as well as multiple raters (patient, therapist, independent observer, supervisor), and sessions (3, 8, 16, and 22). Also, the therapist sample is experienced, and alike in orientation, permitting a comparison of therapists who are ostensibly similar.

Method

The Vanderbilt II Study provided the data for this project. The Vanderbilt II Study was designed to examine the effect of therapist training on psychotherapy outcome, and is reported in Strupp (1993), Henry, Strupp, Butler, Schacht and Binder (1993), and Henry, Schacht, Strupp, Butler and Binder (1993).

Therapist subjects. The 16 therapists included 8 psychologists and 8 psychiatrists, all with at least two years experience post-degree, an average experience level of 5.6 years, and previous training and supervision in psychodynamic principles. All were recommended by previous supervisors as caring, empathic clinicians. There were 10 males and 6 females, all Caucasian.

Patient subjects. The 80 patients, recruited via newspaper announcements, ranged in age from 24–64. Subjects were selected to be comparable in severity to ordinary outpatient samples (*i.e.*, SCL-90R Global Severity Index within one standard deviation of outpatient normative mean; Derogatis, 1983). Subjects were excluded if they had substance use disorders, severe medical problems, or needed psychiatric medication or inpatient treatment. Based on the computerized version of the NIMH Diagnostic Interview Schedule (administered by a trained clinical interviewer), 87% received an Axis I diagnosis and 67% an Axis II diagnosis.

Study design. Each therapist treated 5 outpatients in 25 sessions, once weekly, as part of the Vanderbilt II Study: 2 patients pre-training (cohort 1), 1 during training (cohort 2), and 2 post training (cohort 3). The training focused on Time Limited Dynamic Psychotherapy (TLDP; Strupp & Binder, 1984), led by senior members of the Center for Psychotherapy Research during 12 months of weekly 2-hour sessions.

Measures. Three sets of measures were used, corresponding to three main data analyses. The first, "pretreatment characteristics," were a set of measures to check whether therapist caseloads were similar before treatment began. These were: the Capacity for Dynamic Process Scale (CDPS; Thackrey, Butler & Strupp, 1985), a measure of patients' suitability for TLDP; patient demographic variables (education, age, number of marriages, number of previous jobs, job length); and severity of patient pathology. The latter was a summary score across four measures of patient symptomatology: the SCL-90R; Structural Analy-

sis of Social Behavior (SASB; Benjamin, 1983)[1] Global Assessment Scale, patient's lowest level of functioning in the past week (GAS; Endicott, Spitzer & Fleiss, 1976); and the Problem Severity Scale (PSS), a rating from 0 to 100 of the patient's most difficult problem. All were rated by the patient, except the GAS, which was rated by the therapist and independent observer.

Second, "Therapist Effectiveness Measures" were used to identify therapists as more effective or less effective. Two scores were developed for this purpose: "outcome" and "length of stay in treatment." "Outcome" refers to the improvement in patients' level of pathology by the end of treatment, based on six measures: the SCL-90 R; SASB; PSS; GAS; Post-Therapy Evaluation, 17 items on a 1 to 5 scale of the therapist's rating of outcome (PTE; Strupp, Fox & Lessler, 1969); and Global Outcome Rating, a one-item rating from -5 to $+5$ indicating amount of change since therapy began (GOR; rated by patient, therapist, and independent observer). The other effectiveness variable was "Length of stay in treatment," that is, the number of patients who remained with the same therapist for at least 16 of the 25 available sessions. While somewhat arbitrary, this criterion was chosen to indicate whether the patient had remained in treatment for most of the study (at least four-fifths of the sessions), and had not switched therapists. Also, for patients who departed prior to session 16, their stated reason for leaving was classified as either "due to dissatisfaction with the therapist" or "due to reasons other than dissatisfaction with the therapist." Examples of the latter would include moving away from the area or time constraints.

Third, measures of "Therapists' In-Session Behavior" addressed various aspects of therapists' functioning during therapy. Scales rated by independent observer were: the Luborsky Helping Alliance Scale, session 3 (HA; Luborsky et al., 1983); Vanderbilt Psychotherapy Process Scale, session 3 and 16 (VPPS: Suh, Strupp & O'Malley, 1986); Vanderbilt Negative Indicators Scale, session 3 (VNIS; Suh et al., 1986); and Vanderbilt Therapeutic Strategies Scale, sessions 3 and 16 (VTSS; Butler, Lane & Strupp, 1986). All of the ratings by independent observers were based on a video tape of the second fifteen minutes of the session. Scales rated by the patient were: the Barrett-Lennard Relationship Inventory, sessions 3, 8, 16 (BARLEN; Barrett-Lennard, 1962); Retrospective Assessment of Therapy Experience, after treatment, a 1 to 5 rating scale of how well treatment went (RATE; Strupp, Fox & Lessler, 1969); and SASB, sessions 3, 8, 16, 22.[2] Measures rated by the therapist were: SASB, sessions 3, 8, 16, 22; the Post-Session Rating, a single item of how well the session went, on a ten-point scale rated after sessions 3 and 16 (PSR); and Therapist Regrets, a yes/no rating of whether the therapist reported having made mistakes during session 3 (TR). Finally, supervisors rated the therapist after training on six items (*e.g.*, competence, motivation) (SR). The independent observers were advanced clinical psychology graduate students or practicing clinicians, trained to use the measures, blind to the hypotheses of the current study, and formed into teams of at least two raters per measure (from which their average score was taken to improve reliability). Personnel on each team was held consistent throughout the rating process, and at least two teams were available for all measures, except the VNIS and the HA, for which each had only one team. Assignment of teams to videotape segments was based on a goal of equal numbers of segments for each team, as well as availability of the team for the rating. Where data were available from multiple sessions, an average was also taken.

Data analysis. Three main questions were addressed in the analyses. In the first analysis, "pretreatment characteristics," therapists were compared for similarity of their caseloads before

[1] SASB data were average cluster scores of patients' ratings of their introject (across "introject at best" and "introject at worst"). Clusters 1 and 5 were omitted because they had no clear positive or negative direction. SASB is believed appropriate as an outcome measure as it typically highly correlated with other outcome measures (Strupp, 1988).

[2] SASB process data were averaged to obtain a global score for each therapist per cluster (across sessions, cohorts, and relationships), with ratings by the patient and by the therapist kept separate. Specifically, an average was obtained per therapist caseload across sessions (3, 8, 16 and 22), cohorts (1, 2 and 3), and relationships 31, 32, 35, 36, 39 and 40. These relationships represent the patients' ratings of the therapist "at best" (relationships 31 and 32), "at worst" (35 and 36), and "as is" (39 and 40) on the two SASB surfaces #1 ("other"; 31, 35, 39) and #2 ("self"; 32, 36, 40). The same procedure was also used to calculate therapists' ratings of the their own process behavior. The final result were the variables Cluster 1–8 (patient ratings), and Cluster 1–8 (therapist ratings).

treatment began, using one-way ANOVA on patient demographic variables, patient "capacity for dynamic process" (CDPS) score, and patient pretreatment pathology composite (across all of the "Therapist Effectiveness" variables, standardized). One-way ANOVA was chosen over MANOVA due to missing data on seven patients. The one-way ANOVA design was a between-subjects analysis, with therapist as independent variable (each therapist as a level). In the second analysis, "Therapist Effectiveness," therapists were ranked on the two effectiveness measures: outcome and length of stay. Outcome refers to the average amount of positive change per therapist caseload (based on residualized change scores pooled across the patient sample, pre- to post-treatment, on a composite of all the "Therapist Effectiveness" outcome measures). Length of stay in treatment refers to the number of patients who left treatment before session 16 or who switched therapists.[3] Also, to test therapists' consistency, Cronbach's alpha was run, first on outcome and then length of stay (using a therapist by patient matrix, pooling across therapists and patients).

Finally, the last set of analyses, "Therapist In-Session Behavior," explored the link between therapists' effectiveness and their in-session behaviors. The effectiveness measures (outcome and length of stay) were studied separately on three questions: (a) What in-session behaviors were related to effectiveness at the beginning of treatment (session 3), and across treatment (sessions 3, 8, 16, and 22)? Discriminant analysis and *F* tests were used. (b) Who could distinguish therapists' effectiveness (a comparison of therapist, patient, independent observer, and supervisor)? For each point of view, a total score was created across all of the in-session behavior data. The points of view were then compared with each other via correlation, and to the effectiveness variables via MANOVA. (c) What in-session behaviors are associated with the therapeutic alliance? Therapists were divided into "more alliance" or "less alliance" groups, based on the amount of alliance within caseloads. A comparison of the two groups was explored with discriminant analysis and *F* tests. All discriminant analyses used simultaneous entrance of variables, with tolerance set at .001.

Results and Discussion

I. Equivalence of Therapist Caseloads Prior to Treatment

Therapists' caseloads were equivalent before treatment began. ANOVAs compared therapists' caseloads at pre-treatment on severity of patient pathology, demographic variables, and "capacity for dynamic process." Virtually no significant differences were found.[4] This indicates that differences later found between therapists could not be attributed to initially imbalanced caseloads.

II. Therapists' Differences in Effectiveness

A subgroup of "most effective" therapists was identified. Three therapists could be identified as "most effective." That is, they ranked at the top in outcome, had no negative outcome cases, and had no patients leave before session 16. The binomial probability that three therapists would score high across the effectiveness measures was $p = .07$.,[5] a trend. On the lower end, three therapists were in the lower half on outcome, had at least one negative outcome case, and had two or more patients leave treatment before session 16 (with at least one departure due to "dissatisfaction with the therapist"). The binomial probability was non-significant, however. These findings suggest that even in an experienced therapist sample, differences in performances may occur. Our results

[3] "Length of stay" includes one patient who left treatment after session 16 since the patient reported ending due to dissatisfaction with the therapist. Three patients who moved out of state were not included. Patients with fewer than five sessions were not included as no outcome data were available (they were replaced with new patients).

[4] The only significant findings were as follows. Two therapists, #18 and #22, had significantly less severe patient pathology in their caseloads ($F(15, 64) = 1.92, p < .04$) using Duncan's multiple range test. Thus, therapist #22 was dropped from all further analyses; therapist #18 was retained since s/he emerged very low on effectiveness despite the easier caseload. Note too that therapist #16 was dropped from further analyses of outcome because of an unusually large inconsistency (standard deviation) on outcome. Finally, therapists 16 and 28 had significantly higher caseload means on the patient demographic variable "average length of job." This did not affect further analyses since both therapists still emerged low on effectiveness.

[5] Only outcome and length of stay were used in the analysis, since negative outcome is not independent of the outcome measure.

probably underestimate differences which would be likely "in the real clinical world," given that the sample was small and likely restricted in range (in that all therapists were highly recommended, had at least two years post-degree experience, and agreed to have their work observed).

Therapists were inconsistent on "outcome," but moderately consistent on "length of stay". On outcome across all therapists, Cronbach's alpha (a measure of consistency) was low (−.58). For further analyses on outcome the top four were designated as the more effective group and the bottom four therapists the less effective group, as the goal of the project was to compare relatively extreme therapists. For further analysis on length of stay, therapists were also split into two groups. The more effective group had no patients leave treatment before the 16th session, $n = 5$. The less effective group had at least 2 of their 5 patients leave prior to session 16, at least one of whom left due to "dissatisfaction with the therapist," $n = 7$. Cronbach's alpha on length of stay yielded .36, indicating some consistency.

In order to better understand the length of stay variable, patients who left prior to session 16 were compared on overall pathology with those who remained in treatment. Before treatment began no difference between the two groups was found (using a *t*-test). At the point of leaving treatment, however, the group of patients who left treatment before session 16 showed significantly more pathology (−.36, $n = 20$) than the group who remained (.18, $n = 61$), $t(78) = -3.18$, $p < .001$. This might be interpreted as a therapist effect. That is, patients left prior to session 16 because they were not getting better, or perhaps because what might be normal regression during treatment was not able to be managed adequately. Alternatively, we cannot rule out whether life stresses outside of therapy might account for these results.

16% of patients ($n = 13$) showed negative outcomes. "Negative outcome" refers to patients who become worse during treatment. In this study, patients were classified as having had negative outcomes if their end-of-treatment pathology score was higher than the pre-treatment score. Three therapists (#12, #18, and #28) showed a rather high rate of such negative outcome patients—each had two from their caseload of five. Also, it is interesting that 9 of the 13 negative outcome cases left treatment before session 16. This finding might indicate that patients who were feeling more symptomatic did not believe therapy could help them.

Training affected therapist performance. Since the Vanderbilt II Study focused on therapist training, we also sought to explore the impact of training on effectiveness. Therapist rankings on outcome in cohort 1 (the two pre-training cases) were compared to cohort 3 (two post-training cases) by the Wilcoxon test for matched samples, with a significant $z = 2.56$, $p < .05$. Most therapists (13 of 16) showed *decreased* performance after training, a rather unexpected finding. If one split the therapists into "top" and "bottom" halves, however, the majority (10 of 16) remained in the same half before and after training. This would suggest that while training decreased therapist effectiveness, it did not generally affect therapist groupings for the purposes of the current study. On length of stay, 7 patients left treatment before session 16 in cohort 1 compared to 11 in cohort 3, also suggesting slightly lower therapist performance after training. Reasons for the both the variability in therapists' response to training, and the observed decrements in therapist performance are discussed at length in Henry, Schacht, Strupp, Butler and Butler (1993).

III. In-session Behavior of Therapists

More effective therapists showed more positive behaviors and fewer negative behaviors than less effective therapists. Examples of positive variables were warmth, affirmation and understanding, and helping and protecting. Examples of negative variables were belittling and blaming, ignoring and neglecting, attacking and rejecting. Moreover, the consistency of these results was quite remarkable: 18 of the 19 significant findings in Table 1 and Table 2 followed this pattern.

More effective therapists were more self-critical than were less effective therapists. This finding can be observed by examining the variable "therapist regrets" in Tables 1 and 2. "Therapist regrets" was therapists' rating of whether they made mistakes during session 3. Thus, the more effective therapists actually rated their sessions as having gone less well than did the less effective therapists (*cf.* Flasher et al., 1987).

The helping alliance was shown to be associated with several specific in-session behaviors. Table 3 shows that therapists with higher scores on helping alliance (HA) displayed more warmth and friendliness, and more affirmation and understanding than therapists lower on alli-

TABLE 1. Behaviors Distinguishing More- versus Less-Effective Therapists (across sessions 3, 8, 16, 22)

I. Outcome: No significant differences.*
II. Length of stay:** 8 of 34 variables significant.***

Variable	Mean per group: More Effective	Mean per group: Less Effective	*F*	*p*
Warmth (HA)	9.00	6.25	6.06	.03
Warmth/friendliness (VPPS)	.45	−.15	6.00	.03
Patient view (RATE)	1.37	1.12	7.80	.02
Therapist regrets (TR)	.80	.14	19.20	.001
Affirm/understand (SASB, Pt)†	74.58	65.58	6.80	.03
Affirm/understand (SASB, Th)	74.60	64.90	7.70	.02
Watch/manage (SASB, Th)	10.90	15.80	5.20	.05
Belittle (SASB, Th)	2.30	7.40	5.50	.05

* Discriminant analysis not significant. None of 34 process variables significant on *F* tests ($df = 1, 6$). For all outcome analyses (Tables 1–3), $n = 8$.

** For this and all additional tables, length of stay was scaled positively to match all other variables in the study; that is, higher equals more positive performance. For all length of stay analyses (Tables 1–3), $n = 12$.

*** Discriminant analysis significant, χ^2 (10, $N = 12$) 18.9, $p < .04$. Eight of 34 variables were significant (all $df = 1, 10$). Note that in this and all subsequent analyses, the number of significant variables was always higher than the 1 in 20 expected by chance, indicating Type I error insufficient to account for the results.

† See Benjamin (1986) for a description of SASB. All variable names are from surface 2 (focus on other) to simplify presentation of results; however, SASB data used are in fact an average across surfaces 1 (focus on self) and 2 (focus on other). "Pt" refers to patient ratings, "Th" to therapist ratings.

ance. They also showed less belittling and blaming, and less attacking and rejecting. All means were in the expected direction, with higher HA therapists showing more positive behaviors and fewer negative behaviors. It is noteworthy too that of the negative behaviors, active hostility toward the patient (attacking and rejecting, belittling and blaming) was significant in differentiating the therapist groups, while passive hostility (ignoring and neglecting) was not significant. It is also noteworthy that both patient and therapist data were in agreement in that the same SASB clusters were significant for both. The case of therapist #18 remains an unresolved anomaly, however: this therapist obtained the highest alliance score of the entire sample, yet fit the category of "least effective" therapists. Such a finding might suggest that, on occasion, less effective therapists may compensate by developing a style which is superficially therapeutic, yet which may be unrelated to the actual outcome of their patients (highlighting the need for multiple evaluations of therapist performance).

Even the best therapists were not entirely free of negative behaviors. Examination of the means in Tables 1–3 shows that all therapists showed some negative behaviors. For example, in Table 2, under "ignoring and neglecting," the more effective group had a mean of 16.91. This might suggest that even the most effective thera-

TABLE 2. Behaviors Distinguishing More- versus Less-Effective Therapists (at session 3 only)

I. Outcome: 4 of 29 variables significant:

Variable	Mean per group: More Effective	Mean per group: Less Effective	*t*	*p*	*df*
Negative attitude (VNIS)	.35	−.51	2.12	.05	16
Affirm/understand (SASB, Th)	66.74	45.05	4.99	.000	26
Help/protect (SASB, Th)	48.08	38.30	2.98	.005	33
Ignore/neglect (SASB, Th)	16.91	28.03	−2.74	.01	33

II. Length of stay: 8 of 29 variables significant:

Relationship (BARLEN)	1.85	1.55	1.99	.05	58
Warmth/friendliness (VPPS)	.44	−.20	2.33	.02	45
Therapist regrets (TR)	.80	.14	4.09	.001	22
Affirm/understand (SASB, Pt)	70.98	59.47	−3.68	.001	58
Belittle/blame (SASB, Th)	2.76	7.33	−2.05	.02	58
Attack/reject (SASB, Pt)	1.02	7.01	2.71	.003	58
Ignore/reject (SASB, Pt)	3.69	9.19	−2.01	.03	58
Attack/reject (SASB, Th)	15.72	10.92	2.40	.02	54

TABLE 3. Behaviors Distinguishing Therapists with More versus Less Helping Alliance (6 of 34 variables significant)*

Variable	Mean per Group: More Helping Alliance	Mean per Group: Less Helping Alliance	*F*	*p*
Warmth/friendliness (VPPS)	.52	−.31	4.34	.01
Affirm/understand (SASB, Pt)	73.23	63.73	4.91	.05
Belittle/blame (SASB, Pt)	2.50	7.09	5.61	.04
Attack/reject (SASB, Pt)	1.05	6.32	6.50	.03
Belittle/blame (SASB, Th)	2.74	7.09	5.03	.05
Attack/reject (SASB, Th)	1.23	6.32	6.00	.03

* Discriminant analysis not significant; $df = 1, 13$.

pists at times may be susceptible to negative countertransference, or to reenactment of patients' negative interpersonal patterns (*see* Strupp & Binder, 1984, for a discussion of the latter).

Therapists were differentiated almost entirely by relationship ("nonspecific") variables rather than technical ("specific") ones. This finding appeared on all of the main analyses of in-session behavior (Tables 1–3). Thus, basic capacities of human relating—warmth, affirmation, and a minimum of attack and blame—may be at the center of effective psychotherapeutic intervention. Theoretically based technical interventions were not nearly as often significant in this study. Of course, it must be noted too that there were fewer technical than relationship variables in the study.

Therapists, independent observers, and patients were able to distinguish more effective from less effective therapists, whereas supervisors were largely unable to do so. For each point of view, a total score on the in-session behavior data was created, and then analyzed in relation to outcome and length of stay (using MANOVA). On outcome, no significant differences were found ($df = 1,6$). On length of stay, the therapist and independent observer points of view were significant, and the patient point of view showed a trend on individual *F* tests. The means for more effective and less effective therapists were: .58 versus −.28 (therapist point of view, $F = 5.12$, $p < .05$); .35 versus −.11 (independent observer, $F = 7.67$, $p = .02$); .64 versus −.22 (patient, $F = 4.05$, $p = .07$). The supervisor point of view was not significant; and the overall MANOVA was not significant, $df = 1,11$.

Correlations helped to further pursue this line of inquiry. Results were similar to those found above. Therapist, patient, and observer process ratings were highly intercorrelated but supervisor process ratings correlated only with that of observers. The significant correlations were: patient–therapist ($r = .94$, $p < .000$), patient–observer ($r = .58$, $p < .02$), therapist–observer ($r = .60$, $p < .02$), supervisor–observer ($r = .73$, $p < .001$). It might be noted that supervisory ratings were completed after training, which may have decreased their validity.

The very high correlation between therapists' and patients' points of view, and therapists' willingness to rate their own behaviors as negative were quite remarkable (*cf.* the review by Newman, 1983, who also found evidence for therapists' ability to be self-critical of their work).

Outcome and length of stay appeared unrelated. Length of stay was far more often related to therapists' in-session behavior. It was surprising to find that the correlation between outcome and length of stay was nonsignificant. Moreover, in Tables 1 and 2, length of stay was significantly related to in-session behavior in 16 of the 63 separate *F*-tests (25%) while outcome was significantly related in only 4 of 63 *F*-tests (6%). These results might suggest that length of stay could be a valuable addition to psychotherapy outcome studies. Clinically, therapists who have difficulty retaining their patients in treatment might be targeted for additional supervision and on-going monitoring.

Future Research

Many questions remain to be addressed in future work. How stable is effectiveness across therapists' careers? What is the range of effectiveness in the clinical community? Might subtypes of more effective and less effective therapists be identified? How do patient characteristics, such as diagnosis, affect therapist performance? What results might be found for mid-range therapists (in contrast to the extreme ends)? To what extent are our results generalizable to therapists practicing other theoretical orientations? Are certain treatment orientations more difficult to master than others? In addition, it might be illuminating to explore several questions that were raised in

the current study: Why were supervisors largely unable to differentiate more effective from less effective therapists? Are length of stay and outcome more strongly related than they appeared in this study? How does training affect therapist skill? Further studies of therapist effectiveness might offer a vista to help make psychotherapy more helpful, to more patients, more often.

References

Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs*, **76**, No. 43 (Whole No. 562).

Benjamin, L. S. (1983). *The INTREX user's manual, Part I and II*. Madison, WI: Intrex Interpersonal Institute, Inc., P.O. Box 5218, Madison, WI 53705.

Berman, J. S. & Norton, N. C. (1985). Does professional training make a therapist more effective? *Psychological Bulletin*, **98**, 401-407.

Beutler, L. E., Crago, M. & Arizmendi, T. G. (1986). Therapist variables in psychotherapy process and outcome. *In* S. L. Garfield and A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed, pp. 257-309). New York: John Wiley.

Butler, S. F., Lane, R. W. & Strupp, H. H. (1988). *Patterns of therapeutic skill acquisition as a result of training in Time-Limited Dynamic Psychotherapy*. Paper presented at the annual meeting of the Society for Psychotherapy Research, Santa Fe, New Mexico, June.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., Luborsky, L., McLellan, A. T., Woody, G. E., Thompson, L., Gallagher, D. & Zitrin, C. (1991). Meta-analysis of therapist effects in outcome studies. *Psychotherapy Research*, **1**, 81-91.

Crits-Christoph, P. & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, **59**, 20-26.

Derogatis, L. R. (1983). *SCL-90R administration, scoring and procedures manual*. Baltimore: Johns Hopkins University Press.

Endicott, J., Spitzer, R. L. & Fleiss, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, **33**, 766-771.

Flasher, L. V. (1987). *Negative factors in short-term psychotherapy: Focus on therapist interventions*. Doctoral dissertation, Vanderbilt University.

Frank, J. D. (1961). *Persuasion and healing*. Baltimore: Johns Hopkins University Press.

Freud, S. (1923/1957). *In* J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth.

Garfield, S. L. & Bergin, A. E. (eds.). (1986). *Handbook of psychotherapy and behavior change*. New York: John Wiley.

Hattie, J. A., Sharpley, C. F. & Rogers, H. J. (1984). Comparative effectiveness of professional and paraprofessional helpers. *Psychological Bulletin*, **95**, 534-541.

Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F. & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Mediators of therapists' responses to training. *Journal of Consulting and Clinical Psychology*, **61**, 441-447.

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E. & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. *Journal of Consulting and Clinical Psychology*, **61**, 434-440.

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, **65**, 110-136.

Lafferty, P., Beutler, L. E. & Crago, M. (1989). Differences between more and less effective psychotherapists: A study of select therapist variables. *Journal of Consulting and Clinical Psychology*, **57**, 76-80.

Lambert, M. J. (1989). The individual therapist's contribution to psychotherapy process and outcome. *Clinical Psychology Review*, **9**, 469-485.

Lambert, M. J., Shapiro, D. A. & Bergin, A. E. (1986). The effectiveness of psychotherapy. *In* S. L. Garfield and A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change*. (3rd ed, pp. 157-211). New York: John Wiley.

Luborsky, L. L., Crits-Christoph, P., Alexander, L., Margolis, M. & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy: A counting signs vs. a global rating method. *Journal of Nervous and Mental Disease*, **171**, 480-486.

Luborsky, L. L., Crits-Christoph, P., McLellan, A. T., Woody, G., Piper, W., Liberman, B., Imber, S. & Pilkonis, P. (1986). Do therapists vary much in their success? Findings from four outcome studies. *American Journal of Orthopsychiatry*, **56**, 501-512.

Luborsky, L. L., McLellan, A. T., Woody, G. E., O'Brien, C. P. & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, **42**, 602-611.

Miller, W. R., Taylor, C. A. & West, J. C. (1980). Focused versus broad-spectrum behavior therapy for problem drinkers. *Journal of Consulting and Clinical Psychology*, **48**, 590-601.

Newman, F. L. (1983). Therapist's evaluation of psychotherapy. *In* M. J. Lambert, E. R. Christenson and S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 498-536). New York: John Wiley.

Orlinsky, D. E. & Howard, K. I. (1980). Gender and psychotherapeutic outcome. *In* A. M. Brodsky and R. T. Hare-Mustin (Eds.), *Women and psychotherapy* (pp. 3-34). New York: Guilford.

Ricks, D. F. (1974). Supershrink: Methods of a therapist judged successful on the basis of adult outcomes of adolescent patients. *In* D. F. Ricks, A. Thomas and M. Roff (Eds.), *Life history research in psychopathology* (vol. 3, pp. 275-297). Minneapolis: University of Minnesota Press.

Rilke, R. M. (1905/1984). *In* S. Mitchell (Trans.), *Letters to a young poet*. New York: Random House.

Sledge, W. H., Moras, K., Hartley, D. & Levine, M. (1990). Effect of time-limited psychotherapy on patient drop-out rates. *American Journal of Psychiatry*, **147**, 1341-1347.

Strupp, H. H. (1960). Nature of psychotherapist's contribution to treatment process: Some research results and speculations. *Archives of General Psychiatry* **3**, 219-231.

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

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

STRUPP, H. H., FOX, R. E. & LESSLER, K. (1969). ***Patients view their psychotherapy.*** Baltimore: Johns Hopkins University Press.

THACKREY, M., BUTLER, S. F. & STRUPP, H. H. (1985). Measurement of patient capacity for dynamic process. Paper presented at the annual meeting of the Society for Psychotherapy Research, Evanston, Illinois, June.