

## **THE EFFECTS OF TRAINING IN TIME-LIMITED DYNAMIC PSYCHOTHERAPY: CHANGES IN THERAPEUTIC OUTCOME**

Edward Bein

Vanderbilt University

Timothy Anderson

Ohio University

Hans H. Strupp

Vanderbilt University

William P. Henry

University of Utah

Thomas E. Schacht

James H. Quillen College of Medicine

Jeffrey L. Binder

Georgia School of Professional Psychology

Stephen F. Butler

Innovative Training Systems, Inc.

The present study explored the effects on therapeutic outcomes of training therapists in brief manualized therapy. As part of the Vanderbilt II project, each of 16 therapists (8 psychiatrists and 8 clinical psychologists) treated 2 moderately disturbed adult patients using his or her customary short-term treatment methods; they then received a year of training in a manualized form of brief dynamic therapy, Time-Limited Dynamic Psychotherapy (TLDP); finally, they administered TLDP to 2 additional patients. It was hypothesized that training would result in improved outcomes generally and that differentially greater improvement would be seen in patients commonly considered less suitable for brief dynamic therapy. Outcome data obtained at termination failed to support either hypothesis. Measurements of interpersonal dependency obtained at a one-year follow-up were consistent with the first hypothesis, but the follow-up data were inconsistent with the second. A systematic review of the 32 posttraining cases suggested that the majority of the therapists had not achieved basic competence at TLDP.

During the past 20 years, psychotherapy outcome research has paid particular attention to manualized psychotherapies, studying their relative efficacy with respect to

---

This work was supported in part by National Institute of Mental Health Grants MH16247 and MH20369. We thank Douglas Vakoch, Nana Landenberger, Jessica Londa-Jacobs, John Dugan, and Kristin Nelson for their assistance with this research.

Correspondence concerning this article should be addressed to Edward Bein, Center for AIDS Prevention Studies, UCSF, 74 New Montgomery Street, Suite 600, San Francisco, California 94105. E-mail: ebein@psg.ucsf.edu.

one another and to psychopharmacotherapy. These forms of psychotherapy are defined by a treatment manual that attempts to specify, in an explicit and detailed manner, the steps practitioners should take in order to conceptualize and treat their patients' psychological disorders. The intent is that the use of treatment manuals in guiding training and practice within research projects will promote treatment integrity and thus allow more precise and valid conclusions to be drawn about the usefulness of the treatments studied (Goldfried, Greenberg, & Marmar, 1990).

Would psychotherapists in clinical settings be more effective if they were trained in and faithfully practiced manualized treatments instead of their customary treatment approaches? This question has particular significance for the contemporary therapeutic community: As the demand for accountability rises and third-party payments shrink, more therapists are likely to seek out continuing education in manualized short-term treatment approaches. Wilson (1996) and Addis (1997) have outlined the central issues involved in assessing the worth of manualized treatments versus treatment as usual in clinical settings, but to date there are only a few empirical studies directly addressing the question (e.g., Schulte, Kunzel, Pepping, & Schulte-Bahrenberg, 1992).

The Vanderbilt II project (Strupp, 1993) was designed to address this question with regard to one form of brief manualized therapy, Time-Limited Dynamic Psychotherapy (TLDP; Strupp & Binder, 1984). It specifically examined whether training therapists in this modality improves their ability to skillfully manage the vicissitudes of the therapeutic relationship and increases their therapeutic effectiveness, by comparing the performance of the same set of therapists before and after a year of training in a manualized TLDP protocol.

TLDP was created in the wake of the Vanderbilt I study (Strupp & Hadley, 1979; Strupp, 1980a, 1980b, 1980c, 1980d; Gomes-Schwartz, 1978) to be an effective treatment for a wider range of patients than prior brief psychodynamic therapies (e.g., Sifneos, 1979; Malan, 1976) had been. A loosening of such treatments' selection criteria was deemed possible because "in TLDP, a 'good' therapeutic relationship is not a prerequisite for treatment; rather, the *obstacles* to developing such a relationship, as they emerge in the patient-therapist interaction, are considered the primary area of work" (Strupp & Binder, 1984, p. 58). Patients with an unrealized potential to work collaboratively in treatment would be helped to develop this capacity through the active facilitation of a positive therapeutic alliance and through the early and consistent examination, in the here-and-now, of hostile, suspicious, and resistant behaviors toward the therapist.

Previous analyses of Vanderbilt II data have focused on the effects of the TLDP training on therapist in-session behaviors. Henry, Strupp, Butler, Schacht, and Binder (1993) reported that following training, therapists were more active and exhibited greater adherence to TLDP-consistent interventions, but they also exhibited as much or more negativistic behavior toward patients as before training. Thus, the training did not have the hypothesized effect of reducing negative countertransference reactions. Henry, Schacht, Strupp, Butler, and Binder (1993) investigated trainer teaching style, therapist experience level and introject, and patient suitability as mediators of therapist response to training. They found that therapists with more punitive introjects tended to show greater technical adherence but poorer interpersonal process. They also found that the therapists supervised by one TLDP trainer exhibited significantly greater posttraining adherence to TLDP than did the supervisees of the other trainer. The former trainer was judged to be more directive and task-specific than the latter.

The present study focuses on whether the TLDP training increased therapeutic effectiveness. Because TLDP was specifically developed to help therapists become more

effective with patients viewed by prior dynamic approaches as less suitable for brief therapy, we were particularly interested in the effects of training on therapist effectiveness with these patients. Thus, we tested two hypotheses in this study: (1) TLDP training improves therapist effectiveness; and (2) TLDP training is differentially more beneficial in improving therapist effectiveness with supposedly less suitable patients.

## METHOD

### PARTICIPANTS

*Psychotherapists.* Sixteen licensed psychotherapists, half psychologists and half psychiatrists, participated in the Vanderbilt II project. They were recommended by senior clinicians practicing in the community and faculty members in the Psychiatry Department who had been asked by the project director to nominate worthy participants. Ten were men, six were women, and all were Caucasian. Therapists were on average 4.3 years postinternship or postresidency ( $SD = 3.4$ , range 0–11) and were in private practice. Eleven of the therapists considered themselves at least partly psychodynamic in orientation but none of the 16 had previously had formal training in brief dynamic therapy.

*Patients.* The patient sample was comprised of 64 patients, half of whom received treatment before their therapists received TLDP training (the *pretraining cohort*) and half of whom were treated after their therapists had received training (the *posttraining cohort*). They were selected from adults responding to advertisements for low-cost psychotherapy. Selection criteria were designed to identify a patient population that (a) had genuine and clinically significant symptomatic distress, as evidenced by a Symptom Checklist 90, revised (SCL-90-R) Global Severity Index T-score at screening of at least 40 on outpatient norms (SCL-90-R; Derogatis, 1983); (b) gave an indication in the screening interview that their problems had a clear interpersonal component for which TLDP would be an appropriate treatment (Strupp & Binder, 1984); and (c) did not have strong indicators for alternative treatments such as pharmacotherapy, substance abuse treatment, or hospitalization. The mean age was 42 years ( $SD = 11.1$ , range 24–64); 78% were female; and one was African-American, one Asian, and the rest Caucasian. Seventy-three percent had done college course work, and 50% were college graduates. Forty-four percent were married, 16% were single, 34% were divorced or separated, and 6% were widowed. Seventy-two percent had been in therapy before.

*Diagnostic and Statistical Manual of Mental Disorders, third edition* (DSM-III; American Psychiatric Association, 1980) diagnoses were assigned by a trained interviewing clinician using the computerized version of the NIMH Diagnostic Interview Schedule, and all patients received an Axis I and/or Axis II diagnosis. Ninety-two percent of the patients received an Axis I diagnosis. Twenty-seven percent of the sample were diagnosed with an anxiety disorder, and 70% additionally or alternately were diagnosed with an affective disorder (36% of the sample had a diagnosis of major depression). Sixty-six percent received an Axis II personality disorder (PD) diagnosis (38% of the sample mixed or other PD, 2% cluster A PD, 5% cluster B PD, 22% cluster C PD). The mean Global Severity Index T-score at the start of treatment was 45.5 ( $SD = 8.6$ , range 19–77). The mean Global Assessment Scale score (Endicott, Spitzer, Fleiss, & Cohen, 1976) at the start of treatment, as rated by intake interviewer, was 59.4 ( $SD = 6.1$ , range 45–68).

## MEASURES

*Global Assessment Scale (GAS; Endicott et al., 1976).* The GAS is a widely used 100-point anchored scale for assessing a patient's lowest overall level of functioning during the previous week. It is completed by a clinician. High scores indicate better functioning.

*Global Outcome Rating (GOR).* The GOR is an 11-point scale indicating how much improvement a patient has made since beginning therapy (+5 = very greatly improved, 0 = no change, -5 = very much worse). It can be completed by the patient or by a clinician.

*Interpersonal Dependency Inventory (IDI; Hirschfeld, Klerman, Gough, Barrett, Korchin, & Chodoff, 1977).* The IDI, containing 48 four-point items, is a self-report measure of maladaptive interpersonal dependency. It consists of three nonoverlapping scales: Emotional Reliance on Another Person (ER); Lack of Social Self-Confidence (SSC); and Assertion of Autonomy (AA). Frank, Kupfer, Jacob, and Jarrett (1987) found autocorrelations of .77, .85, and .61, respectively, for these scales on IDI administrations 17 weeks apart, which can be taken as lower bounds on the scales' test-retest reliabilities. No generally accepted total score formula has been established (Bornstein, 1994). The Bornstein, Manning, Krukonis, Rossner, & Mastro Simone (1993) total score formula (ER + SSC - AA) is used here. High scores indicate greater dependency.

*Millon Comprehensive Multiaxial Inventory (MCMI; Millon, 1983).* The MCMI, containing 175 true-false items, is a self-report measure of personality style and symptomatology. It contains 11 overlapping scales measuring personality styles and disorders and nine scales measuring symptomatology. Millon (1983) reported one-week test-retest reliabilities between .81 and .91 and KR-20 reliabilities between .73 and .95 for the personality scales. There are no generally accepted schemes for combining scale scores into total scores (Choca, Shanley, and Van Denberg, 1992). Four personality scales (Dependent, Borderline, Passive-Aggressive, and Avoidant) had mean pretreatment elevations above 60 (the outpatient median) in both the pretraining and posttraining cohorts, and these were chosen to assess change in interpersonal functioning.

*SCL-90-R (Derogatis, 1983).* The SCL-90-R, containing 90 five-point items, is a multidimensional self-report symptom inventory. The Global Severity Index (GSI), which is the mean item score, reflects a patient's overall level of symptomatic distress. Edwards, Yarvis, Mueller, Zingale, and Wagman (1978) reported a coefficient alpha of .95 and test-retest reliability of .94 for the GSI of the SCL-90. GSI raw scores were converted here to normalized T-scores based on separate norms for male and for female outpatients.

*INTREX Introject Questionnaire (Benjamin, 1983).* The INTREX Introject Questionnaire, containing 36 items rated 0-100, is a self-report measure of the way patients treat themselves. Its eight nonoverlapping scales represent the octants of the structural analysis of social behavior (SASB) circumplex, which is characterized by an affiliation dimension and an orthogonal autonomy dimension. An attack pattern coefficient (ATK) and a control pattern coefficient (CON) can be derived from the pattern of octant scores, both taking values between +1 and -1. ATK represents the degree to which the patient is self-loving (negative scores) or self-hating (positive scores). CON represents

the degree to which the patient is self-emancipating (negative scores) or self-controlling (positive scores).

*Capacity for Dynamic Psychotherapy Scale (CDPS; Thackrey, Butler, & Strupp, 1993).* The CDPS, containing nine Likert-scale items (1 = minimal, 5 = maximal, with half-point increments), measures a patient's current willingness and ability to collaboratively address problems of an affective and interpersonal nature. Items assess psychological mindedness, motivation, and hostility or suspiciousness. The CDPS total score was used to assess suitability for brief dynamic therapy, as traditionally understood.

Coefficient alpha for 31 pretraining cases for CDPS completed by intake interviewers was .96, and the correlation between CDPS total scores completed by intake interviewers and other raters (working from the videotape of the interview) was .72 (Thackrey et al., 1993). Thackrey et al. also reported positive correlations between CDPS total scores and several measures of posttreatment outcome for these nonTLDP cases.

#### PROCEDURE

Therapy sessions in the study were conducted in the therapists' offices and were audiotaped, except for sessions 3 and 16, which were conducted and videotaped in a research office. Patients who completed fewer than five sessions were considered dropouts and replaced. This occurred twice with each cohort. There was also a case lasting four sessions that mistakenly was not replaced. The 64 cases in the study averaged 20.8 sessions (SD = 6.6, range 4–25).

In the pretraining phase, therapists were assigned two patients and asked to conduct brief (25-session limit) therapy with these patients, using their customary treatment methods. As in the posttraining phase of the project, most therapists were assigned a patient above and below the CDPS sample median. When these cases terminated, a year-long training in TLDP commenced. Therapists were divided into training groups of four, with the two authors of the training manual each responsible for training two of the groups. Training consisted of approximately 50 weekly two-hour seminar and supervision sessions, which were audiotaped. The program included: (a) didactic presentations of TLDP principles and techniques, accompanied by readings from the manual and illustrated by excerpts from transcripts and audio- or videotapes of the seminar participants and the supervisors; and (b) treatment of a training case (whose data are not included in this study), with small-group supervision. Audio- and videotaped excerpts of training case sessions were discussed in detail by the seminar leaders and the group to highlight key TLDP concepts. As the training cases neared termination, therapists were assigned two additional patients and expected to conduct TLDP (again with a 25-session limit). On average, five training sessions were held after a posttraining cohort case commenced, and during these sessions posttraining cases were often discussed. Therapists completed the GAS after the first session of each therapy, and completed the GAS and GOR after the final session.

Prior to starting treatment, patients completed the SCL-90-R, MCMI, IDI, and INTREX Introject Questionnaire (rating how they treated themselves when they were at their worst) and participated in a videotaped research intake interview with a staff psychologist. At the completion of treatment, patients filled out the same measures plus the GOR and participated in another videotaped research interview. The interviewers completed the CDPS and GAS after the intake interview and completed the GAS and GOR after the termination interview.

Approximately one year after the termination of each therapy, an attempt was made to invite each patient to return for a follow-up interview and to complete the same measures completed at termination. The interviewer again completed the GAS and GOR.

Table 1 contains descriptive statistics for the sample at pretreatment, termination, and one-year follow-up.

**TABLE 1. Descriptive Statistics for Pretraining and Posttraining Cohorts**

Measure	Pretraining cohort			Posttraining cohort		
	<i>M</i> (SD)			<i>M</i> (SD)		
	Pre	Term	F/U	Pre	Term	F/U
GSI <sup>a</sup>	45.52 (10.04) <i>n</i> = 31	39.75 (10.67) <i>n</i> = 32	40.00 (9.27) <i>n</i> = 28	45.53 (7.11) <i>n</i> = 32	42.13 (9.15) <i>n</i> = 30	40.52 (9.26) <i>n</i> = 25
IDI total	55.31 (15.46) <i>n</i> = 32	49.19 (16.02) <i>n</i> = 32	48.71 (16.01) <i>n</i> = 28	56.91 (10.61) <i>n</i> = 32	48.50 (14.45) <i>n</i> = 30	43.58 (14.16) <i>n</i> = 24
Dependent <sup>b</sup>	73.87 (30.30) <i>n</i> = 31	63.50 (32.43) <i>n</i> = 32	60.50 (29.75) <i>n</i> = 28	72.59 (28.39) <i>n</i> = 32	63.83 (26.76) <i>n</i> = 30	57.29 (27.33) <i>n</i> = 24
Avoidant <sup>b</sup>	63.39 (24.52) <i>n</i> = 31	48.53 (26.82) <i>n</i> = 32	47.82 (26.05) <i>n</i> = 28	62.12 (22.04) <i>n</i> = 32	45.07 (28.58) <i>n</i> = 30	45.58 (30.42) <i>n</i> = 24
Passive-aggressive <sup>b</sup>	69.87 (23.53) <i>n</i> = 31	50.19 (29.67) <i>n</i> = 32	42.79 (27.34) <i>n</i> = 28	69.38 (22.17) <i>n</i> = 32	51.93 (28.06) <i>n</i> = 30	49.37 (28.18) <i>n</i> = 24
Borderline <sup>b</sup>	72.58 (14.70) <i>n</i> = 31	57.88 (19.69) <i>n</i> = 32	58.36 (21.05) <i>n</i> = 28	71.41 (11.60) <i>n</i> = 32	58.67 (18.34) <i>n</i> = 30	57.96 (19.41) <i>n</i> = 24
GAS (therapist)	60.50 (10.25) <i>n</i> = 32	74.66 (11.05) <i>n</i> = 32		60.97 (10.65) <i>n</i> = 32	70.53 (12.74) <i>n</i> = 32	
GAS (interviewer)	59.38 (6.07) <i>n</i> = 32	67.65 (9.66) <i>n</i> = 31	66.54 (11.21) <i>n</i> = 28	57.34 (6.72) <i>n</i> = 32	68.73 (10.80) <i>n</i> = 30	69.22 (6.45) <i>n</i> = 23
GOR <sup>c</sup> (patient)		3.47 (1.88) <i>n</i> = 32	3.61 (1.42) <i>n</i> = 28		3.29 (1.58) <i>n</i> = 31	3.44 (1.50) <i>n</i> = 25
GOR <sup>c</sup> (therapist)		3.13 (1.16) <i>n</i> = 32			2.56 (1.41) <i>n</i> = 32	
GOR <sup>c</sup> (interviewer)		2.50 (1.50) <i>n</i> = 31	2.04 (1.80) <i>n</i> = 28		2.90 (1.83) <i>n</i> = 30	3.00 (1.38) <i>n</i> = 23

*Note.* Pre = pretreatment; Term = termination of treatment; F/U = 1-year follow-up; GSI = Global Severity Index of SCL-90-R; IDI = Interpersonal Dependency Inventory; GAS = Global Assessment Scale; GOR = Global Outcome Rating.

Low scores indicate better functioning on all measures except GAS and GOR.

<sup>a</sup>T-scores based on outpatient norms.

<sup>b</sup>Millon Comprehensive Multi-axial Inventory (MCMI) scale scores.

<sup>c</sup>Rated on -5 to 5 scale.

## DATA ANALYSIS

We evaluated our two hypotheses by testing the main effect of cohort and the cohort-by-CDPS (i.e., by supposed patient suitability) interaction. We heeded Martindale's (1978) and Crits-Christoph and Mintz's (1989) recommendations to include a therapist random factor in our analyses, as patients treated by the same therapist are likely to have correlated outcomes, and because we wanted to be able to generalize our findings to a population of therapists (of which the 16 project therapists would be considered a random sample). We therefore conducted mixed model analyses, employing pretreatment-to-posttreatment change scores as dependent variables; therapist and the therapist-by-cohort interaction as random effects; and cohort, CDPS, and the cohort-by-CDPS interaction as fixed effects. Following Littell, Milliken, Stroup, and Wolfinger (1996; p. 176), analyses were simplified by dropping nonsignificant effects.

The outcome measures examined via these analyses were the GSI; therapist and interviewer GAS and GOR; and three variables derived from the interpersonal variables IDI total score and MCMI scales Dependent, Borderline, Passive-Aggressive, and Avoidant. We performed a principal components analysis of the pretreatment scores for these variables in order to reduce the number of outcome variables, following Duntzman's (1989; p. 54) suggestion of representing each component by the variables that loaded highest on it. We found an interpretable three-component solution. Standard scores for IDI total score and Dependent were summed and standardized; the resulting variable was labeled Dependency. Standard scores for Passive-Aggressive and Avoidant were similarly combined and denoted PA. Borderline represented the third component. The patient-rated GOR was also examined, but via a Mann-Whitney test because it was highly skewed. Because our analyses were planned and limited in scope, and in order to avoid a considerable loss of power, we decided not to adjust for familywise error.

We also investigated the extent of clinically significant change (Jacobson & Truax, 1991) for several outcome variables.

**RESULTS**

## TESTS OF OUR HYPOTHESES

The cohort by CDPS interaction was not significant for any outcome variable,  $p > .10$ , indicating a lack of support for our second hypothesis. The cohort main effect was significant for only one outcome variable, therapist GAS,  $F(1, 62) = 5.16$ ,  $p = .03$ , and approached significance for therapist GOR,  $F(1, 62) = 3.02$ ,  $p = .09$ . However, for both variables the pretraining cohort had better mean outcomes than the posttraining cohort.

## CLINICALLY SIGNIFICANT CHANGE

We examined GSI raw scores to determine the amount of clinically significant change made in each cohort. Male and female nonpatient and outpatient norms from Derogatis (1983) were used to compute each sex's cutting score (what Jacobson and Truax [1991] denote  $c$ ) between normal and dysfunctional populations. Twelve (20%) of the 61 patients with GSI data had pretreatment scores already within the normal population and therefore could not make clinically significant change. Of the remaining



patients, eight (33%) of 24 from the pretraining cohort and five (20%) of 25 from the posttraining cohort achieved clinically significant change. The Fisher's exact test of cohort differences was not significant,  $p > .10$ .

We also examined IDI scale scores to determine the amount of clinically significant change made in each cohort. These were used because the test-retest reliability of the total score has not been reported. Hirschfeld et al.'s (1977) cross-validating sample data were used to represent the normal population, and our sample data were used to represent the dysfunctional population. Following Hirschfeld et al. and Bornstein et al. (1993), we computed separate *cs* for males and females. Of the 62 patients with IDI data, 21 (34%), 24 (39%), and 35 (56%) had pretreatment scores already within the normal population on ER, SSC, and AA, respectively, and therefore could not make clinically significant change. Of the remaining patients in the pretraining cohort, 3 (15%) of 20, 2 (10%) of 20, and 1 (8%) of 13 made clinically significant change on the respective subscales. For the posttraining cohort, 2 (9%) of 21, 4 (22%) of 18, and 1 (7%) of 14, respectively, made clinically significant change. The three Fisher's exact tests of cohort differences were not significant,  $p > .10$ .

Finally, clinically meaningful improvement in introject at its worst was examined via a Fisher's exact test. Patients were rated as showing or not showing clinically meaningful improvement according to criteria adapted from Henry, Schacht, and Strupp (1990; p. 771). Meaningful improvement was defined as an ATK increase of at least .50 in the direction of self-love or a CON change from negative to positive while ATK remained positive (self-hating). This latter condition was deemed an improvement because the patient had become less self-destructive.

Sixty-two patients had both pretreatment and termination introject data. The previous definition of introject improvement (Henry et al., 1990) implies that patients with pretreatment ATK less than -.50 cannot improve significantly, as they are sufficiently self-loving so that there is little room for meaningful improvement. Six pretraining cohort patients and four posttraining cohort patients met this criterion at pretreatment, and one from each cohort showed a decline to a self-hating (i.e., positive ATK) introject at termination. Of the remaining patients, 7 (27%) of 26 from the pretraining cohort and 10 (38%) of 26 from the posttraining cohort achieved meaningful improvement. The Fisher's exact test of cohort differences was not significant,  $p > .10$ .

## SECONDARY ANALYSES

Because the preceding analyses did not provide support for our hypotheses, we looked for alternative explanations for these results. One possible explanation, of course, was that TLDP is in fact no more effective than treatment as usual. There were, however, two other plausible explanations we wished to examine.

First, it seemed possible that the posttraining treatments would prove superior to the pretraining treatments when time frames longer than the start-of-treatment-to-termination interval were considered. We wished to examine the possibility that sets of outcomes that appeared roughly equal at termination might exhibit different trajectories over longer periods of time. Because some patients received additional treatment between termination and follow-up and because unknown decision processes led patients to participate or not at follow-up, it is difficult to draw clear conclusions about causality from analyses of follow-up data. Nonetheless, we thought that an examination of the one-year follow-up data was important.

Second, it seemed possible that many of the Vanderbilt II therapists did not in fact achieve proficiency in TLDP. If an appreciable number of the therapists did not



achieve at least a modest level of skill in TLDP, then the Vanderbilt II project would constitute a test of the effects of receiving a certain type and amount of training in TLDP but would not be a true test of the effects of learning or mastering TLDP. To the extent that this was the case, it remains to be seen whether being trained to mastery in TLDP enhances therapeutic effectiveness.

We investigated the plausibility of these explanations via the data analyses described below.

*One-year follow-up results.* Fifty-three of the 64 patients in our sample (83%) agreed to participate in the one-year follow-up (28 [88%] from the pretraining cohort, 25 [78%] from the posttraining cohort). For 51 of the returnees (27 pretraining, 24 posttraining), information was obtained on whether they had received any subsequent psychotherapy and/or psychopharmacotherapy in the year following termination of their project therapy. Nineteen (37%) of the 51 indicated they had continued with their project therapist past their 25th session or had initiated treatment with another treatment provider. Eight (30%) of these returnees were from the pretraining cohort and 11 (46%) were from the posttraining cohort; the Fisher's exact test was not significant,  $p > .10$ . Although the information provided about additional treatment was often sketchy, we were able to determine which patients had a meaningful amount of additional treatment, somewhat arbitrarily defined as consisting of at least four psychotherapy sessions or a month of regular use of a psychotropic medication. Fifteen (29%) patients had a meaningful amount of additional treatment, seven (26%) pretraining patients and eight (33%) posttraining patients; the Fisher's exact test was not significant,  $p > .10$ .

In order to test our hypotheses about the effect of TLDP training on outcome at follow-up, we repeated the same kinds of mixed model analyses described above, but using pretreatment-to-follow-up change scores as dependent variables. Because most of the posttraining patients were missing Introject data, an analysis of introject improvement was not performed. Also, because therapists were not involved at follow-up, there were no therapist ratings to examine. Again, none of the cohort-by-CDPS interactions were statistically significant. The cohort main effect was significant for two dependent variables, Dependency,  $F(1, 49) = 5.02, p = .03$ , and interviewer GOR,  $F(1, 48) = 4.85, p = .03$ , and approached significance for interviewer GAS,  $F(1, 49) = 3.45, p = .07$ . For all three variables, the posttraining cohort had better mean outcomes than the pretraining cohort.

*Assessing therapist proficiency at TLDP.* Session tapes from all 32 posttraining cases were reviewed to assess the extent to which the therapists had conducted TLDP with at least a minimally acceptable level of skill. Because therapists had a single training case, we believed that setting a higher standard (e.g., asking whether the therapy had been conducted at a level suitable for a clinical trial) was unrealistic.

Assessments of therapist skill at TLDP were done on a case-by-case basis by the first two authors. Neither was affiliated with the Vanderbilt II project during the years when the therapies were conducted (1984–1987), and both were blind to the outcomes of all but two of the posttraining cases. Each rater worked separately in reviewing and rating session tapes. For each posttraining case, tapes of session 3 and a session from the latter half of the therapy (session 16 if available, otherwise the second-from-last session) were viewed/listened to in their entirety and rated. After viewing/listening to a tape, a rater reviewed a set of 12 items adapted from a TLDP adherence measure, the Vanderbilt Therapeutic Strategies Scale (Butler, Henry, & Strupp, 1996), that as-

essed key technical aspects of TLDP. The rater then assessed the therapist's performance during the session on a five-point scale (very unskilled at TLDP, unskilled, minimally skilled, skilled, and very skilled). After rating both sessions, the raters met to discuss the case and voted whether or not the therapy was at least minimally skillful TLDP. When a disagreement occurred, the tie was broken by rating a third tape from a session midway between the other two sessions, reviewing videotapes together, or by showing and discussing tape excerpts with colleagues.

The raters were in agreement on 81% of the cases, with a kappa of .54, 95% CI = (.21, .86). In all cases of disagreement, the rater who gave the higher rating assessed the level of TLDP skill as minimally skilled. Two possible rating biases should be noted. In light of the frequently inelegant TLDP offered by the therapists (detailed below), the raters may have been inordinately lenient in rating cases conducted with borderline levels of skill as minimally skilled rather than as unskilled. There also may have been a countervailing bias: The more pervasively therapists were judged as unskillful, the less conclusively the lack of support for the study's hypotheses could be seen as indicative of the therapeutic value of TLDP.

Nine of the 32 posttraining cases (28%) were judged to have been conducted with at least a minimal level of skill at TLDP. Nine of the therapists did not conduct a TLDP case with at least minimal skill, five conducted one of their two cases with at least minimal skill, and two therapists conducted both cases skillfully.

## DISCUSSION

Our analyses of the termination data, in sum, provided no support for either of our hypotheses about the beneficial effects of TLDP training on outcome. Examination of clinically significant change also failed to give clear evidence of superior posttraining cohort functioning.

At the one-year follow-up, the posttraining cohort had improved more than had the pretraining cohort on a composite measure of dependency and on ratings by interviewers. However, because interviewers were members of the research staff and were not blind to the cohort a patient belonged to, they may have been biased in the direction of overestimating posttraining outcomes and/or underestimating pretraining outcomes. It should also be noted that this study did not control for therapist maturation effects, which might partly account for better posttraining outcomes (K. Holroyd, personal communication, March 12, 1997). There was no support for our hypothesis that the TLDP training would be differentially helpful for improving outcomes with supposedly less suitable (i.e., low CDPS) patients.

We believe our semiformal assessment of the skillfulness of the posttraining therapies provides reasonably clear evidence that most of the project therapists failed to achieve a minimally acceptable level of TLDP mastery. Note that our ratings were of therapist performance, not competence. The former term refers to the level of skill evidenced in a particular session or case, while the latter refers to a general capacity, across cases, to perform skillfully. Even very competent therapists may perform poorly in particular sessions or with particular patients. Because project therapists had two posttraining cases, we were reluctant to assess individual therapists for the possession of a general but less-than-universal capacity to perform TLDP skillfully: If a therapist was rated skillful on one case but not the other, it would be hard to determine which case was more indicative of the therapist's usual level of skill (note that this precludes comparisons between TLDP learners and nonlearners).

However, we believe that it is not strictly necessary to distinguish the learners from the nonlearners, as the total number of posttraining cases judged to have been conducted with acceptable skill gives a rough indication of the extent to which TLDP had been learned by the therapist sample at large. Even if our tally of nine TLDP-skillful cases is only approximate, it still seems implausible that the majority of the project therapists had attained a reasonable level of competence. We believe that the nine therapists who failed to conduct a minimally skillful posttraining case were not TLDP learners, and that some of the remaining seven therapists may have been learners.

In our review of posttraining sessions, we frequently found therapists asking patients about the therapeutic relationship out of the blue, or offering hasty transference interpretations that were not adequately grounded in prior session material (see Butler and Strupp [1993; p. 203] for an example). Sometimes therapists' persistence in inappropriately or prematurely focusing on the transference seemed to confuse patients and heighten their anxiety. In a few instances, transference interpretations appeared to be used to express the therapist's frustration or irritation with the patient. We also found that when therapists did offer appropriate transference interpretations, they often seemed unsure of how to follow up so as to deepen the patient's emotional experience and self-understanding. These common behaviors suggested that therapists were still struggling to integrate a new way of conceptualizing and offering treatment (note, though, that because most of the sessions rated were held on the research staff's turf and videotaped rather than audiotaped, therapists may have performed in a somewhat more self-conscious manner than customary). In addition, while therapists were expected to follow the TLDP protocol, they were not constrained to do so. In sum, TLDP that was both skillfully and faithfully conducted was not administered to most of the posttraining patients.

Why was the year-long training insufficient to instill a higher level of skillfulness? As noted above, it seems likely that trainer style (Henry, Schacht, et al., 1993; Anderson, 1997) had an impact on what therapists learned. It is also likely that divided loyalties between customary treatment methods and the TLDP protocol (Weissman, Rounsaville, and Chevron, 1982) limited several therapists' commitment to mastering TLDP. However, it appears that one training case was insufficient for most therapists to achieve autonomous mastery (i.e., without the need of ongoing supervision) of the new treatment modality. The decision to provide a single training case was made in the early 1980s, and was partly based on the exigencies of funding and scheduling. It was also based on a belief in the particular effectiveness of small-group supervision and on an expectation that because most project therapists would have had prior psychodynamic training, the TLDP training would offer a technical adjustment of customary treatment methods rather than a wholly new approach. This expectation was evidently mistaken. While most of the Vanderbilt II therapists identified their orientation as psychodynamic, our review of their videotaped sessions indicated that many had a rather modest command of basic psychodynamic technique. For most project therapists, then, the TLDP training may have been closer to learning a new treatment modality than to reconfiguring or refining a well-learned repertoire of therapeutic skills.

We conclude with a caution about training and a call for further research. If research validates the worth of some forms of manualized treatment in clinical settings, we believe it is questionable whether these therapies can be adequately learned by practicing clinicians if current levels of investment in continuing education are maintained. If our experience with TLDP is at all representative, then reading a manual

on one's own, attending a workshop, and trying out new methods without supervision will probably not suffice for most clinicians trying to master a new treatment modality.

In addition, we call for more research on the effective implementation of manual-guided training. The primary aim of manual-guided training, as with any educational enterprise, is to help students master a body of knowledge and a set of skills, and the training is successful to the extent that it realizes this. Moras (1993), for example, has offered valuable recommendations on how to improve manuals and related teaching aids, and we hope that these and similar ideas will be further explored.

## REFERENCES

- Addis, M. E. (1997). Evaluating the treatment manual as a means of disseminating empirically validated psychotherapies. *Clinical Psychology: Science and Practice, 4*, 1–11.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: APA.
- Anderson, T. (1997, June). Educating therapists: The training phase of the Vanderbilt II Psychotherapy Research Project. In K. Moras (Chair), *The influence of Hans Strupp on psychotherapy research: Training therapists*. Symposium conducted at the annual international meeting of the Society for Psychotherapy Research, Gelso, Norway.
- Benjamin, L. S. (1983). *The INTREX user's manual, Part I and II* [Computer program manual]. Madison, WI: Intrex Interpersonal Institute, Inc.
- Bornstein, R. F. (1994). Construct validity of the Interpersonal Dependency Inventory: 1977–1992. *Journal of Personality Disorders, 8*, 64–76.
- Bornstein, R. F., Manning, K. A., Krukonis, A. B., Rossner, S. C., & Mastrosimone, C. C. (1993). Sex differences in dependency: A comparison of objective and projective measures. *Journal of Personality Assessment, 61*, 169–181.
- Butler, S. F., Henry, W. P., & Strupp, H. H. (1996). Measuring adherence in time-limited dynamic psychotherapy. *Psychotherapy, 32*, 629–638.
- Butler, S. F., & Strupp, H. H. (1993). Effects of training experienced dynamic therapists to use a psychotherapy manual. In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 191–210). New York: Basic Books.
- Choca, J. P., Shanley, L. A., & Van Denburg, E. J. (1992). *Interpretative guide to the Millon Clinical Multiaxial Inventory*. Washington, DC: American Psychological Association.
- Crits-Christoph, P., & Mintz, J. (1991). Implications of therapists 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, MD: Johns Hopkins University Press.
- Duntleman, G. H. (1989). *Principal components analysis* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-069). Thousand Oaks, CA: Sage.
- Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test-taking and the stability of adjustment scales: Can we assess patient deterioration? *Evaluation Quarterly, 2*, 275–291.
- 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*, 766–771.
- Frank, E., Kupfer, D. J., Jacob, M., & Jarrett, D. (1987). Personality features and response to acute treatment in recurrent depression. *Journal of Personality Disorders, 1*, 14–26.
- Goldfried, M. R., Greenberg, L. S., & Marmar, C. (1990). Individual psychotherapy: Process and outcome. *Annual Review of Psychology, 41*, 659–688.
- Gomes-Schwartz, B. (1978). Effective ingredients in psychotherapy: Prediction of outcome from process variables. *Journal of Consulting and Clinical Psychology, 46*, 1023–1035.
- Henry, W. P., Schacht, T. E., & Strupp, H. H. (1990). Patient and therapist introject, interpersonal process, and differential psychotherapy outcome. *Journal of Consulting and Clinical Psychology, 58*, 768–774.
- 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.
- Hirschfeld, R. M., Klerman, G. L., Gough, H. G., Barrett, J., Korchin, S. J., & Chodoff, P. (1977). A measure of interpersonal dependency. *Journal of Personality Assessment*, 41, 610–618.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for mixed models*. Cary, NC: SAS Institute Inc.
- Malan, D. H. (1976). *The frontier of brief psychotherapy: An example of the convergence of research and clinical practice*. New York: Plenum Press.
- Martindale, C. (1978). The therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 46, 1526–1530.
- Millon, T. (1982). *Millon Clinical Multiaxial Inventory Manual*. Minneapolis, MN: National Computer Systems.
- Moras, K. (1993). The use of treatment manuals to train psychotherapists: Observations and recommendations. *Psychotherapy*, 30, 581–586.
- Schulte, D., Kunzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Therapy and Research*, 14, 67–92.
- Sifneos, P. (1979). *Short-term dynamic psychotherapy: Evaluation and technique*. New York: Plenum Press.
- Strupp, H. H. (1980a). Success and failure in time-limited psychotherapy: A systematic comparison of two cases (Comparison 1). *Archives of General Psychiatry*, 37, 595–603.
- Strupp, H. H. (1980b). Success and failure in time-limited psychotherapy: A systematic comparison of two cases (Comparison 2). *Archives of General Psychiatry*, 37, 708–716.
- Strupp, H. H. (1980c). Success and failure in time-limited psychotherapy: With special reference to the performance of a lay counselor (Comparison 3). *Archives of General Psychiatry*, 37, 831–841.
- Strupp, H. H. (1980d). Success and failure in time-limited psychotherapy: Further evidence (Comparison 4). *Archives of General Psychiatry*, 37, 947–954.
- 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 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.
- Thackrey, M., Butler, S. F., & Strupp, H. H. (1993). The Capacity for Dynamic Process Scale (CDPS). In M. L. Canfield & J. E. Canfield (Eds.), *A collection of psychological scales* (pp. 57–63). Bartlesville, OK: Research, Evaluation & Statistics.
- Wilson, G. T. (1996). Manual-based treatments: The clinical application of research findings. *Behaviour Research and Therapy*, 34, 295–314.
- Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1982). Training psychotherapists to participate in psychotherapy outcome studies. *American Journal of Psychiatry*, 139, 1442–1446.

### Zusammenfassung

Die hier beschriebene Studie untersucht die Wirkungen eines Trainings in manualisierter Kurzzeittherapie auf das Therapieergebnis. Als Teil des Vanderbilt II Projektes behandelten jeweils 16 Therapeuten (8 Psychiater und 8 klinische Psychologen) zwei mittelgradig beeinträchtigte erwachsene Patienten mit den ihnen vertrauten Kurzzeitbehandlungsmethoden. Danach wurden sie über ein Jahr in einer manualisierten Form psychodynamischer Kurzzeittherapie ausgebildet und wandten diese Therapie auf zwei weitere Patienten an. Es wurde angenommen, dass die Ausbildung in besseren Ergebnissen resultieren würde und dass Patienten, die man üblicherweise für eine psychodynamische Kurzzeittherapie eher ungeeignet erachten würde, deutlichere Verbesserungen erreichen. Die Ergebnisdaten, die bei Ende der Therapie erhoben wurden, konnten diese Hypothese nicht bestätigen. Maße für die interpersonale Abhängigkeit zu einem Einjahreskatamnesezeitpunkt waren mit der ersten Hypothese konform, aber inkonsistent mit der zweiten. Eine systematische Untersuchung der 32 nach der Ausbildung behandelten Fälle legt nahe, dass die Mehrzahl der Therapeuten keine grundlegende Kompetenz in zeitlich limitierter dynamischer Psychotherapie erworben hatte.

### Résumé

Cette étude explore les effets sur les résultats thérapeutiques de la formation des thérapeutes dans une thérapie brève, manualisée. Dans le cadre du projet Vanderbilt II, chacun des 16 thérapeutes (8 psychiatres et 8 psychologues cliniciens) a traité deux patients adultes modérément perturbés, chacun par sa méthode

habituelle de thérapie brève ; ensuite ils ont reçu une formation d'une année dans une méthode manualisée de thérapie brève, la Psychothérapie Dynamique Limitée dans le Temps (TLDP) ; pour finir, ils ont appliqué la TLDP à deux patients supplémentaires. Les hypothèses prédisaient que la formation allait aboutir à de meilleurs résultats en général, et qu'une amélioration comparativement plus importante sera observée chez des patients habituellement considérés comme moins adaptés à une thérapie dynamique brève. Les résultats obtenus à la terminaison n'ont pu confirmer aucune de ces hypothèses. Des mesures de la dépendance interpersonnelle une année après étaient consistantes avec la première hypothèse, mais les données catamnestiques étaient inconsistantes avec la deuxième. Une revue systématique des 32 cas après formation a suggéré que la majorité des thérapeutes n'avait pas acquis une compétence de base en TLDP.

### **Resumen**

Este estudio exploró la efectividad del entrenamiento de terapeutas en una terapia breve manualizada. Como parte del proyecto Vanderbilt II, dieciséis terapeutas (ocho psiquiatras y ocho psicólogos clínicos) trataron dos pacientes adultos moderadamente perturbados con sus métodos habituales de tratamiento a corto plazo. Estos terapeutas recibieron luego un año de entrenamiento en terapia dinámica breve manualizada que se llamó Psicoterapia Dinámica de Tiempo limitado (TLDP). Finalmente, administraron esta TLDP a dos pacientes más. La hipótesis fue que el entrenamiento llevaría, en general, a resultados mejores y que habría mayor mejoramiento en pacientes menos aptos para una terapia dinámica breve. Los resultados a la terminación no apoyaron ninguna de las dos hipótesis. Las mediciones de dependencia interpersonal al año de seguimiento resultaron consistentes con la primera hipótesis, mientras que los resultados de seguimiento fueron inconsistentes con la segunda. Una revisión sistemática de los treinta y dos casos luego del entrenamiento sugiere que la mayoría de los terapeutas no había logrado una competencia básica en el TLDP.

*Received March 13, 1998*

*Revision received May 27, 1999*

*Accepted June 3, 1999*