

## Training in Cognitive, Supportive–Expressive, and Drug Counseling Therapies for Cocaine Dependence

Paul Crits-Christoph, Lynne Siqueland, Jesse Chittams, Jacques P. Barber, and Aaron T. Beck  
University of Pennsylvania

Arlene Frank  
Brookside Hospital

Bruce Liese  
University of Kansas

Lester Luborsky, David Mark, and Delinda Mercer  
University of Pennsylvania

Lisa Simon Onken  
National Institute on Drug Abuse

Lisa M. Najavits  
McLean Hospital and Harvard Medical School

Michael E. Thase  
Western Psychiatric Institute and Clinic,  
University of Pittsburgh

George Woody  
University of Pennsylvania

This study assessed the effects of training on the performance of 65 therapists in delivering manual-guided therapies to 202 cocaine-dependent patients. Changes in ratings of therapists' adherence and competence was assessed in 3 treatment modalities: supportive–expressive dynamic therapy (SE), cognitive therapy (CT), and individual drug counseling. Effects of manual-guided training on the therapeutic alliance were also assessed. Training effects were examined through a hierarchical linear modeling approach that examined changes both within cases and across training cases. A large effect across cases was detected for training in CT. Supportive–expressive therapists and individual drug counselors demonstrated statistically significant learning trends over sessions but not over training cases. Training in SE and CT did not have a negative impact on the therapeutic alliance, although alliance scores for trainees in drug counseling initially decreased but then rebounded to initial levels.

The nature and success of training in psychotherapy is an issue that not only affects various training programs (e.g., clinical psychology, psychiatric residencies, and social work training) but also has implications for how best to disseminate new psychosocial treatments to existing practitioners. Stein and Lam-

bert (1995) have reviewed the literature related to the impact of graduate training in psychotherapy on therapist behaviors and patient outcome. Research related to clinical supervision, a key element in any training program, has also been recently reviewed by Holloway and Neufeldt (1995).

Paul Crits-Christoph, Lynne Siqueland, Jesse Chittams, Jacques P. Barber, Aaron T. Beck, Lester Luborsky, David Mark, Delinda Mercer, and George Woody, Department of Psychiatry, University of Pennsylvania; Arlene Frank, Brookside Hospital, Nashua, New Hampshire; Bruce Liese, Department of Family Medicine, University of Kansas; Lisa Simon Onken, National Institute on Drug Abuse (NIDA), Bethesda, Maryland; Lisa M. Najavits, Department of Psychiatry, McLean Hospital, Cambridge, Massachusetts, and Department of Psychiatry, Harvard Medical School; Michael E. Thase, Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh.

The NIDA Collaborative Cocaine Treatment Study is an NIDA-funded cooperative agreement involving four clinical sites, a coordinating center, and NIDA staff. The Coordinating Center at the University of Pennsylvania, Department of Psychiatry includes Paul Crits-Christoph (Principal Investigator), Lynne Siqueland (Project Coordinator), Karla Moras (Assessment Unit Director), Jesse Chittams (Director of Data Management/Analysis), and Larry Muenz (Statistician). The collaborating scientists at the Treatment Research Branch, Division of Clinical and Research Services at NIDA include Jack Blaine and Lisa Simon Onken. The four participating clinical sites are University of Pennsylvania, Department of Psychiatry (Lester Luborsky, Principal Investigator;

Jacques P. Barber, Co-Principal Investigator; and Delinda Mercer, Project Director), Brookside Hospital/Harvard Medical School (Arlene Frank, Principal Investigator; Stephen F. Butler, Co-Principal Investigator; Sarah Bishop, Project Director), McLean/Massachusetts General Hospital–Harvard University Medical School, Department of Psychiatry (Roger D. Weiss, Principal Investigator; David R. Gastfriend, Co-Principal Investigator; Lisa M. Najavits and Margaret L. Griffin, Project Directors), University of Pittsburgh/Western Psychiatric Institute and Clinic, Department of Psychiatry (Michael E. Thase, Principal Investigator; Dennis Daley and Ishan M. Salloum, Co-Principal Investigators; and Judy Lis, Project Director). The Training Unit includes Heads of Cognitive Therapy Training Unit Aaron T. Beck (University of Pennsylvania, Department of Psychiatry) and Bruce Liese (University of Kansas Medical Center, Department of Family Medicine); Heads of Supportive–Expressive Therapy Training Unit Lester Luborsky and David Mark (University of Pennsylvania, Department of Psychiatry); Heads of Individual Drug Counseling George Woody (Veterans Administration Medical Center/University of Pennsylvania, Department of Psychiatry) and Delinda Mercer; and Group Drug Counseling Unit Delinda Mercer (Head), Dennis Daley, and Gloria Carpenter (Assistant Head, Treatment Research Unit—University of Pennsylvania, Department of Psychiatry).

Only a few studies, however, have investigated the effects of training psychotherapists in specific modalities using treatment manuals. These studies have compared adherence or competence on pretraining to posttraining cases or compared such ratings for sequential training cases. With regard to interpersonal therapy (IPT) for depression, Rounsaville, Chevron, Weissman, Prusoff, and Frank (1986) reported on the training of IPT therapists for both the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP; Elkin, Parloff, Hadley, & Autrey, 1985; Elkin, Shea, et al., 1989) and for the Frank et al. (1990) maintenance therapies for recurrent depression trial. The TDCRP sample involved 11 therapists, and the maintenance therapies for recurrent depression sample consisted of 7 therapists. Competence ratings made by supervisors were compared for the first training case versus subsequent training cases for each therapist. For the TDCRP sample, the IPT therapists were found to perform at a high level on their first supervised case following review of the treatment manual and a 5-day didactic seminar, but no improvement in competence was found with subsequent cases, perhaps because of a ceiling effect (i.e., the therapists were already performing adequately so further improvement may not have been necessary). The therapists in the maintenance therapies for depression study, who were significantly younger and less experienced than the TDCRP therapists, demonstrated acceptable competence on their first case and improvement on their second case (Cohen's *d* effect size = .54), although the statistical significance of the finding was not reported.

Henry, Schacht, Strupp, Butler, and Binder (1993) studied the effects of training of 16 therapists in manual-guided time-limited dynamic therapy. Ratings on a number of dimensions were made for two sessions drawn from each of two pretraining cases and two posttraining cases. Training consisted of weekly didactic-supervisory group meetings for 1 year, with each therapist having one training case. After training, therapists showed increased use of specific techniques (e.g., exploration of the therapeutic relationship), as well as improvement in interviewing style and general therapeutic techniques not particularly targeted in the training (expression of in-session affect, greater use of open-ended questions and improved participant-observer stance). Therapists also showed an increase in complex thought units (possibly indicating implied criticism) after training, compared with before training. This latter findings has been interpreted to suggest that treatment manuals may have a negative impact on the therapeutic relationship.

---

The Monitoring Board includes Larry Beutler, Jim Klett, Bruce Rounsaville, and M. Tracie Shea. The contributions of John Boren (NIDA), the project officer for this cooperative agreement, are also gratefully acknowledged.

The preparation of this article was funded in part by Grants NIDA U18-DA07090, NIDA U18-DA07663, NIDA U18-DA07673, NIDA U18-DA07693, NIDA U18-DA07085, National Institute of Mental Health (NIMH) Clinical Research Center P30-MH-45178, NIMH Career Development Award K02-MH00756, and NIDA Research Scientist Award K05-DA00168.

Correspondence concerning this article should be addressed to Paul Crits-Christoph, Room 700, 3600 Market Street, Philadelphia, Pennsylvania 19104. Electronic mail may be sent to crits@landru.cpr.upenn.edu.

One pilot study on the effectiveness of training in experiential therapy has been reported by Daldrup, Beutler, Engle, and Greenberg (1988). Four experienced psychotherapists were trained in focused expressive therapy (Daldrup et al., 1988). Ratings of work samples submitted as part of the selection process were compared with three therapy samples (per therapist) observed during the final phase of the training process using a focused expressive therapy compliance scale. The post-training scores showed better compliance than pretraining scores (Cohen's *d* effect size = .93).

Although the aforementioned studies provide some mixed support for the value of training in manual-guided psychotherapies, limitations because of relatively small samples of therapists or few training cases are apparent. The relatively small samples of therapists in these previous studies raise issues of generalizability as well as statistical power to detect effects. Small numbers of training cases per therapists may also limit the ability to detect a training effect if, in fact, several training cases are needed before therapists are able to implement a manual-guided psychotherapy with skill.

Despite the relative scarcity of data on the effects of manual-guided psychotherapy training, it has become a prerequisite for an outcome trial to conduct a preliminary therapist training phase typically consisting of supervised casework as well as didactic seminars. It is important to begin to document the success or failure of such training phases for different treatment modalities and training programs so that future studies can be designed accordingly. Moreover, as mentioned earlier, the relative success of training in manual-guided psychotherapies has relevance to the dissemination of such treatments to clinical training programs and clinical practice.

The purpose of the present study was to evaluate the effects of training on therapist competence for three manual-guided psychosocial treatment modalities for cocaine dependence: cognitive therapy (CT), supportive-expressive dynamic therapy (SE), and individual drug counseling (IDC). In addition to therapist competence at delivering the manual-guided treatments, the present study examined the impact of training on the therapeutic alliance to see if a focus on manual-based techniques might have a negative impact on the therapeutic relationship.

This study overcomes some of the limitations of previous studies in terms of number of therapists and number of training cases. Moreover, recent advances in statistical techniques, such as hierarchical linear modeling (HLM), permit more sophisticated analyses of the effects of training over time. In particular, these techniques allow for an examination of whether therapists improve in the implementation of a manual-guided therapy as each case develops and whether learning is carried over from training case to training case. We hypothesized that training would increase the competence in the delivery of the interventions outlined in the treatment manuals both within training cases (across sessions) and across training cases for each of the three treatment modalities.

## Method

### Overview of Training Phase

Therapists were selected for participation in a training phase study, with therapists who achieved adequate levels of competence during train-

ing later invited to participate in a subsequent clinical trial evaluating the efficacy of three psychosocial treatments for patients diagnosed with cocaine dependence (see Crits-Christoph et al., 1997, for a description of the subsequent clinical trial). The three individual treatments studied were SE (Mark & Faude, 1995; Mark & Luborsky, 1992), a psychodynamic treatment derived from Luborsky (1984); CT (Beck, Wright, Newman, & Liese, 1993), based on Beck's cognitive model (Beck, 1976); and IDC (Mercer & Woody, 1992), based on the 12-step addiction model. In the training study, patients were randomized to one of the treatment conditions after a brief stabilization phase in which patients had to achieve a period of initial abstinence measured by three consecutive drug-free urine screens.

The active phase of the treatment was 6 months long and consisted of sessions twice weekly for the first 3 months of treatment and once weekly for Months 4 through 6. Three monthly booster sessions were offered to patients after the 6-month active phase. In addition to the individual therapy, patients also received group drug counseling. A control condition of group drug counseling alone was also piloted during the training phase of the project, but the limited number of therapists in this condition limited any statistical analysis of training effects; therefore, data from this control condition are not reported here.

The goal of the training phase was to have each individual therapist and counselor treat 4 patients who stayed in treatment for at least 1 month. If a supervisor judged that a therapist or counselor needed more training, then additional training cases were assigned. Participation in this project was demanding of therapists, both in terms of their time and in terms of the expectation that they would learn the respective treatment approach adequately so that they would be certified as competent enough to continue into the subsequent main clinical trial.

## Patients

The inclusion criteria for the study consisted of a primary diagnosis of current cocaine dependence diagnosis, the use of cocaine in the past 30 days, ages 18–60, and a stable mailing address with plans to reside in the area for the next 2 years. Exclusion criteria for diagnoses included primary diagnosis of current polysubstance dependence or alcohol dependence (patients could have current other substance dependence diagnoses as long as the cocaine dependence was primary), current opioid dependence, bipolar or psychotic disorder, dementia or organic brain disorder. Other exclusion criteria included imminent suicidal or homicidal risk, need for or unwillingness to discontinue current psychotropic medications, life-threatening illness that would prevent attendance at sessions, impending incarceration, being mandated to treatment, or residence in a halfway house.

The patient sample included 202 patients who were randomized to treatment and had a minimum number of adherence–competence scores (discussed later). At intake, the average age of the sample was 33.3 years ( $SD = 6.7$ ), and 31% of the patients were female. Fifty-seven percent were Caucasian, and 43% were from ethnic minority groups, primarily African American (3% Hispanic or Native American). Sixty-two percent of the patients were employed. Seventy-five percent of the patients lived alone, and 25% were married or lived with a partner. Seventy-three percent were primarily crack smokers, 23% were primarily intranasal drug users, and 4% primarily injected cocaine. On average, at the time of intake, patients were using cocaine 8.9 days/month ( $SD = 7.8$ ) and had been using cocaine for 6.2 years ( $SD = 4.5$ ). Fifty-two percent of patients had other current substance dependence diagnoses, with 37% meeting criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., revised; *DSM-III-R*; American Psychiatric Association, 1987) for current alcohol dependence and 11% for current cannabis dependence. In addition, 28% had at least one Axis I disorder other than substance dependence, including 15% with a current depressive disorder (10% current major depression). Forty-four percent

of the sample qualified for a personality disorder diagnosis based on the Structural Clinical Inventory for the *DSM-III-R* (2nd ed.; SCID-II; Spitzer, Williams, Gibbon, & First, 1990) with the most common diagnosis being antisocial personality disorder (14% of the sample).

## Therapists

Therapists were selected by the respective supervisors for each therapy modality on the basis of a combination of background education and training, letters of reference, and two audiotaped samples of their therapy work, which were rated for the quality of the therapy delivered. For SE, 50% of the applicants were selected to begin training. Of those who applied for CT, 71% were judged suitable. For IDC, 74% were selected to begin training.

A total of 65 therapists were included in the analysis, with 20 in CT, 25 in SE, and 20 in IDC. To be included in the analysis, a therapist had to have at least one training case that was rated on three sessions for adherence–competence. To have three sessions rated, a therapist had to have a case that generally included at least eight sessions (see schedule of ratings below). This criterion of three rated sessions was set so that a slope (and associated standard error) could be estimated across sessions. Therapist training data were included in our study, regardless of whether the therapist subsequently dropped out of the study or was not certified for continuing into the main clinical trial. Table 1 presents demographic and descriptive information on the therapists, categorized by treatment modality.

## Raters

*Clinical supervisors.* The main raters used to assess adherence–competence were the clinical supervisors for the respective therapies. Supervisors were chosen by the heads of the training units on the basis of

Table 1  
*Descriptive Characteristics of Therapists and Counselors*

Characteristic	CT ( <i>n</i> = 20)	SE ( <i>n</i> = 25)	IDC ( <i>n</i> = 20)
Gender			
Male	15	15	8
Female	5	10	12
Race			
Caucasian	19	24	15
African American	1	0	5
Asian	0	1	0
Degree			
PhD, PsyD, or EdD	16	17	0
MD	0	1	0
MSW	4	2	2
MA	0	4	6
BA, Associate, or RN	0	0	12
Age			
<i>M</i>	40.1	40.0	40.7
<i>SD</i>	6.1	5.4	6.7
Years of experience			
<i>M</i>	9.6	8.7	8.2
<i>SD</i>	7.7	5.0	5.5

*Note.* CT = cognitive therapy; SE = supportive–expressive dynamic therapy; IDC = individual drug counseling; PhD = doctor of philosophy; PsyD = doctor in psychology; EdD = doctor of education; MD = doctor of medicine; MSW = master of social work; MA = master of arts; BA = bachelor of arts. Years of experience = years since completing graduate degree for SE and CT and years practicing as drug counselor for IDC.

their knowledge and experience within dynamic, CT or drug counseling modalities, and most were coauthors of the treatment manuals for each modality. There were three SE supervisors, four CT supervisors, and four IDC supervisors. All supervisors had extensive supervision experience within the general treatment modality, and most had previous experience in training therapists with the specific modality described in the treatment manual. Supervisors coordinated their training through conference calls and meetings in preparation for training workshops. Each supervisor rated the therapies for which he or she provided clinical supervision.

**Independent judges.** Separate ratings were also obtained by judges who were unaware of the clinical progress on each case and unaware of the sequence of training cases for each therapist. The same pool of supervisors made these ratings, but each supervisor rated only patients for whom he or she was not the clinical supervisor. Supervisors rated only patient tapes from the modality for which the supervisor was an expert (i.e., CT supervisors rated only CT cases, etc.). To limit the rating task, we rated two sessions of 3 different patients per therapist. Because only two sessions per patient were rated, these ratings could not be used to examine within-case trends. However, these ratings did serve as a check on reliability and did confirm trends across training cases. For analysis of learning trends, the ratings from the clinical supervisor of each case were examined separately from the ratings made by the independent judges who did not supervise the cases that they rated.

### Training Process

Once accepted into the study, therapists were first given a copy of the manual and other training materials to review. Over the course of the training period, the therapists also attended four training workshops that consisted of didactic presentations, role plays, and discussion of case examples and difficulties encountered implementing the treatment. The training unit directors, all supervisors, and all therapists attended these 2-day workshops. The first workshop (held before the assignment of the first training case) was primarily didactic and focused on the theoretical foundations of the treatment. The second workshop, and especially the third and fourth workshops, focused on reviewing techniques and case formulations using patients with cocaine dependence who had been or were currently in treatment with the pilot study therapists. Through the use of videotapes and audiotapes, problematic and successful cases were discussed.

Each therapist was assigned four training cases (of at least 1-month duration), usually one at a time, over the course of the training phase. All therapy sessions were audiotaped and mailed to the supervisor within 2–3 days. Supervision occurred over the telephone and was focused on the feedback from the ratings of adherence and competence made by the supervisor for that particular session and over the course of training. Sessions 2, 4, 8, 12, 18, 24, 30, and 36 were scheduled to be rated by the therapist's supervisor after the supervisor listened to the entire tape. Therefore, eight sessions were rated for adherence–competence for patients who completed the full treatment. However, clinical (i.e., patient crisis) and training needs (therapist below acceptable adherence–competence) allowed supervisors to listen to and rate other sessions out of the above sequence. The average patient in this study had 5.3 ( $SD = 1.7$ ) sessions rated by the supervisor.

### Measures

The Cognitive Therapy Scale (CTS; Young & Beck, 1980) was used in rating competence in the CT condition. Previous research has shown that the CTS adequately assesses variations in session quality (Vallis, Shaw, & Dobson, 1986). The CTS was also used to evaluate competence of CT therapists during the training phase of the TDCRP (Shaw, 1984): This scale includes items on general therapeutic skills (such as setting

an agenda) and items on cognitive conceptualization, strategy, and technique (e.g., guided discovery, focusing on key cognitions, strategy for change, and homework). CTS scores on each item range from 0 (*poor*) to 6 (*excellent*). Preliminary reliability analyses with the judges and patients used for the present study indicated that 4 of the original 11 items could not be rated reliably. Internal consistency (Cronbach's alpha) for the CTS total score based on the remaining 7 items was .92. In addition to rating the CTS items, CT judges also rated patient difficulty, because it was hypothesized that the implementation of standard CT techniques would be hindered with a difficult patient, as had been found by Foley, O'Malley, Rounsaville, Prusoff, & Weissman (1987) in IPT.

The Penn Adherence and Competence Scale for Supportive–Expressive Therapy (PACSE) was modified for this project from a general adherence and competence scale designed for SE (Barber & Crits-Christoph, 1996). Only the Expressive subscale was of interest for the present study. Because therapists generally began training with adequate levels of use of supportive techniques and training, supervision focused on the use of expressive techniques. In addition, the use of supportive techniques has not been found to be specific to SE (Barber & Crits-Christoph, 1996; Luborsky, Woody, McLellan, O'Brien, & Rosenzweig, 1982). Both frequency (adherence) and quality (competence) of interventions are rated separately for each item. The Expressive subscale includes items related to the focus on interpersonal relationships and interpretation of the core conflictual relationship themes. Recent research has shown a relationship between competence in the use of expressive techniques, but not supportive techniques, and the outcome of short-term SE for depression (Barber, Crits-Christoph, & Luborsky, 1996).

On the basis of preliminary reliability analyses of items, a final scale of 15 items for adherence and 6 items for competence was used for our study. Item scores ranged from 1 (*not present or poor quality*) to 7 (*high frequency or excellent quality*). Internal consistency in this study assessed by Cronbach's alpha was .85 for the Expressive adherence score and .89 for the Expressive competence score. The Expressive adherence and competence scores were correlated (.62), and both were retained for analysis. In addition to the technique items, SE judges also rated patient difficulty.

The Adherence and Competence Scale for Addiction Counseling (Mercer, Calvo, Krakauer, & Barber, 1994) was developed for this project and includes 17 items categorized into 6 primary sections: monitoring drug use behaviors, encouraging abstinence, supportively confronting negative behaviors, encouraging 12-step participation, relapse prevention, and educating the client. Each of the items was rated for both adherence and competence on scores that ranged from 1 (*not present or poor quality*) to 7 (*highly frequent or excellent quality*). Adherence and competence summary scores were calculated by averaging over the 17 items. Using the supervisors' ratings, we found that the internal consistency scores (Cronbach's alphas) of the adherence and competence scales were .84 and .87, respectively. The correlation between the adherence and competence scores was very high (.90); therefore, only the competence scale was retained for analysis. The authors of the IDC manual and scale did not consider patient difficulty to be an important issue for the implementation of the manual. Consequently, a rating of patient difficulty was not included within the IDC adherence–competence scale.

The therapeutic alliance was assessed using patient-rated versions of two widely used alliance scales, the Helping Alliance Questionnaire—Revised (HAQ; Luborsky et al., 1996) and the California Psychotherapy Alliance Scale (CALPAS; Gaston, 1991). These measures were completed by patients after the 2nd, 5th, and 24th treatment session. Using Session 2 data, we found that the internal consistency (Cronbach's alpha) for the HAQ total score was .89. The CALPAS consists of four subscales and a total score. For the Working Capacity subscale, the internal consistency using Session 2 data was .87; for the Patient Commitment subscale, it was .88; for the Working Strategy subscale, it was .86; for the Therapist

Understanding subscale, it was .87; for the total score, it was .83. Using the Session 2 data, we found that the correlation between the HAQ and CALPAS total scores was .58 ( $p < .0001$ ) in this sample.

### Statistical Analyses

HLM (Bryk & Raudenbush, 1992) was used to examine the effects of training on competence and alliance scores. Several features of HLM were advantageous for analyzing our data. HLM can examine individual, person-specific time trends, as well as group trends, basing its estimates on whatever data are available for each individual. The nature of our data, with varying numbers of sessions per patients, and varying number of patients per therapist due to attrition, cannot be accommodated with standard repeated measures analysis of variance (ANOVA). Moreover, the supervisors did not always keep to the schedule of sessions to be rated because of clinical or training needs to examine other sessions, as well as problems with some tapes. Another advantage of HLM is that it allows for residual autocorrelation. It seems likely that competence in one session might be more highly related to competence in the immediately preceding session than to sessions further away in time, thus producing some autocorrelation that would tend to violate the equal correlation assumption of repeated measures ANOVA.

The HLMs conducted were three-level analyses (sessions within patients within therapists). At the first level, a slope and intercept were estimated for the relation between competence scores and session number. A test of the group trend, as well as individual slopes and intercepts for each patient, is provided. At the second level, the individual patient slopes and intercepts are examined with therapist as the higher level variable. This analysis examined whether the slopes or intercepts show a linear trend over the multiple patients for each therapist. The test for change in intercepts addressed the question of whether therapist learning carried over from case to case, with a therapist beginning each new case (Session 0) at a higher competence level than in previous cases. Using Level 1 (over sessions) slopes as a dependent variable at Level 2 (over patients) examines the question of whether slopes over sessions systematically decrease with each subsequent patient seen by each therapist. This would occur if therapists learn the necessary techniques over the course of the first training case; however, with each subsequent training case, there is less room for improvement over sessions, leading to lower slopes with later training cases. A separate HLM was performed for each treatment modality. Analyses of adherence-competence scores were performed (for CT and SE) with and without patient difficulty as a covariate. Results were virtually identical; therefore, we present here only the results without patient difficulty as a covariate. HLM analyses were also performed using therapeutic alliance measures as the dependent variables.

## Results

### Reliability

In addition to the scale internal consistency reliabilities reported above, a comparison of the supervisors' adherence-competence ratings to the second set of independent ratings allowed for an examination of interjudge reliability. Reliability was assessed after calculation of variance components specifying rater, therapist, patient, and sessions as random effects in an unbalanced design and using restricted maximum likelihood estimation. The analyses revealed consistent mean differences between raters, as well as consistent trends for raters to have different preferences (i.e., higher adherence-competence scores) for different therapists (Rater  $\times$  Therapist interaction). More specifically, there was a strong tendency, especially for SE and CT, for

the rater to give higher ratings to a therapist when the rater had served as the ongoing clinical supervisor for the therapist. Therefore, we calculated intraclass correlation coefficients not including the rater and Rater  $\times$  Therapist components. The exclusion of these components limits the generalizability of any results and in particular restricts any statements about absolute scale values. However, because supervisors rated all sessions within each case and mostly rated all patients for each therapist, statements about relative changes can be made. Because it is likely that in practice the same supervisor would train a given therapist, such relative statements are still meaningful.

For the CTS, the interjudge (single judge) reliability was .69; for IDC competence, it was .70; for SE expressive techniques adherence, it was .59; for SE expressive techniques competence, it was .56. Although these reliabilities are marginal to modest, they are consistent with the single-judge reliabilities found in other uses of these scales for SE (Barber, Crits-Christoph, & Luborsky, 1996) and IDC (Barber, Mercer, Krakauer, & Calvo, 1996). Combining the original (supervisor) and second set (independent judge) of ratings would have allowed for higher reliabilities, but given that the first rater was aware of clinical progress of each case, as well as the lack of ability to examine within case trends with the second set of ratings, we chose to analyze the ratings separately, using the second set to confirm any trends found with the first set of ratings.

### Training Effects on Competence

*Supervisors' ratings.* Using the supervisors' ratings,  $p$  values for the results of the HLM analysis for each of the three treatment modalities are presented in Table 2. Mean values for each successive training case for each modality are presented in Table 3. For IDC, a significant linear effect was found for the first-level analysis for competence ( $p = .03$ ), indicating that therapists performed better as each case progressed (i.e., from early to later sessions). In terms of magnitude, this linear effect

Table 2  
Hierarchical Linear Modeling Results ( $p$  Values) for Training of Cognitive and Supportive-Expressive Therapists and Individual Drug Counselors

Effect	$p$			
	CT	IDC (comp.)	SE	
			Adh.	Comp.
Level 1				
Linear slope over sessions within patients	.10	.03	.004	.79
Level 2				
Linear slope over training cases using Level 1 intercepts as dependent variable	.001	.49	.32	.64
Linear slope over training cases using Level 1 slopes as dependent variable	.005	.29	.17	.05

Note. CT = cognitive therapy; SE = supportive-expressive dynamic therapy; IDC = individual drug counseling; Adh. = adherence; Comp. = competence.

Table 3  
 Mean Competence and Alliance Ratings and Standard Deviations  
 Across Training Cases for CT, IDC, and SE

Rating	Training cases							
	First		Second		Third		Fourth	
	M	SD	M	SD	M	SD	M	SD
<b>CT</b>								
Competence	36.9	11.3	44.6	10.7	45.3	10.6	46.2	9.6
HAQ	5.2	0.4	5.2	0.5	5.3	0.4	5.3	0.4
CALPAS total	6.1	0.6	5.9	0.7	6.0	0.5	5.9	0.5
<b>IDC</b>								
Competence	3.6	1.2	3.7	1.1	3.7	1.1	3.3	0.8
HAQ	5.3	0.3	5.2	0.7	5.1	0.7	5.3	0.5
CALPAS total	6.0	0.5	5.8	0.7	5.7	0.7	6.0	0.6
<b>SE</b>								
Adherence	2.8	0.8	2.8	0.9	2.9	0.8	2.7	0.8
Competence	3.9	0.8	3.7	0.9	3.7	0.8	3.5	1.0
HAQ	5.2	0.6	5.2	0.5	5.3	0.5	5.2	0.6
CALPAS total	6.0	0.7	5.9	0.7	6.0	0.6	5.8	0.6

Note. Adherence and competence scores for SE and IDC are averages of items rated on a scale ranging from 1 to 7. Competence scores for CT are totals on a 7-item Cognitive Therapy Scale (CTS) prorated to compare with the standard 11-item CTS. HAQ scores are averages of items rated on a scale ranging from 1 to 6. CALPAS totals are averages of items rated on a scale ranging from 1 to 7. CT = cognitive therapy; IDC = individual drug counseling; SE = supportive-expressive dynamic therapy; HAQ = Helping Alliance Questionnaire—Revised; CALPAS = California Psychotherapy Alliance Scale.

over sessions was estimated to be a change of 0.18 units (on the scale of 1 to 7) on competence ratings for every 10 sessions. No significant effect was evident in IDC for a trend across training cases, using either the Level 1 intercepts or slopes as dependent variables.

The results for CT showed a nonsignificant effect ( $p = .10$ ) for learning over sessions but a highly significant ( $p < .001$ ) linear effect of Level 1 intercepts over training cases. Moreover, the lack of an overall effect over sessions is clarified by highly significant ( $p < .005$ ) linear change in Level 1 slopes over training cases. This effect was in the expected direction (i.e., early training cases evidenced a linear increase in competence over sessions), but this effect is reduced in later training cases. The size of this effect can be expressed in terms of units of improvement in competence for each new training case. On the CTS total score, the rate of change over training cases was estimated to be 4.5 units of improvement in competence scores for each new patient seen.

As can be seen in Table 2, a significant linear effect over sessions (Level 1) was found for frequency of use of expressive techniques in SE. In terms of amount of change, this effect was small, with the rate of change estimated to be 0.08 adherence scale units (on a scale ranging from 1 to 7) for every 10 sessions. No effect for changes in Level 1 intercepts over training cases was apparent for SE. There was, however, a significant ( $p = .05$ ) effect for Level 1 slopes changing over training cases for SE competence. An examination of the slopes over sessions for each individual training case revealed a small linear improvement for the first training case but relatively flat lines (no improvement over sessions) for subsequent training cases.

*Independent judges.* Data from the independent raters who were unaware of case progress were also analyzed by means of

HLM, but because there were not enough data points within each training case (only two sessions per patient), the analysis consisted of patients at Level 1 (scores for the two sessions per patient were averaged) and therapists at Level 2. Results from these analyses confirmed the findings from the original supervisor. For CT, a significant ( $p < .02$ ) linear improvement was found across training cases, but no effects for SE or IDC (all  $ps > .25$ ) were apparent. The estimate rate of change on the CTS, on the basis of the slope coefficient, was 3.9 units of improvement per training case, slightly less than the rate of change seen with the data based on supervisors' ratings of the cases they had supervised.

#### Training Effects on Alliance

The total score on the HAQ and the four subscales and the total score of the CALPAS were examined for training effects within and across training cases by using HLM. Mean HAQ and CALPAS total scores over training cases are given in Table 3. As can be seen, patient alliance ratings (averaging across sessions) were very positive, with HAQ average scores about 5 on a scale ranging from 1 to 6 and with CALPAS scores averaging about 6 on a scale ranging from 1 to 7. No statistically significant linear changes within or across training cases were detected for either IDC or SE on any of the alliance measures. The pattern of mean scores for successive training cases for IDC, however, suggested the possibility of a nonlinear effect. Using the CALPAS Working Strategy subscale and entering a quadratic term into the Level 2 model (across training cases), we found a statistically significant ( $p < .05$ ) curvilinear effect, with scores decreasing for the second and third IDC training cases but increasing back to initial levels for the fourth training

Table 4  
*Mean Alliance Scores and Standard Deviations in Cognitive Therapy  
 for Sessions 2, 5, and 24*

Session	CALPAS							
	HAQ		Working strategy		Therapist understanding		Total	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
2	5.1	0.5	6.0	0.8	6.2	0.6	5.9	0.6
5	5.2	0.5	6.0	0.9	6.2	0.7	6.0	0.7
24	5.5	0.4	6.3	0.7	6.6	0.4	6.1	0.6

*Note.* CALPAS = California Psychotherapy Alliance Scale; HAQ = Helping Alliance Questionnaire—Revised.

case. The means (and standard deviations) for the four training cases on the Working Strategy subscale were 6.2 (0.6), 5.8 (0.9), 5.6 (0.8), and 6.1 (0.7). No significant quadratic effects were found for the other CALPAS subscales, CALPAS total score, or HAQ total score.

Although no across-case training effect was evident for CT, a significant within-case (over sessions) training effect for CT was found for the HAQ ( $p < .0001$ ), CALPAS Working Strategy subscale ( $p < .01$ ), CALPAS Therapist Understanding ( $p < .0005$ ), and CALPAS total ( $p < .01$ ) scores. Mean scores for each session on these measures are given in Table 4. As can be seen, the effect was largely a function of patients in CT rating the alliance more positive at Session 24.

### Discussion

In this study of training effects for manual-guided CT, SE, and IDC, only CT demonstrated learning that carried over from training case to training case. The learning demonstrated within IDC and SE occurred only within training cases and these effects were small. Thus, it appears as though SE and IDC therapists were able to respond somewhat to supervision as a case progressed, but they were not able to transfer this learning to new cases.

The learning effect for CT found here has important potential implications for the training of therapists, not only within research studies but also in graduate training programs. Increasingly, graduate training programs have gravitated to the teaching of empirically supported treatment modalities that typically involve treatment manuals (Crits-Christoph, Frank, Chambless, Brody, & Karp, 1995). To the extent that CT for cocaine dependence turns out to be an efficacious treatment, the teachability of the therapy provides further justification for the dissemination of the approach to clinical training programs. Moreover, the fact that the learning effect over cases for CT was evident not only in supervisors' ratings but also from ratings by independent judges not familiar with each case lends confidence to our finding and suggests that supervisors' ratings can be used to assess ongoing learning over cases. This is especially important because in actual training programs it is likely that clinical supervisors typically provide the ongoing assessments of the extent to which a trainee is learning.

At the same time, sources of bias in supervisors' ratings should not be ignored. In our investigation, supervisors rated their own trainees higher than independent judges did, suggesting some bias in their ratings. Bias in supervisors' ratings has been found in previous research, such as Dodenhoff's (1981) finding that trainees who were interpersonally attracted to their supervisors were rated as more effective by their supervisors. Further work on competence scales is clearly needed so that reliability is enhanced and bias minimized.

There may be several reasons for the lack of learning effects across training cases for SE and IDC. First, it may be that these therapists were generally good to begin with and had little room for improvement. Thus, the results for SE and IDC may be more in accordance with Rounsaville's et al.'s (1986) results in training experienced therapists to deliver IPT in the NIMH collaborative study than with Henry, Strupp, Butler, Schacht, and Binder's (1993) results. Rounsaville et al. (1986) suggested that their results were due to a ceiling effect because they carefully selected experienced therapists from the outset and eliminated therapists who did not meet stringent acceptance criteria. In contrast, in the Henry, Strupp, et al. (1993) study, therapists were not selected (but were assessed) on the basis of their pretraining treatment skills.

In the present study, we selected therapists and counselors who were familiar with a particular therapy modality, and training focused on specific interventions for the population and refinement in technique and delivery. The fact that, for SE, learning occurred over sessions for the first case but not for subsequent cases is possibly consistent with a ceiling effect where SE therapists had little room for improvement after the first case. It is also possible that the ceiling effect is imposed by the therapists themselves. A therapist might implement just enough techniques to obtain positive progress with a patient and then back off from the manual-based techniques, resorting perhaps to supportive interventions, as sessions continue with a patient who is improved.

In absolute terms, however, the SE and IDC therapists were rated during the training phase as performing techniques only at an acceptable level (average item rating of 3.7 for SE, 3.6 for IDC) on a scale ranging from 1 to 7. In addition, the SE supervisors did not consider the group of SE trainees to be

outstanding on average, although a couple of the SE therapists performed at a higher level of competence than the others. Moreover, the SE supervisors considered SE to be difficult to implement skillfully. In particular, these supervisors believed it is difficult to master the ability to formulate the patient's central relationship pattern on line during a therapy session and tactfully make inquiries and interpret this clinical material in the context of patient anxiety. A common fear among SE trainees was that interpretations would interfere with a positive therapeutic alliance. Despite these difficulties in training SE and IDC, most therapists ended the training phase at an acceptable level of competence to continue into the main clinical trial.

Another possible explanation of the lack of training effects for SE and IDC is that these approaches might not be conducive to the use of treatment manuals. It may be, particularly with SE, that treatment manuals somehow constrain the therapist or hinder their abilities to be spontaneous, thereby compromising the quality of treatment. Along these lines, the study by Henry, Strupp, et al. (1993) is often cited in connection with the potential pitfalls of manual-guided psychotherapies. In the Henry, Strupp, et al. (1993) study, a nonsignificant trend for therapists to have more hostile messages after training in brief dynamic therapy was detected, as well as a significant increase in therapist complex thought units (possibly indicating implied criticism) after training.

In the present study, however, there was no evidence that training using treatment manuals compromised the therapeutic alliance for SE or CT. For IDC, an initial negative impact on the alliance was detected in an exploratory analysis of curvilinear effects over training cases. Alliance scores, however, rebounded to initial levels by the fourth training case. Moreover, alliance scores throughout training were at the upper end of the scale, suggesting that patients viewed their therapeutic relationships very positively. These data suggest that the negative impact of manual-guided training reported by Henry, Strupp, et al. (1993) may not be generalizable to other treatment manuals. In addition, our findings regarding IDC indicate that such negative impacts, when they occur, may be transitory. If this study used only one training case, as had been done by Henry, Strupp, et al. (1993), then the transitory nature of the negative impact on the alliance in IDC would not have been evident.

We do concur with Henry, Schacht, et al. (1993) that, at least in regard to dynamic therapy, new models of training need to be considered. The existing training methods and treatment manual for cognitive therapy, however, appeared to work successfully. It may be that different forms of manuals and training procedures will be needed for therapies that are quite different in their assumptions about human nature, techniques, and degree of reliance on intuition and inference. Training methods, however, may need to vary not only depending on the type of treatment but also depending on aspects of the particular trainee and even a given patient's current clinical state. For example, Holloway and Neufeldt (1995), on the basis of previous research (e.g., Tracey, Ellickson, & Sherry, 1989), indicate that greater structure in supervision is important for less experienced trainees, but this relationship is moderated by the reactance potential of the trainee as well as the extent to which a client is currently in crisis. Without guidelines, different supervisors may handle these issues differently. Indeed, Henry, Schacht, et al. (1993)

reported that the degree to which therapists learned the treatment manual was influenced by different supervisory styles. To deal with these sorts of issues, Holloway and Neufeldt (1995) have suggested that there needs to be more attention to the training of supervisors, and at least one manual for training supervisors has appeared (Neufeldt, Iversen, & Juntunen, 1995).

At this point in time, it is not possible to unravel whether our lack of an effect over training cases in SE and IDC is due to inadequate training procedures, lack of room for improvement on the scales, poor sensitivity of the scales to change, poor trainers, difficulty in teaching or learning SE and IDC in particular, or sampling error with a small sample of therapists and counselors. Because competence in expressive techniques in SE therapy has been found to predict treatment outcome with depressed patients (Barber, Crits-Christoph, & Luborsky, 1996), it appears important to continue to attempt to maximize such competence through either selection or training procedures.

Despite the overall training effects found for CT, considerable individual therapist variability in improvement was found. Such variability suggests that it may be fruitful to examine potential predictors (e.g., demographic and other background characteristics) of improvement with an eye toward modifying selection criteria or concentrating training efforts with certain trainees. A further report will explore these potential mediators of training effects.

## References

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- Barber, J. P., & Crits-Christoph, P. (1996). Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report. *Psychotherapy Research, 6*, 81-94.
- Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Therapist competence and treatment outcome in dynamic therapy. *Journal of Consulting and Clinical Psychology, 64*, 619-622.
- Barber, J. P., Mercer, D., Krakauer, I., & Calvo, N. (1996). Development of an adherence/competence rating scale for individual drug counseling. *Drug and Alcohol Dependence, 43*, 125-132.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.
- Beck, A. T., Wright, F. D., Newman, C. F., & Liese, B. S. (1993). *Cognitive therapy of substance abuse*. New York: Guilford Press.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Crits-Christoph, P., Frank, E., Chambless, D. L., Brody, C., & Karp, J. (1995). Training in empirically-validated treatments: What are clinical psychology students learning? *Professional Psychology, 26*, 514-522.
- Crits-Christoph, P., Siqueland, L., Blaine, J., Frank, A., Luborsky, L., Onken, L. S., Muenz, L., Thase, M. E., Weiss, R. D., Gastfriend, D., Woody, G., Barber, J. P., Butler, S., Daley, D., Salloum, I. M., Bishop, S., Griffin, M. L., Lis, J., Mercer, D., Najavits, L. M., Moras, K., & Beck, A. T. (1997). The NIDA Collaborative Cocaine Treatment Study: Rationale and methods. *Archives of General Psychiatry, 54*, 721-726.
- Daldrup, R. J., Beutler, L. E., Engle, D., & Greenberg, L. (1988). *Focused expressive psychotherapy*. New York: Guilford Press.
- Dodenhoff, J. T. (1981). Interpersonal attraction and direct-indirect supervisor influence as predictors of counselor trainee effectiveness. *Journal of Counseling Psychology, 28*, 47-52.



- Elkin, I., Parloff, M. B., Hadley, S. W., & Autrey, J. H. (1985). National Institute of Mental Health Treatment of Depression Collaborative Research Program: Background and research plan. *Archives of General Psychiatry*, 42, 305-316.
- Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S., & Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971-982.
- Foley, S. H., O'Malley, S., Rounsaville, B., Prusoff, B. A., & Weissman, M. M. (1987). The relationship of patient difficulty to therapist performance in interpersonal psychotherapy of depression. *Journal of Affective Disorders*, 12, 207-217.
- Frank, E., Kupfer, D. J., Perel, J. M., Cornes, C., Jarrett, D. B., Malinger, A. G., Thase, M. E., McEachran, A. B., & Grochocinski, V. J. (1990). Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry*, 47, 1093-1099.
- Gaston, L. (1991). Reliability and criterion-related validity of the California Psychotherapy Alliance Scales—Patient Version. *Psychological Assessment*, 3, 68-74.
- 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' response 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.
- Holloway, E. L., & Neufeldt, S. A. (1995). Supervision: Its contributions to treatment efficacy. *Journal of Consulting and Clinical Psychology*, 63, 207-213.
- Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment*. New York: Basic Books.
- Luborsky, L., Barber, J. P., Siqueland, L., Johnson, S., Najavits, L. M., Frank, A., & Daley, D. (1996). The revised Helping Alliance Questionnaire (HAQ-II): Psychometric properties. *Journal of Psychotherapy Practice and Research*, 6, 260-271.
- Luborsky, L., Woody, G. E., McLellan, A. T., O'Brien, C. P., & Rosenzweig, J. (1982). Can independent judges recognize different psychotherapies? An experience with manual-guided therapies. *Journal of Consulting and Clinical Psychology*, 50, 49-62.
- Mark, D., & Faude, J. (1995). Supportive-expressive therapy for cocaine abuse. In J. P. Barber & P. Crits-Christoph (Eds.), *Dynamic therapies for psychiatric disorders, Axis I* (pp. 294-331). New York: Basic Books.
- Mark, D., & Luborsky, L. (1992). *A manual for the use of supportive-expressive psychotherapy in the treatment of cocaine abuse*. Unpublished manuscript, Department of Psychiatry, University of Pennsylvania.
- Mercer, D., Calvo, N., Krakauer, I. D., & Barber, J. P. (1994). *Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence*. Unpublished measure, Center for Psychotherapy Research, Department of Psychiatry, University of Pennsylvania Medical School.
- Mercer, D., & Woody, G. (1992). *Addiction counseling*. Unpublished manuscript, Center for Studies of Addiction, University of Pennsylvania/Philadelphia Veterans Affairs Medical Center.
- Neufeldt, S. A., Iversen, J. N., & Juntunen, C. L. (1995). *Supervision strategies for the first practicum*. Alexandria, VA: American Counseling Association.
- Rounsaville, B. J., Chevron, E. S., Weissman, M. M., Prusoff, B. A., & Frank, E. (1986). Training therapists to perform interpersonal psychotherapy in clinical trials. *Comprehensive Psychiatry*, 27, 364-371.
- Shaw, B. (1984). Specification of the training and evaluation of cognitive therapists for outcome studies. In J. B. W. Williams & R. L. Spitzer (Eds.), *Psychotherapy research: Where are we and where should we go?* (pp. 173-188). New York: Guilford Press.
- Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured Clinical Interview for DSM-III-R—Patient Edition (with psychotic screen): Personality disorders (SCID-II, Version 1.0)*. Washington, DC: American Psychiatric Press.
- Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology*, 63, 182-196.
- Tracey, T. J., Ellickson, J. L., & Sherry, P. (1989). Reactance in relation to different supervisory environments and counselor development. *Journal of Counseling Psychology*, 36, 336-344.
- Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 54, 381-385.
- Young, J., & Beck, A. T. (1980). *The Cognitive Therapy Scale*. Unpublished manuscript, Center for Cognitive Therapy, Philadelphia.

Received October 28, 1996  
Accepted November 18, 1997 ■