The National Institute on Drug Abuse Collaborative Cocaine Treatment Study was designed to assess the efficacy of four different psychosocial interventions (cognitive therapy, supportive–expressive dynamic therapy, and individual and group drug counseling) for cocaine dependence. This report addresses the treatment integrity and discriminability of the three individual treatments. Therapists’ adherence and competence for all three
Randomized clinical trials that test the efficacy of different medications do not need to be concerned either with the amount of active ingredient found in each pill given to the patient or with the possibility that nonintended active ingredients have been included in the pills. Pharmaceutical companies have already taken care of ensuring such aspects of internal validity for these clinical trials. Of course, mistakes could occur (e.g., the wrong medication may be given to a patient), but for the most part, these potential errors are not of great concern. In contrast, when studying the efficacy of psychotherapeutic interventions in a randomized clinical trial, there is concern whether the treatment was indeed delivered and whether it was delivered adequately and as intended (treatment integrity) and whether techniques from other treatments were included (treatment discriminability). Thus, the assessment of treatment integrity is of central importance to the understanding and interpretation of psychosocial clinical trial results (Waltz, Addis, Koerner, & Jacobson, 1993). If the psychotherapy treatments were not delivered adequately, this might be an important explanation of poor outcomes. Alternatively, if the treatment delivered was not the intended treatment, then conclusions from the clinical trial would be misleading. Moreover, if treatments intended to be different in a study were not discriminable, the study hypotheses would not have been adequately tested.

To ensure the internal validity of psychotherapeutic interventions, state-of-the-art randomized clinical trials comparing the efficacy of these different treatments have employed elaborate technologies (e.g., Elkin et al., 1989). For example, treatment manuals, which describe and specify the treatment to be delivered in detail, have arguably become a minimal requirement for a clinical trial to be deemed credible. In addition, adherence scales have been developed to use in conjunction with treatment manuals to allow researchers to assess whether therapists followed the treatment techniques specified in the manual. Adherence scales can be used not only to demonstrate that different treatments are distinguishable from one another (treatment differentiability; Dobson & Shaw, 1988; Kazdin, 1986) but also to assess the extent to which practitioners incorporate techniques from a variety of treatments.

The National Institute on Drug Abuse (NIDA) Collaborative Cocaine Treatment Study (CCTS; Crits-Christoph et al., 1997, 1999, 2001) is a multicenter project investigating the efficacy of four treatments: cognitive therapy (CT; Beck, Wright, Newman, & Liese, 1993) plus group drug counseling (GDC; Mercer, Carpenter, Daley, Patterson, & Volpicecci, 1994), supportive–expressive (SE; Mark & Luborsky, 1992; Mark & Faude, 1995) psychotherapy plus GDC, individual drug counseling (IDC; Mercer & Woody, 1992) plus GDC, and GDC alone. Crits-Christoph et al. (1999) found that all four treatments decreased drug use substantially, with cocaine use in the past 30 days improving from a mean of 10.4 days ($SD = 7.8$; $Mdn = 8.0$) at intake to a mean of 3.4 days ($SD = 6.5$; $Mdn = 0$).
12 months after randomization. They also reported that IDC + GDC produced statistically and clinically superior outcomes compared with the other treatments. At the 6-month assessment, for example, only 39% of patients in IDC + GDC had used cocaine in the past month, compared with 57% of patients in CT + GDC, 49% in SE + GDC, and 52% in GDC alone. The CCTS investigators specified the treatment variables (therapeutic techniques) of each treatment modality included in their study via the development and use of detailed treatment manuals. Similar procedures had been used in previous large-scale comparative trials, such as the National Institute of Mental Health (NIMH) Treatment for Depression Collaborative Research Program (TDCRP; Elkin et al., 1989) and Project Matching Alcoholism Treatments to Client Heterogeneity (MATCH; Project MATCH Research Group, 1993; Carroll et al., 1998). Hill, O’Grady, and Elkin (1992) showed that the treatment used in the TDCRP could be easily discriminated. Similarly, Carroll et al. (1998) were able to discriminate among the treatments used in Project MATCH.

Because adherence to specific techniques does not necessarily ensure that the treatment was delivered appropriately, the NIDA CCTS implemented therapist competence scales to allow for the assessment of how skillfully the techniques were implemented (Barber, Liese, & Abrams, 2003; Barber, Krakauer, Calvo, Badgio, & Faude, 1997; Barber, Mercer, Krakauer, & Calvo, 1996). The NIDA CCTS represents the first major clinical trial in which, in addition to adherence ratings, competence scales rated by independent judges have been implemented for all of the individual treatment conditions. Because experts in the treatment delivered are needed to rate competence (e.g., Barber, Crits-Christoph, & Luborsky, 1996; O’Malley et al., 1988; Waltz et al., 1993), we required a specific team of judges for each treatment condition used.

This report has two aims: (a) to assess the interjudge reliability and internal consistency of the adherence and competence measures used in the NIDA CCTS and (b) to determine whether the treatments could be discriminated using these scales.

**Method**

**Patients**

Participants in the NIDA CCTS were 487 cocaine-dependent outpatients. Of these, 364 were randomized to one of the three individual treatments at five research sites: 92 at the University of Pittsburgh (PA) Western Psychiatric Institute and Clinic, 89 at the University of Pennsylvania (Philadelphia, PA), 93 at Brookside Hospital (Nashua, NH), 54 at Massachusetts General Hospital (Boston, MA), and 36 at McLean Hospital (Belmont, MA). The participants were recruited from the following sources: 46% by newspaper or flyer, 22% from other substance abuse treatment centers, 18% by a friend or acquaintance, 8% from other mental health centers, and 6% from private mental health providers. Patients between the ages of 18 and 60 were included in the study if they met Structured Clinical Interview for DSM–IV (First, Spitzer, Gibbon, & Williams, 1994) criteria for a principal diagnosis of cocaine dependence (current or in early partial remission), had used cocaine in the past 30 days, and had a stable living situation, with plans to live in the area for the next 2 years. Patients were excluded if they had a principal diagnosis of alcohol dependence or alcohol abuse, opioid dependence (current or in early partial remission), or polysubstance dependence/abuse; dementia or other irreversible organic brain syndrome; psychotic symptoms; a history of bipolar I disorder; a current imminent suicide or homicide risk; a need to be maintained on a psychotropic medication; or a life-threatening or unstable medical illness. Patients were also excluded if they were unwilling
to discontinue a current psychotherapeutic treatment, had been hospitalized for more than 10 days in the past 30 days, were legally mandated to attend treatment, or were more than 12 weeks pregnant.

Patients were screened, usually by telephone, and, if appropriate, invited for an intake visit. Following the intake visit and signing of informed consent, the patients began an orientation phase that included both attendance and assessment requirements. They were required to attend three clinic visits within 14 days, including one group session and two case management visits, before being randomized to treatment. In the orientation phase, group counselors suggested attendance at self-help groups such as Cocaine Anonymous or Alcoholics Anonymous, promoted HIV risk reduction, and addressed housing, job, and financial needs. Patients meeting the attendance requirements then moved on to the 1–2 day postorientation assessments. The purpose of the orientation phase was to select for treatment only those patients with enough motivation to attend at least a few sessions.

Following satisfactory completion of the postorientation assessments, patients were randomized to treatment using a computerized “urn” randomization procedure, with sex, marital status, employment status, mode of cocaine use, psychiatric severity, and antisocial personality trait score included in order to balance the treatment conditions on these potential prognostic factors (see Crits-Christoph et al., 1997, 1999). The urn randomization was done separately for each site to ensure that treatment conditions were balanced on the relevant factors within each location.

Treatments

The three individual therapies were CT (Beck et al., 1993), based on Beck’s cognitive model (Beck, 1976); IDC (Mercer & Woody, 1992), based on the 12-step disease model of addiction; and SE dynamic therapy (Mark & Faude, 1995; Mark & Luborsky, 1992), a time-limited psychodynamic treatment based on Luborsky’s (1984) model. In addition to receiving their individual treatment, all patients also received GDC (Mercer et al., 1994), a psychoeducational and problem-solving group treatment that is grounded in the 12-step addiction model as well. A fourth treatment condition consisting of GDC only was also part of the study; however, it was not included in the present report since it was not an individual treatment and patients did not always have the same counselor. The active phase of the individual treatments lasted 6 months and consisted of twice-weekly sessions for the first 3 months of treatment and weekly sessions throughout months 4–6. Three monthly booster sessions were offered to patients who had stayed in treatment for the 6 months of the active phase. Group drug counseling sessions were held once a week for the full 6 months of individual treatment.

Therapists and Their Training

Therapists were selected by their respective training units on the basis of their background information, education, clinical training, letters of reference, and two audiotaped samples of their therapy work (see Crits-Christoph et al., 1998). Therapists went through an extended training period under the leadership of a coordinating center located at the University of Pennsylvania (Crits-Christoph et al., 1998). The training included an initial 2-day workshop during which the background and rationale for the study were presented. In addition, the specific treatment manuals and the difficulties implementing these manuals in that patient population were reviewed. During the remainder of the training phase, three additional workshops were conducted. Videotaped and audiotaped sessions from
actual therapies were presented for learning purposes and were reviewed extensively.
During the training phase of the study, clinicians were required to see four patients for at
least 1 month. These sessions were audiotaped and sent to the coordinating center for
supervision. Supervisors contacted their supervisees within a few days of receiving the
tapes and provided them with phone supervision for all of their training cases. Crits-
Christoph et al. (1998) reported that the average patient in the training phase had 5.3
\(SD = 1.7\) sessions rated by the supervisor. During the clinical phase of the CCST, which
is the focus of the present report, therapists continued to receive supervision at a somewhat lower rate.

Clinicians who demonstrated their ability in delivering treatment as intended and in a
competent manner were certified to participate in the trial. Fifteen cognitive therapists
\(M = 13.4\) years of experience), 13 SE therapists \(M = 11.7\) years), and 12 IDC counselors
\(M = 13.9\) years) were enrolled in the clinical trial.

Measures

Cognitive Therapy Adherence-Competence Scale (CTACS; Barber, Liese, & Abrams,
2003; Liese, Barber, & Beck, 1995). The 21-item CTACS is a new scale developed for this project that was derived from the Cognitive Therapy Scale (Young & Beck, 1980). It addresses the main interventions in Beck et al.’s (1993) manual for CT for cocaine dependence. Each item was rated separately for adherence to the manual, appropriateness of the frequency of the intervention (or lack of intervention), and quality of the intervention on a Likert-type scale ranging from 0 (low) to 6 (high). If a certain intervention did not occur (corresponding to an adherence of 0), judges rated the appropriateness of the therapist’s decision not to carry out that specific intervention. In those cases, the quality of the intervention was equated with the appropriateness of the non-intervention. Because of the high intercorrelations among these scale scores, an overall competence score was computed as the mean of the appropriateness and quality scores. Preliminary reliability data were based on a sample of patients who had received CT \((n = 92)\), SE \((n = 20)\), or IDC \((n = 22)\); these data were found to be acceptable (Barber et al., in press). More specifically, the intraclass correlations (ICCs[2, 2]) (Shrout & Fleiss, 1979, p. 426) for the 21-item mean CTACS were .80 and .80 for adherence and competence, respectively.

Adherence/Competence Scale for IDC for Cocaine Dependence (ACS-IDCCD; Barber, Mercer, et al., 1996). This 43-item scale covers the main components of the IDC manual. Adherence and competence were rated separately on a scale from 1 (low) to 7 (high). Preliminary reliability data based on ratings from sessions of the training/pilot phase of the NIDA CCTS by expert judges of IDC \((n = 41)\), SE \((n = 10)\), and CT \((n = 11)\) indicated excellent reliability (ICCs[2, 2] of .90 and .85 for adherence and competence, respectively).

Adherence/Competence Scale for SE for Cocaine Dependence (ACS-SEC; Barber et al., 1997). The ACS-SEC consists of three major subscales: Supportive (13 items), Expressive (31 items), and Cocaine Abuse (11 items). The Supportive subscale contains items that focus on therapist interventions intended to build and maintain the therapeutic alliance. The Expressive subscale covers the interpretative and clarifying techniques used in SE. The Cocaine Abuse subscale contains items pertaining to the therapist’s interventions that address drug abuse from a dynamic and interpersonal perspective. Each item is rated on a 1 (low) to 7 (high) Likert-type scale, separately for adherence (frequency of the intervention), appropriateness (of the frequency), and quality (skillful deliv-
ery of the intended intervention). As with the CTACS, we averaged the appropriateness and quality scores into a competence score because of their high intercorrelations. On the basis of SE therapy expert ratings of 32 SE sessions, 10 CT sessions, and 10 IDC sessions from the pilot phase of the NIDA CCST, acceptable to good reliability was found for the different adherence subscales (ICCs[2, 2] of .70 to .90), but weaker reliability was found for competence (ICCs[2, 2] of .32 to .65). The reliabilities of the adherence and competence average scale scores were .86 and .50 in this pilot phase sample.

**Expert Raters and Audiotapes**

Experts in the specific treatment modalities were needed to evaluate whether treatment was delivered in a competent manner. The experts used as judges for each treatment condition (two for each modality) were not otherwise associated with the NIDA CCTS. Each set of expert judges rated audiotaped sessions from their designated treatment condition as well as from the two alternative treatment conditions using the adherence/competence rating scale corresponding to their area of expertise. For example, the CT experts rated a smaller number of randomly selected sessions from SE and IDC using the CT scale in addition to rating a larger number of CT sessions.

Because of the large number of patients included in the study and the use of expert judges from three different treatment modalities, only a portion of the therapy sessions completed were selected for rating. Each judge rated two audiotapes from every patient—therapist/counselor dyad from the treatment condition of their expertise. The two tapes that were rated were selected in the following manner: One tape was randomly drawn from Sessions 2 through 11, while the second tape was randomly drawn from Session 12 to termination. In addition, judges rated one to two treatment sessions of each therapist/counselor from the treatment conditions in which he or she lacked expertise. Because a large treatment effect size was expected, only a relatively small number of tapes of sessions selected randomly from the two alternative treatment conditions were included.

IDC raters evaluated 183 randomly selected audiotapes of IDC sessions from 99 patients seen by one of 12 counselors who participated in the clinical phase of the NIDA CCS. In addition, IDC raters judged 54 randomly chosen tapes from 37 patients seen by one of 14 SE therapists, and 71 randomly chosen tapes from 48 patients seen by one of 16 CT therapists. Ninety percent of rated sessions came from between Sessions 2 and 17, with the remainder from Sessions 18 to 29 (Mdn = 5).

SE raters evaluated 214 audiotapes of SE sessions randomly chosen from 112 patients seen by the 16 SE therapists. SE raters also rated 53 CT sessions involving 32 patients and 16 therapists, and 35 IDC sessions involving 22 patients and 12 therapists. Ninety percent of rated sessions were between Sessions 2 and 20, with the remainder from Sessions 21 to 35 (Mdn = 6).

CT raters evaluated 196 randomly selected audiotapes of CT sessions involving 105 patients and 16 therapists, 42 tapes of SE sessions from 25 patients seen by one of 13 therapists, and 35 tapes of IDC sessions from 21 patients seen by one of 11 therapists. Ninety percent of rated sessions were between Sessions 2 and 19, with the remainder from Sessions 20 to 30 (Mdn = 6).

**Results**

**Interjudge Reliability**

An assessment of variance components was conducted using the mixed-models procedure in SAS (SAS/STAT Software, 1996) to partition the variance of the judges’ ratings.
For each scale or subscale, variance components were estimated using a restricted maximum-likelihood estimation method for the following sources of variance: site, treatment (CT, IDC, or SE), Site × Treatment, therapist (nested within treatment and site), patient (nested within therapist, treatment, and site), session (nested within patient, therapist, treatment, and site), rater, and error. Each term was treated as a random effect in these analyses. The term treatment captures the effect of specific treatments on the adherence/competence ratings derived from each modality-specific measure; the term rater captures systematic differences that are attributable to an individual judge’s tendency to score high (or low) on a given scale. These analyses were conducted separately for each set of expert judges (i.e., separately for the CTACS, ACS-IDCCD, and ACS-SEC). The estimates of variance were transformed into proportions of variance on the basis of estimates of the total variance. As expected, the majority of the variance in the judges’ ratings was accounted for by treatment modality for each of the modality-specific measures (ranging from .09 to .73, Mdn = .40; see Table 1). This indicates that ratings of adherence and competence varied as a function of the type of treatment being rated by all three sets of expert judges. Table 1 also shows that for each of the modality-specific scales and subscales, therapist accounted for approximately 5% of the variance for each rating system, reflecting that all three sets of expert judges determined that some therapists applied therapeutic techniques more frequently and more competently than other therapists. In addition, a substantial amount of variance (roughly 18%) was explained by session, particularly for the SE-specific scales. Hence, not only the frequency but also the competent use of therapeutic techniques fluctuated across sessions according to all judges, especially the SE judges. Finally, patient accounted for a notable amount of variance, but this was specific to SE measures of adherence, indicating that SE judges, but not CT or IDC judges, believed that SE techniques were used more frequently with some patients and less frequently with other patients.

The variance estimates generated from the analytic model described above were used to calculate intrarater reliability coefficients, with rater specified as a random effect. Table 1 presents the intrarater reliability estimates for the two judges’ mean score (i.e., ICC [2, 2]; Shrout & Fleiss, 1979, p. 426). More specifically, these estimates reflect the reliability of the mean rating from two expert judges and their generalizability to other random samples of expert judges. Finally, Cronbach’s alpha coefficient was computed for each subscale to assess internal consistency. All ICCs were acceptable (ranging from .54 to .94, Mdn = .85), except the SE Competence Cocaine Abuse subscale (.54). All scales had acceptable internal consistency (ranging from .76 to .98, Mdn = .93).

**Criterion Validity**

We also examined criterion validity, or the ability of a measure to distinguish between different therapeutic modality. Adherence and competence ratings were compared for each modality-specific measure to examine whether therapists adhered more to their own treatments than did the therapists from the alternative modalities. Following the overall F test (all three omnibus F tests were significant at p < .0001), the adjusted means from the mixed-model analysis were compared using pairwise comparisons with a Bonferroni-adjusted alpha of .0014 (see Table 2). As expected, adherence and competence ratings discriminated between the treatments. This is consistent with the results in Table 1, where the treatment effect explained most of the variance of the ratings. More specifically, expert IDC judges rated IDC sessions higher than CT or SE sessions (and rated CT sessions higher than SE sessions; see Table 2) on both adherence and competence scales.
Table 1

*Proportion of Variance Explained by Model Terms, Interjudge Reliability, and Internal Consistency for All Treatments*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Model term</th>
<th>ICC (two judges’ mean score)</th>
<th>Reliability (Cronbach’s alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDC Scale (n = 307)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence mean scale score</td>
<td>.00   .47   .00   .08   .05   .19   .01   .20</td>
<td>.88</td>
<td>.86</td>
</tr>
<tr>
<td>Competence mean scale score</td>
<td>.00   .38   .00   .07   .06   .09   .22   .19</td>
<td>.74</td>
<td>.93</td>
</tr>
<tr>
<td>CT Scale (n = 273)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence mean scale score</td>
<td>.02   .66   .00   .05   .03   .08   .10   .06</td>
<td>.91</td>
<td>.94</td>
</tr>
<tr>
<td>Competence mean scale score</td>
<td>.01   .73   .00   .05   .04   .05   .00   .12</td>
<td>.94</td>
<td>.96</td>
</tr>
<tr>
<td>SE Scale (n = 302)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence supportive subscale</td>
<td>.01   .37   .00   .08   .02   .19   .02   .31</td>
<td>.80</td>
<td>.88</td>
</tr>
<tr>
<td>Competence supportive subscale</td>
<td>.01   .48   .00   .06   .01   .08   .00   .35</td>
<td>.79</td>
<td>.95</td>
</tr>
<tr>
<td>Adherence expressive subscale</td>
<td>.01   .39   .00   .05   .09   .23   .02   .20</td>
<td>.88</td>
<td>.92</td>
</tr>
<tr>
<td>Competence expressive subscale</td>
<td>.01   .41   .00   .04   .02   .21   .04   .27</td>
<td>.82</td>
<td>.96</td>
</tr>
<tr>
<td>Adherence cocaine abuse subscale</td>
<td>.00   .11   .00   .04   .20   .47   .00   .17</td>
<td>.90</td>
<td>.76</td>
</tr>
<tr>
<td>Competence cocaine abuse subscale</td>
<td>.01   .09   .00   .06   .04   .19   .12   .51</td>
<td>.54</td>
<td>.76</td>
</tr>
<tr>
<td>Adherence mean scale score</td>
<td>.01   .33   .00   .06   .12   .25   .01   .23</td>
<td>.87</td>
<td>.93</td>
</tr>
<tr>
<td>Competence mean scale score</td>
<td>.01   .44   .00   .05   .02   .15   .04   .29</td>
<td>.81</td>
<td>.98</td>
</tr>
</tbody>
</table>

*Note.* SE = supportive-expressive psychodynamic therapy; CT = cognitive therapy; IDC = individual drug counseling; Tx = treatment; ICC = intraclass correlation. Competence is an average composite score of appropriateness and quality.
Table 2
Least Square Means of Adherence and Competence Ratings Across Treatments

<table>
<thead>
<tr>
<th>Scale</th>
<th>SE</th>
<th>CT</th>
<th>IDC</th>
<th>F</th>
<th>SD</th>
<th>Cohen’s d effect size indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDC Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence mean scale score</td>
<td>1.50↓</td>
<td>1.98↑</td>
<td>2.18</td>
<td>77.9*</td>
<td>0.26</td>
<td>IDC&gt;SE 2.62 IDC&gt;CT 0.77 CT&gt;SE 1.85</td>
</tr>
<tr>
<td>Competence mean scale score</td>
<td>3.07</td>
<td>3.59</td>
<td>4.04</td>
<td>97.6*</td>
<td>0.33</td>
<td>2.94 1.36 1.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence mean scale score</td>
<td>1.29↓</td>
<td>3.52↑</td>
<td>1.52</td>
<td>198.1*</td>
<td>0.51</td>
<td>CT&gt;SE 4.37 CT&gt;IDC 3.92 IDC&gt;SE 0.45</td>
</tr>
<tr>
<td>Competence mean scale score</td>
<td>1.78</td>
<td>4.01</td>
<td>2.00</td>
<td>211.4*</td>
<td>0.48</td>
<td>4.65 4.19 0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence Supportive subscale</td>
<td>3.13</td>
<td>2.63</td>
<td>2.14</td>
<td>50.6*</td>
<td>0.46</td>
<td>SE&gt;DC 2.15 SE&gt;CT 1.09 CT&gt;IDC 1.07</td>
</tr>
<tr>
<td>Competence Supportive subscale</td>
<td>3.90</td>
<td>3.05</td>
<td>2.77</td>
<td>102.5*</td>
<td>0.43</td>
<td>2.63 1.98 0.65</td>
</tr>
<tr>
<td>Adherence expressive subscale</td>
<td>2.31</td>
<td>1.89</td>
<td>1.40</td>
<td>49.4*</td>
<td>0.42</td>
<td>2.17 1.00 1.17</td>
</tr>
<tr>
<td>Competence Expressive subscale</td>
<td>3.63</td>
<td>3.09</td>
<td>2.92</td>
<td>73.2*</td>
<td>0.33</td>
<td>2.15 1.64 0.52</td>
</tr>
<tr>
<td>Adherence Cocaine Abuse subscale</td>
<td>1.90</td>
<td>2.33</td>
<td>2.29</td>
<td>12.0*</td>
<td>0.51</td>
<td>-0.76 -0.84 0.08</td>
</tr>
<tr>
<td>Competence Cocaine Abuse subscale</td>
<td>3.69</td>
<td>3.60</td>
<td>3.44</td>
<td>11.2*</td>
<td>0.23</td>
<td>1.09 0.39 0.70</td>
</tr>
<tr>
<td>Adherence mean scale score</td>
<td>2.66</td>
<td>2.36</td>
<td>2.01</td>
<td>35.1*</td>
<td>0.34</td>
<td>1.91 0.88 1.03</td>
</tr>
<tr>
<td>Competence mean scale score</td>
<td>3.71</td>
<td>3.17</td>
<td>2.98</td>
<td>86.7*</td>
<td>0.30</td>
<td>2.43 1.80 0.63</td>
</tr>
</tbody>
</table>

Note. For Cohen’s d, a small effect is .2, a medium effect is .5, and a large effect is .8. SE = supportive-expressive psychodynamic therapy; CT = cognitive therapy; IDC = individual drug counseling. Means with different subscripts are significantly different at the Bonferroni adjusted alpha of .0014. A mixed-model analysis was used to assess the treatment effect on outcome while adjusting for a fixed period effect and a random subject effect.
*\( p < .0001 \).
of the ACS-IDCCD. Similarly, the expert CT judges rated CT sessions higher than either SE or IDC sessions on the CTACS. Finally, the expert SE judges rated SE sessions higher than either CT or IDC sessions on the Supportive and Expressive subscales and overall scales of the ACS-SEC. However, note that the difference between SE and IDC sessions was significant only for adherence ratings. Furthermore, the findings for the SE adherence Cocaine Abuse subscale were actually in the wrong direction, although this might be explained by the questionable reliability of this subscale or by the fact that counselors in IDC dealt more with cocaine-related issues than did SE therapists.

Discussion

The results demonstrated that expert raters can, on average, reliably rate adherence and competence for each of the three individual treatments utilized in the NIDA CCTS. We have also shown that the therapists and counselors in the NIDA CCTS implemented the interventions and techniques belonging to their intended treatment modality more often than they implemented those from other treatments. Furthermore, we were able to show that interventions belonging to a treatment modality were more competently delivered by therapists who used that modality than by therapists who followed an alternative modality.

We did not, however, assess overall therapeutic skillfulness. Carroll et al. (1998) reported that Project MATCH therapists did not differ in overall therapeutic skillfulness across treatment conditions. In our study, we assessed therapists’ skillfulness as related to the specific delivery of a treatment. The closer measurement of nonspecific aspects of treatment assessed across treatment conditions, which bears some relations to therapist skillfulness and is available as part of the NIDA CCTS, is the alliance ratings filled out by patients and therapists. Crits-Christoph et al. (1999) reported that the therapeutic alliance did not differ among treatment groups. Barber et al. (1999) reported that although alliance improved from Session 2 to Session 5, the improvement was very small and alliance did not increase differentially across treatments.

Although all treatments were clearly distinguishable, as suggested by the large effect sizes that generally characterized the discriminability analysis and the fact that treatment modality explained most of the variation in the ratings in the variance decomposition analysis, there was some evidence that therapies differed in the degree to which techniques from other modalities were used. Quite clearly, SE dynamic therapists tended to score lowest on both the CT and IDC scales, suggesting that CT and IDC might have more in common with each other than with SE. This result might reflect the fact that CT and IDC are more focused directly on drug use than is SE therapy. On the SE scale, CT therapists received higher scores than IDC counselors. This finding is likely to be consistent with the fact that both CT and SE are more supportive and interpretive treatments than IDC. In light of the findings that IDC was the most efficacious treatment, it is noteworthy that the absolute levels of adherence and competence in the IDC condition were low. It is not clear whether this is a property of the rating scales (i.e., for IDC, inclusion of a large number of items assessing techniques that, although important to the overall treatment method, do not need to occur within every session for competent delivery of the treatment) or whether adherence/competence in IDC was inadequate. However, the clinical supervisors’ impressions were that all treatments were delivered as intended. That is, there was no indication, that IDC, SE, or CT was not delivered competently.

Of course, it is possible that the discrimination of the three treatments was due to, or exaggerated by, the demand characteristics facing our expert judges (i.e., judges viewed the task as one in which their job was to demonstrate that psychotherapies are different,
particularly their own preferred therapy in comparison with other therapies). Because it was generally impossible to keep the expert judges blind to the type of treatment upon hearing a taped session, it could be that judges were biased to provide higher adherence and competence ratings to the treatment of their expertise. While it might be possible to diminish this demand characteristic by using a nonexpert judge, it is unclear whether nonexperts could adequately judge whether a therapist was using a specific therapeutic intervention and, if so, whether it was competently delivered. If it is the case that only an expert can sufficiently judge whether an intervention was delivered competently, perhaps this demand characteristic could be lessened by using judges who are expert in multiple treatments and who have less allegiance to a particular treatment. It seems unlikely, however, that such judges would be readily available for discriminability studies (given that specialization generally occurs in one area) or that this method would be cost effective (given these judges’ level and breadth of expertise). Of course, these are empirical questions that future discriminability studies could address.

As was true for other contemporary discriminability studies (e.g., Hill et al., 1992; Hogue et al., 1998), it is important to keep in mind that the reliability of the adherence and competence ratings from the present study refers to the mean rating from two expert raters. Preliminary analyses of the present study indicated that reliabilities estimated for one expert judge were decent, but they were not uniformly as acceptable as the average ratings, indicating that discriminability studies require the use of more than one expert judge given the findings to date. Despite the use of two expert judges, our judges could not reliably implement the SE Competence Cocaine Abuse subscale. Preliminary analyses found that these estimates were improved only negligibly by eliminating items from this subscale; thus, it is unclear whether we failed to provide sufficient training on these items or whether the items themselves are problematic. Thus, future studies will need to either provide more elaborate training on these items or use a different measure. In sum, the present results suggest that the manual-guided individual treatments delivered as part of the NIDA CCST were implemented as intended, with a high level of fidelity and with clear delineation from each other.

References
Carroll, K.M., Connors, G.J., Cooney, N.L., DiClemente, C.C., Donovan, D.M., Kadden, R.R.,


