Integrative Cognitive Therapy for Depression: A Preliminary Investigation

Louis G. Castonguay, Alexander J. Schut, Deane E. Aikins, and Michael J. Constantino
The Pennsylvania State University

Jean-Philippe Laurenceau
University of Miami

Laura Bologh
The Pennsylvania State University

David D. Burns
Stanford University

This study is a preliminary investigation of an integrative treatment aimed at improving the efficacy of cognitive therapy (CT) for depression. The development of the treatment protocol was based on process findings, which suggested that strategies used in CT to resolve alliance ruptures may actually exacerbate problems in the therapeutic relationship. The protocol integrates, within the traditional CT treatment manual, procedures to repair alliance ruptures that are derived from or consistent with humanistic and interpersonal therapies. Although conducted by inexperienced therapists, the integrative treatment led to greater improvement than a waiting-list condition. The results also compare favorably to previous findings for CT.

4

Cognitive therapy (CT) is one of the most frequently used and empirically tested treatments for depression. There is ample evidence, however, that a number of depressed individuals either fail to respond adequately to this treatment or do not respond at all (Robins & Hayes, 1993). In the recent National Institute of Mental Health (NIMH) collaborative study, for example, the rate of recovery of the clients who completed CT varied from 51% to 65% (depending on the criteria used for clinical recovery; Elkin et al., 1989). As demonstrated by Elkin, Gibbons, Shea, and Shaw (1996), the NIMH findings are similar to those obtained in most comparable studies of CT for depression.

These results suggest that the efficacy of CT can be improved. As recently argued by Grawe (1997), a fruitful way to improve the outcome of psychotherapies is to modify existing treatments on the basis of process research. A recent investigation of the process of change in CT suggests that part of its limited efficacy might be due to the procedures used in this approach to deal with ruptures in the therapeutic alliance. On the basis of a manualized CT protocol conducted by experienced therapists, Castonguay, Goldfried, Wiser, Raue, and Hayes (1996) found that although the quality of the therapeutic alliance was positively related to client change, therapists' focus on intrapersonal issues such as the causal relationship between the client's thoughts and his or her emotions (a crucial aspect of cognitive techniques) correlated *negatively* with outcome. Qualitative analyses conducted to understand the latter, unexpected finding revealed that the cognitive therapists in this study attempted to resolve alliance ruptures (e.g., reluctance to engage in prescribed interventions, lack of trust toward the therapist) by increasing their adherence to cognitive techniques. They did so either by trying to persuade the client of the validity of the cognitive rationale or, as prescribed in the CT treatment manual (Beck, Rush, Shaw, & Emery, 1979), by identifying negative therapeutic reactions as evidence of clients' distorted thoughts (about therapy or the therapist) that need to be challenged. These interventions, however, seemed to worsen the alliance and thus potentially interfere with client change.

Of course, these findings do not suggest that the therapeutic relationship is disregarded in CT. The importance of a positive and collaborative relationship between therapist and client has been emphasized by the pioneers of CT (see Beck et al., 1979). Leading cognitive therapists have also demonstrated the beneficial impact of empathy on the client's symptoms (Burns & Nolen-Hoeksema, 1992). What these findings suggest, however, is that the efficacy of CT may be improved by the adoption of more appropriate strategies aimed at repairing alliance problems. Consistent with the contribution of client-centered therapy (Rogers, 1957), Burns (1989; Burns & Auerbach, 1996) has developed several techniques to address alliance ruptures during treatment with CT. These include three

listening skills, called thought and feeling empathy (paraphrasing the client's criticisms and feelings of anger, frustration, or disappointment about the therapist or the therapy), inquiry (asking gentle, probing questions to learn more about the client's negative feelings), and the disarming technique (finding truth in the client's criticisms even when these criticisms may seem unreasonable and unfair). Burns also describes two self-expression skills called "I feel" statements (if appropriate, the therapist expresses his or her feelings using tactful language) and stroking (expressing respect for the client, even if the client seems hostile). According to Burns, if the therapist uses these methods skillfully, the client will frequently feel validated and understood and will also resume his or her engagement with the therapeutic task prescribed in CT. Burns (1989; Burns & Auerbach, 1996; Burns & Nolen-Hoeksema, 1992) also advocates the use of session-by-session assessment of the therapeutic alliance with a self-administered Empathy Scale (ES; Burns, 1989, 1995; Persons & Burns, 1985) completed by the client between therapy sessions. Changes in scores on the ES allow therapists to track alliance failures more accurately and rapidly.

Influenced by the contribution of humanistic and interpersonal therapists, Safran (Safran & Segal, 1990) has suggested the use of similar therapeutic tools to help cognitive therapists recognize and resolve relationship ruptures. Like Burns (1989), he has emphasized the importance of the therapist's empathy to facilitate the client's description of alliance ruptures and expression of negative feelings associated with them. Similar to the procedure of disarming, he advocates the therapist's exploration of his or her potential contributions to the problems in the therapeutic relationship (e.g., inadequate presentation of the therapeutic rationale, misunderstanding of client experience, negative emotions toward the client). Safran has also argued that therapists' willingness to recognize their mistakes or empathic failures frequently facilitates clients' openness and exploration of their own role in the alliance ruptures. Such exploration is believed to foster clients' awareness of core schema and long-standing, maladaptive patterns of relationship with others. According to Safran, the emergence and resolution of alliance ruptures can thus provide unique opportunities for deep and global therapeutic change. Other well-known cognitive–behavioral therapists have recommended the use of humanistic and/or interpersonal strategies to repair alliance ruptures. Included in Linehan's (1993) "techniques of acceptance," for example, is the use of empathy and direct validation, which, like Burns's disarming technique, involves the therapist's ability to see reasonableness in the client's dysfunctional behaviors, accept the client's hostile affect, and recognize his or her own mistakes. Like Burns and Safran, Linehan has argued that alliance problems are frequent and that their resolution can lead to the acquisition of skills

that can be used by the client when dealing with interpersonal difficulties outside the sessions.

The goal of the present study is to provide a preliminary investigation of the efficacy of an integrative version of CT that, on the basis of Burns (1989) and Safran's (Safran & Segal, 1990) work, integrates humanistic and interpersonal interventions to resolve problems in the therapeutic relationship. Although the integration movement in psychotherapy has received considerable attention in the last 2 decades, up until now it has been primarily based on clinical observations and has suffered from a paucity of empirical findings (Kendall, 1982; Lambert, 1992; Wolfe & Goldfried, 1988). As a first step in a process of empirical validation, the present study compares the effectiveness of this integrative form of CT (ICT) to a waiting-list condition (WL). It was predicted that ICT would lead to greater decrease of symptomatology than WL.

Comparisons of a treatment with a no-treatment condition represent a crucial step in building evidence for the empirical support of such treatment (Chambless & Hollon, 1998). As Chambless and Hollon (1998) cogently argued,

> The fundamental question clients presenting for treatment are likely to pose is, "Does the treatment you propose for me actually work?" This question is addressed by the comparison of treatment with some type of minimal or no-treatment condition (e.g., waiting list or assessment-only control) in which clients receive the assessment procedures but no significant amount of active treatment. (p. 8)

As specified by Chambless and Hollon (1998), a treatment that has been found to be more beneficial than no treatment in two or more studies conducted by different research teams is to be considered *efficacious*. When defined by the criteria delineated by Chambless and Hollon (1998), the present study is an attempt to determine whether ICT is *possibly efficacious* (i.e., treatment that has received support for its efficacy by one study or by numerous studies conducted by the same team). Assuming that our hypothesis is supported (i.e., ICT is superior to a WL condition), the present study could then be followed by a number of investigations. A replication study conducted in another lab would be required for ICT to meet the *efficacious* criteria. At a later phase, studies comparing ICT with other treatments or with a placebo could then assess whether ICT is both *efficacious and specific* (Chambless & Hollon, 1998)—including the comparison of ICT and CT to test whether the strategies to repair alliance ruptures specific to ICT represent active mechanisms of change. Future research could also be conducted to assess the effectiveness and cost effectiveness of ICT (Chambless & Hollon, 1998). As the first phase of our research program, however, the present study is a logical step to begin evaluating whether ICT works and whether it deserves empirical investigations in the future.

METHOD

Participants

Participants, recruited by advertisements, met the following inclusion criteria: (a) between 18 and 55 years old, (b) *DSM-IV* criteria for major depressive disorder (MDD), and (c) a score of at least 20 on the Beck Depressive Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Exclusion criteria included (a) current treatment for depression; (b) substance abuse disorder for the last 12 months, panic disorder, manic or hypomanic episode (past or present), or psychotic disorder (past or present); and (c) too distressed for (or unwillingness to accept) potential assignment to a 12- to 15-week WL condition. Of 239 individuals who responded to the advertisements, 29 met the study's criteria after completing the assessment procedures. One person, however, declined to participate in the study after completion of the assessment procedures.

Of the 28 individuals who participated in the study, 14 were assigned to ICT and 14 to WL. Six (2 in ICT [14.3%] and 4 in WL [28.6%]) participants dropped out before the end of the comparison period. One participant assigned to the ICT condition was deleted from the analyses because of her therapist's failure to adhere to the treatment protocol. This left 11 participants in the ICT condition and 10 in the WL condition for outcome analyses (the main analyses were also conducted with the participant for whom the therapist did not adhere to the treatment protocol).

No significant differences were found between the two groups in terms of demographic variables. Participants were mostly middle aged ($M = 38.8$ years old, $SD = 10.9$), female (76.2%), White (100%), and educated (89% had college education or graduated from college), and more than half were married (57.1%).

Procedures

Participants were first screened over the phone. Those who met inclusion criteria (with the exception of the BDI) and failed to meet exclusion criteria were assessed for a second time in a clinical interview using the mood and psychosis sections of the Structured Clinical Interview (SCID; Spitzer, Williams, Gibbon, & First, 1989) for the *DSM–III–R* (modified to reflect the changes in the diagnostic criteria of MDD for the *DSM–IV*) and parts of the Anxiety Disorders Interview Schedule (Brown, DiNardo, & Barlow, 1994; to rule out panic disorder and substance abuse), as well as the Global Assessment of Functioning scale (GAF, or Axis V of *DSM–IV*).

Individuals who met the study's criteria (including a BDI score of at least 20) were asked to come for another assessment session, which included the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960). On the basis of the assignment procedure used by Paivio and Greenberg (1995), participants were then alternatively assigned to either ICT or WL. (Although such a sequential assignment is not a deliberate randomized procedure, it is unlikely that systematic bias in condition assignment occurred because unknown and random factors determined when participants contacted us.)

During the waiting period, participants in WL were contacted by phone weekly by a therapist. Participants were asked a list of specific questions designed to monitor their depression symptoms, level of functioning, life crises, and suicidality. The weekly phone calls lasted from less than 1 min to 15 min. Participants in ICT received a minimum of 12 and a maximum of 20 sessions of individual therapy (biweekly for the first 2 to 4 weeks, and weekly for the rest of the treatment). At the end of WL and ICT, participants were assessed on the MDD section of the SCID, the GAF, and the HDRS.

Although the treatment was planned to be 12 to 15 weeks long, cancellations and the difficulties in setting weekly and/or make-up sessions increased the length of the treatment to an average of 17.3 weeks ($SD = 3.55$), which is significantly longer than the WL period ($M = 13.00$, $SD = 1.89$), $t(19) = 3.39$, $p < .01$.

Treatment

Therapy was conducted by seven graduate students in clinical psychology at The Pennsylvania State University who had not been previously trained in CT for depression before serving as therapists in this study. Before being assigned a client in ICT, therapists were trained for a period of 8 to 18 months. In addition to extensive reading, observation of cases, and role playing, the training involved a 5-day intensive workshop provided by David Burns, which focused on CT and the techniques he developed to resolve alliance ruptures emerging in this treatment (all but one therapist participated in Burns's workshop).

The protocol required therapists to conduct CT following the guidelines of Beck et al.'s (1979) treatment manual, unless problems in the therapeutic relationship emerged during treatment. The main elements of the CT manual are (a) behavioral techniques to increase activities, (b) identification of automatic thoughts and their impact on the client's emotions, (c) challenge of automatic thoughts, (d) identification and challenging of core underlying schema, and (e) rehearsal of coping strategies to deal with depressive episodes after completion of treatment. Therapists identified

alliance ruptures by reviewing clients' scores on an empathy scale (Burns, 1989, 1995) filled out after each session and by observing alliance rupture markers during therapy sessions, as described in Safran and Segal (1990). When signs of alliance ruptures were identified, therapists applied strategies described in Burns and Auerbach (1996) and Safran and Segal (1990), which included (a) invitation for the client to explore the potential rupture, (b) empathic response to the client's emotional reaction toward the therapist and/or therapy, and (c) disarming (exploration and validation of at least some aspects of the client's perception of the therapist's contribution to the alliance rupture). After addressing the alliance rupture, therapists resumed the application of CT, either by continuing to apply the technique that was used before the rupture emerged or by changing to another technique included in the CT treatment manual (e.g., shifting from a challenge of distorted thoughts to the use of skill training techniques to facilitate behavior change). Weekly group or individual supervision was conducted by Louis G. Castonguay to monitor therapists' adherence to the treatment protocol.

Measures

Before and at the end of ICT and WL, participants were assessed using two well-validated instruments of depressive symptomatology: the BDI, a self-report measure, and HDRS (24-item version, transformed to allow comparison with the 17-item version used in many studies), a semistructured interview. The GAF was used to provide a general measure of psychological, social, and occupational functioning.

Interrater Agreement

The clinical interviews at pretreatment were conducted by one of nine graduate students, all trained in the administration of the SCID. An independent assessor (also a trained graduate student) reviewed the audiotaped recordings of approximately 25% of the study's participants (randomly selected),[1] along with an equivalent number of taped interviews randomly selected from individuals who did not participate in the study because they failed to meet the study's criteria. Kappa for MDD diagnosis

---

[1]The last 4 participants in the study were not included when tapes were selected for the calculation of interrater estimates. The percentages of tapes selected, however, are based on the entire sample.

was .80 (with perfect agreement for the individuals who participated in the study), and although the intraclass *r* for the GAF was low (.31), the Finn's *r* (an index of interrater agreement that takes into account restricted variance in the scores; Whitehurst, 1984) for the GAF was .95. At pretreatment, the HRSD interviews were conducted by one of three trained graduate students who showed an acceptable level of interrater agreement (intraclass $r = .74$; Finn's $r = .99$) on the basis of an independent review of approximately 25% of the study participants' audiotaped HDRS (randomly selected and assigned to coders). At the end of WL and ICT, participants were assessed by one of two graduate students (blind to the experimental conditions) on the SCID section of MDD, the GAF, and the HDRS (these two assessors did not assess MDD and GAF at pretreatment but were part of the HDRS assessment group). Independent reviews of approximately 25% of the postassessment tapes showed a perfect agreement on MDD criteria among these two assessors, as well as a high level of interrater agreement for the GAF (intraclass $r = .81$) and HDRS (intraclass $r = .98$).

Additional Comparisons

Although the goal of the current study was to compare ICT with a no-treatment condition, we thought that it would be interesting to compare the results of this preliminary study with findings obtained in CT when conducted with clinically experienced therapists in previous research. The findings of the NIMH collaborative study (Elkin et al., 1989) were used for the purpose of comparison not only because of the exemplary methodological sophistication of this large-scale study but also because a comprehensive effort was made to ensure that CT would be conducted in a consistent, standardized, and competent way (see Elkin, 1999). Experienced and carefully selected therapists were trained by recognized experts (e.g., A. Beck, B. Shaw), received intensive supervision during a training phase (during which they treated several pilot cases), met competence criteria before taking part in the outcome study, received additional monitoring and consultation during the outcome study, and demonstrated adherence to CT. Furthermore, as we mentioned previously, the NIMH findings have been shown to be similar to those obtained in most comparable studies of CT for depression (Elkin et al., 1996). As such, we felt that the NIMH results provided a stringent criterion to compare our results. It should be clear, however, that such comparison is in no way equivalent to a comparative study involving both ICT and CT. In the context of a very first step at supporting the potential efficacy of ICT, such comparisons would simply provide some (albeit indirect and tentative) indication that it shows any

promise of improving CT before a direct comparison of ICT and CT could be considered.

As in the current study, individuals who received treatment in the collaborative study were mostly middle aged, female, and White (Elkin et al., 1989). As compared to the NIMH study, however, participants in the current study appeared to be more educated (89% vs. 75% had college education or graduated from college), and a larger number were married (57.1% vs. 36%; I. Elkin, personal communication, 1999). In contrast with the NIMH study, therapists in the current study were inexperienced and did not have to show competence in the treatment protocol before being assigned a client (only four of them had attempted to apply the protocol with a case before treating a participant in the study). However, the current therapists received weekly supervision, which was not the case in the NIMH study. It should also be noted that although the ICT treatment was longer than the WL, the average number of sessions (16.6) was equivalent to the average number of sessions (16.2) for completers of the treatments in the NIMH study (Elkin et al., 1989, 1996).

RESULTS

Table 1 presents means and standard deviations for each condition on the outcome measures at pre- and posttest. The two conditions did not differ significantly on any of the pretreatment variables. Paired t tests indicated significant improvement from pre- to posttest on all three mea-

Table 1. Symptomatic Measures at Pre- and Posttest

Measure	ICT			WL			CT-NIMH			ICT Post-WL		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD
BDI												
Pretest	11	27.36	4.34	10	27.50	5.28	37	26.8	8.4	3	26.0	5.29
Posttest	11	4.82	3.66	10	17.50	8.80	37	10.81	11.94	3	6.67	2.08
HDRS												
Pretest	11	15.00	4.40	10	18.00	2.66	37	19.2	3.6	3	15.67	3.75
Posttest	11	4.05	2.53	10	11.70	5.89	37	7.81	6.80	3	4.33	2.25
GAF/GAS												
Pretest	11	53.09	7.01	10	54.20	5.73	37	52.8	7.2	3	58.67	3.21
Posttest	11	78.36	8.91	10	67.20	15.11	37	69.19	12.04	3	88.33	2.89

Note. ICT = integrative cognitive therapy; WL = waiting list condition; CT-NIMH = cognitive therapy condition of the National Institute of Mental Health Collaborative Study; ICT post-WL = group of participants who received ICT after completing WL and meeting the study's inclusion criteria before beginning treatment; BDI = Beck Depression Inventory; HDRS = Hamilton Depression Rating Scale; GAF = Global Assessment of Functioning Scale (used in the present study); GAS = Global Assessment Scale (used in National Institute of Mental Health Collaborative Study).

sures for ICT, $ts(10) = 6.15$ to 14.01, all $ps < .001$, and WL, $ts(9) = 2.77$ to 3.5, all $ps < .05$. (Considering the small sample size and the exploratory nature of the present study, it should be noted that we decided not to correct the alpha level of .05 for family error in order not to decrease statistical power; cf. Cohen, 1994. It should also be mentioned that even if no difference between conditions was found on outcome measures at pretest, these pretest variables were controlled in the between-groups analyses to provide base-free assessment of change; cf. Borkovec, Newman, Pincus, & Little, 2002.)

Three analyses of covariance (ANCOVAs; one for each dependent measure) were performed, with pretreatment scores as covariates.[2] All three ANCOVAs were significant: BDI, $F(1, 18) = 18.81, p < .001$; HDRS, $F(1, 18) = 15.49, p < .01$; GAF, $F(1, 18) = 6.10, p < .05$. In all ANCOVAs, the group differences were in favor of ICT.[3] Using posttest means and pooled standard deviations (cf. Lipsey & Wilson, 2001), we found large effect sizes (above .80; Cohen, 1988) for each dependent variable (BDI = 1.91; HDRS = 1.72; GAF = 0.91).[4]

---

[2]The use of a multivariate analysis of covariance was considered for the two depression measures (BDI and HDRS). High correlation of these variables at outcome ($r = .88$), however, suggested a problem of multicollinearity. Further, in light of the fact that a multivariate analysis of variance is often less powerful than an analysis of variance (Tabachnick & Fidell, 1996), the exploratory nature of the study, and the small sample size, led us to conduct separate ANCOVAs.

[3]The main analyses were also conducted with the participant for whom the therapist did not adhere to the treatment protocol (careful monitoring of the first six sessions of the treatment—which represented more than 25% of the entire therapy—revealed only two brief attempts by the therapist to have the client use one behavioral technique, i.e., progressive relaxation, one very brief and unsuccessful intervention to challenge distorted thoughts, no attempt to set and follow a specific session agenda, and no review of the client's homework. The therapist also failed to address several alliance ruptures regarding the tasks and goals of treatment). As was done above, three ANCOVAs were conducted (using pretreatment scores as covariates). The ANCOVAs were significant for the BDI, $F(1, 19) = 15.54, p < .01$, and HDRS, $F(1, 19) = 6.40, p < .05$. For the GAF, the ANCOVA was only marginally significant, $F(1,19) = 3.21, p < .1$. It should be noted, however, that even when this participant is added to the ICT group, the means at posttreatment on the dependent variables compare favorably to what was obtained for the CT of the NIMH (BDI, $M = 5.92, SD = 5.16$; HDRS, $M = 5.33$, $SD = 5.07$; GAF, $M = 76.33, SD = 11.03$; see Table 1 for the NIMH means and standard deviations).

[4]Because the length of treatment was significantly different between ICT and WL, we also conducted three additional ANCOVAs, using the length of treatment as an additional covariate. The ANCOVAs were significant for the BDI, $F(1, 17) = 8.34, p < .05$, and HDRS, $F(1, 17) = 8.33, p < .05$. For the GAF, the ANCOVA failed to reach statistical significance, $F(1, 17) = 2.99, p = .1$. It should be noted, however, that these results are most likely undervaluing the superiority of ICT over WL, as these analyses violated the precept of independence between the independent variable and the covariate (the point-biserial correlation between the length of treatment and group status was .61). As argued by Tabachnick and Fidell (1996), "violation of this precept results in removal of some portion of the effect of the IV on the DV—that portion of the effect that is associated with the covariate" (p. 326). This explains the reduction of F values observed from the main analyses.

At posttest, none of the 11 participants in ICT met *DSM–IV* criteria for MDD, whereas 5 of the 10 participants in WL still met these criteria (chi-square analysis was significant; Fisher's exact test, one-tailed, $p < .05$). We also calculated the number of participants in each group showing clinically significant change (Jacobson & Truax, 1991) on the BDI and HDRS using the same reliability estimates and cutoff scores that were used to determine clinically significant changes in the collaborative data set (Ogles, Lambert, & Sawyer, 1995). As was done in the Ogles et al. (1995) study, pretreatment scores (on both the BDI and the HDRS) were adjusted before reliable change indices were calculated because regression toward the means was detected (Speer, 1992). Also like Ogles et al. (1995) did, we determined clinical significance by comparing the participants to two reference groups: a functional group (representing a nondistressed or general population sample) and an asymptomatic group (without symptoms of distress). With the functional group as the reference group, 91% of the ICT participants met criteria for clinically significant change on the BDI, and 82% met criteria on the HDRS. In contrast, only 20% of WL participants met criteria for the BDI, and 30% for the HDRS. Chi-square analyses revealed that the proportion of participants meeting criteria for clinically significant change was statistically different for both the BDI (Fisher's exact test, one-tailed, $p < .01$) and HDRS (Fisher's exact test, one-tailed, $p < .05$). With the same functional group as the reference group, the percentages of clinically significant change observed for the CT condition of the NIMH were 50% for the BDI and 68% for the HDRS. With an asymptomatic group as the reference group, a substantial number of ICT participants met stringent criteria for clinical significance (64% for both the BDI and HDRS). In contrast, only 20% of WL participants met these stringent criteria for the BDI and HDRS. Although chi-square analyses conducted between ICT and WL were only marginally significant for the BDI and HDRS (Fisher's exact test, one-tailed, $p = .06$), the percentages of ICT participants reaching clinically significant change with the asymptomatic criteria compare favorably to those of the NIMH–CT condition (28% for the BDI and 45% for the HDRS).

As indicated in Table 1, the BDI and GAF[5] pretest scores of the

---

[5]The comparison between the present study and the NIMH collaborative study on observer-rated global functioning should be interpreted cautiously. Whereas the Global Assessment of Functioning scale (GAF) of the *DSM–IV* (Axis V) was used in the present study, the Global Assessment Scale (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976) was used in the collaborative study. However, because the GAF was adapted from the GAS, the scales are very similar. One difference is that the GAS specifically instructs raters to base their assessment on the client's functioning over the past week, whereas the GAF makes no specific instruction for time frame. It seems fair to argue, however, that in both studies what was assessed was the clients' current level of global functioning.

participants who completed the treatment in this study were equivalent to the scores of individuals who completed the CT condition in the NIMH collaborative study (pretest data for the 37 individuals who completed CT were taken from Elkin et al., 1989; posttest data were obtained from I. Elkin, personal communication, 1998, because the means in Elkin et al., 1989, were adjusted for pretest scores and marital status). The mean of the HDRS at pretest, however, was higher for the CT condition of the NIMH collaborative study. (Because the two samples were equivalent in terms of BDI, however, this discrepancy may reflect different calibration used by assessors in rating participants' responses in the two studies; I. Elkin, personal communication, 1998). Table 1 also indicates that for the BDI and GAF, change from pre- to posttreatment obtained in the present study with inexperienced graduate student therapists compares favorably to what was achieved with more clinically experienced therapists in the NIMH study.

Immediately after completing the WL condition, 4 participants received and completed ICT. A 5th participant did not experience symptoms of depression at the end of the WL but 3 months later met criteria for MDD and presented with a moderate to high level of symptoms (BDI = 32; HRSD = 19.5; GAF = 60). At her request, she was treated using ICT. As for the 5 other participants in the WL, 1 did not receive treatment because of an absence of symptoms, 1 chose not to receive treatment, 1 had to interrupt her treatment for 2.5 months because of a medical condition, and 2 dropped out. Of the 5 participants who completed ICT after WL, 3 still met the study inclusion criteria before receiving ICT. The pre- and posttreatment scores for these participants are presented in Table 1. At posttest, none met *DSM-IV* criteria for MDD. Paired *t* tests indicated significant improvement from pre- to posttest on all three outcome measures, $ts(13) = 7.22$ to 14.79, all $ps < .001$, for the 14 participants (11 clients in the experimental group, and 3 WL clients after receiving ICT) who met the inclusion and exclusion criteria before beginning treatment. With a general population as the reference group, 86% of these 14 participants met criteria for clinically significant change on the BDI, and 79% met criteria on the HDRS. When the asymptomatic group was used as the reference group, 50% of the 14 participants met criteria for clinically significant change on the BDI, and 64% met criteria on the HDRS. The percentages favorably compare to the ones obtained for the CT condition of the NIMH study.

Six months after the end of ICT, 12 of the 14 participants meeting the study's criteria before treatment filled out a BDI. One of them, however, had received additional treatment after the completion of therapy. The means with ($M = 8.08, SD = 9.97$) and without ($M = 7.18, SD = 9.93$) this participant were within the general population range (i.e., less than 13.46; Ogles et al., 1995).

DISCUSSION

Although eclectic or integrative therapy is the primary orientation of clinical psychologists (Mahoney, 1990), the effectiveness of such an approach has received almost no attention from researchers. The present study provides preliminary empirical support for an integrative form of psychotherapy for depression, which was based on findings of process research as well as on the clinical work of leaders in the field. Although conducted by inexperienced therapists, this treatment led to significantly greater therapeutic gains than a WL condition. As such, ICT appears to meet Chambless and Hollon's (1998) criteria for *possibly efficacious* treatment (assuming that our sequential assignment of participants in either group is functionally equivalent to a randomized control trial). Also appropriate for the measurement of treatment efficacy in a preliminary investigation, large effect sizes were found for all dependent variables. Furthermore, high rates of clinically significant change were observed for the treatment condition. Follow-up data also suggest that most of the therapeutic gains, at least in terms of depressive symptoms, were maintained 6 months after treatment.

Although no direct statistical comparisons were conducted, the change obtained at the end of treatment compares favorably, at least in terms of a self-report measure of depression and an observer-rated measure of general functioning, to what was observed for individuals who completed the CT condition of the NIMH collaborative study. The results also compare favorably to the findings of a recent meta-analysis of high-quality controlled studies (Gloaguen, Cottraux, Cucherat, & Blackburn, 1998), which revealed an estimate effect size (for the BDI) of .82 when comparing CT for depression and WL or placebo conditions (which is less than half of the effect size obtained in the present study, i.e., 1.92, even after Hedges's, 1981, correction for small sample bias was applied to the present data, i.e., 1.84).

Because there are inevitably several differences between any two studies, the apparent superiority of the present findings over the results of the NIMH study should be considered very cautiously. For example, although the therapists in the NIMH study had more clinical experience, treated several cases and met competence criteria before being assigned actual participants, and were trained by CT leaders, they also received less intense supervision during the outcome phase of the study. Furthermore, none of the clients in the NIMH study were deleted from the analyses for potential lack of therapist adherence. Moreover, a higher percentage of married participants, higher level of education, and a lower pretreatment HDRS scores in the present investigation may explain the difference in outcome between the two studies (although, as mentioned above, the lower HDRS

scores may have resulted from different calibration systems). On the other hand, the outcome differences between the two studies observed for the BDI and GAF are consistent with the empirical evidence supporting the role of a positive alliance in therapy. Identified as one of the most robust predictors of outcome across theoretical orientations, the alliance has been found to explain as much as 35% of the outcome variance in CT (Castonguay et al., 1996). In addition, using structural equation modeling techniques, Burns and Nolen-Hoeksema (1992) have reported that the quality of the therapeutic alliance has a substantial direct causal effect on recovery from depression in patients treated with CBT. Moreover, studies on psychodynamic therapy and CBT have shown that even small amounts of negative transaction patterns between therapist and client are strongly related to negative outcome (Critchfield, Henry, Isaac, Castonguay, & Borkovec, 1996; Henry, Schacht, & Strupp, 1986). At the minimum, the comparison of our findings with those of NIMH provide some indication (albeit indirect and tentative) that the addition of specific interventions aimed at properly repairing relationship problems in CT might improve the efficacy of this approach in treating depression. The next logical step for this research should involve the replication of these findings by another research team and the direct comparison of the efficacy of ICT to CT.

Because of the small sample size, the results should of course be considered with caution. It should be noted, however, that although smaller than the number of CT completers in the NIMH study, the number of individuals who met the study's inclusion criteria and completed ICT (as part of the experimental group and after completing WL) is larger than the average sample size for treatment groups in a large meta-analysis of psychotherapy studies (i.e., 12; Shapiro & Shapiro, 1983). Although large-sample studies have become the expected norm when the goal is to replicate the efficacy of well-established treatments, small-sample studies are still needed to provide preliminary support to innovative treatments (e.g., Foa, Abramowitz, Franklin, & Kozak, 1999; Telch, Agras, & Linehan, 2000). As wisely argued by Katz (1962), a research report may be acceptable "even though not definitive in its results, because it opens up a new and promising area of research. The person who blazes a trail should not necessarily be asked to make it a paved highway" (p. 2).[6]

It should also be mentioned that although we intended ICT and WL to be equal in total number of weeks, ICT was significantly longer. Of course, ethical considerations prohibit the creation of a WL that is overly extensive and/or flexible enough to account for phenomena that are part of most treatment delivery (e.g., missing sessions). For this reason, the present study is not the only outcome investigation for which the treatment con-

---

[6]We are grateful to Marvin Goldfried for directing us to this reference.

dition was longer than the WL. In some recent studies (e.g., Flannery-Schroeder & Kendall, 2000; Kendall et al., 1997), for instance, the WL was only half the length of the treatment condition. Nevertheless, secondary analyses conducted after length was controlled for showed the ICT to be significantly superior to WL with regard to the BDI and HRSD. Moreover, the true effect of treatment on the GAF is more than likely to have been undervalued because of the lack of independence between the independent variable and covariate (i.e. length of treatment; Tabachnick & Fidell, 1996). Considering the removal of some treatment effects, as well as the small sample, it is noteworthy that the ANCOVA (controlling for length of treatment) conducted with the GAF approached statistical significance ($p = .1$).

It should also be noted that objective measures of treatment adherence were not included in this report. The observation of numerous sessions for the purpose of adherence monitoring, however, indicated that alliance ruptures emerged frequently during therapy, that therapists were able to apply the noncognitive strategy prescribed by the treatment protocol, and that they were able to resume the application of CT once the alliance problems had been addressed and resolved.

REFERENCES

Beck, A. T., Rush, J. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy for depression*. New York: Guilford Press.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Borkovec, T. D., Newman, M. G., Pincus, A. L., & Little, R. (2002). A component analysis of cognitive–behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70,* 288–298.

Brown, T. A., DiNardo, P. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV)*. Albany, NY: Graywind Publications.

Burns, D. D. (1989). *The feeling good handbook*. New York: Morrow.

Burns, D. D. (1995). *Therapist's toolkit: Comprehensive assessment and treatment tools for the mental health professional.* Unpublished manuscript.

Burns, D. D., & Auerbach, A. (1996). Therapeutic empathy in cognitive-behavioral therapy: Does it really makes a difference? In P. Salkovskis (Ed.), *Frontiers of cognitive therapy* (pp. 135–164). New York: Guilford Press.

Burns, D. D., & Nolen-Hoeksema, S. (1992). Therapeutic empathy and recovery from depression in cognitive–behavioral therapy: A structural equation model. *Journal of Consulting and Clinical Psychology, 60,* 441–449.

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting outcome in cognitive therapy for depression: A comparison of unique and common factors. *Journal of Consulting and Clinical Psychology, 64,* 497–504.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The Earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Critchfield, K. L., Henry, W. P., Isaac, M. V., Castonguay, L. G., & Borkovec, T. D. (1996,

June). *A preliminary analysis of interpersonal process in cognitive behavioral psycho-therapies: The search for common factors in outcome prediction*. Poster presented at the meeting of the Society for Psychotherapy Research. Amelia Island, GA.

Elkin, I. (1999). A major dilemma in psychotherapy outcome research: Disentangling thera-pists from therapies. *Clinical Psychology: Science and Practice, 6,* 10–32.

Elkin, I. E., Gibbons, R. D., Shea, M. T., & Shaw, B, F. (1996). Science is not a trial (but it can sometimes be a tribulation). *Journal of Consulting and Clinical Psychology, 64,* 92–103.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health treatment of depression collaborative research pro-gram: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982.

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.

Foa, E. B., Abramowitz, J. S., Franklin, M. E., & Kozak, M. J. (1999). Feared consequences, fixity of belief, and treatment outcome in patients with obsessive-compulsive disorder. *Behavior Therapy, 30,* 717–724.

Flannery-Schroeder, E. C., & Kendall, P. C. (2000). Group and individual cognitive-behavioral treatments for youth with anxiety disorder: A randomized clinical trial. *Cog-nitive Therapy and Research, 24,* 251–278.

Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I.-M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders, 49,* 59–72.

Grawe, K. (1997). Research-informed psychotherapy. *Psychotherapy Research, 7,* 1–19.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurological and Neurosurgical Psychiatry, 23,* 56–61.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Henry, W. P., Schacht, T. E., & Strupp, H. H. (1986). Structural analysis of social behavior: Application to a study of interpersonal process in differential psychotherapeutic out-come. *Journal of Consulting and Clinical Psychology, 54,* 27–31.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psy-chology, 59,* 12–19.

Katz, D. (1962). Editorial. *Journal of Abnormal and Social Psychology, 64,* 1–2.

Kendall, P. C. (1982). Integration: Behavior therapy and other schools of thought. *Behavior Therapy, 13,* 559–571.

Kendall, P. C., Flannery-Schroeder, E., Panichelli-Mindel, S. M., Southam-Gerow, M., Henin, A., & Warman, M. (1997). Therapy for youths with anxiety disorders: A second ran-domized clinical trial. *Journal of Consulting and Clinical Psychology, 65,* 366–380.

Lambert, M. J. (1992). Psychotherapy outcome research: Implications for integrative and eclectic therapists. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psycho-therapy integration* (pp. 94–129). New York: Basic Books.

Linehan, M. M. (1993). *Cognitive-behavioral treatment for borderline personality disorder*. New York: Guilford Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Mahoney, M. J. (1990). *Human change processes*. New York: Guilford Press.

Ogles, B. M., Lambert, M. J., & Sawyer, J. D. (1995). Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Journal of Consulting and Clinical Psychology, 63,* 321–326.

Paivio, S. C., & Greenberg, L. S. (1995). Resolving "unfinished business": Efficacy of expe-riential therapy using empty-chair dialogue. *Journal of Consulting and Clinical Psychol-ogy, 63,* 419–425.

Persons, J. B., & Burns, D. D. (1985). Mechanism of action of cognitive therapy: Relative contribution of technical and interpersonal intervention. *Cognitive Therapy and Re-search, 9,* 539–551.

Robins, C. J., & Hayes, A. M. (1993). An appraisal of cognitive therapy. *Journal of Consulting and Clinical Psychology, 61,* 205–214.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21,* 95–103.

Safran, J. D., & Segal, Z. V. (1990). *Interpersonal process in cognitive therapy*. New York: Basic Books.

Shapiro, D. A., & Shapiro, D. (1983). Comparative therapy outcome research: Methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology, 51,* 42–53.

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402–408.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1989). *Structured Clinical Interview for the DSM-III-R*. Washington, DC: American Psychiatric Association.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.

Telch, C. F., Agras, W. S., & Linehan, M. M. (2000). Group dialectical behavior therapy for binge-eating disorder: A preliminary, uncontrolled trial. *Behavior Therapy, 31,* 569–582.

Whitehurst, G. J. (1984). Interrater agreement for journal manuscript reviews. *American Psychologist, 39,* 67–70.

Wolfe, B. E., & Goldfried, M. R. (1988). Research on psychotherapy integration: Recommendations and conclusions from an NIMH workshop. *Journal of Consulting and Clinical Psychology, 56,* 448–451.

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!