Journal of Consulting and Clinical Psychology
1996, Vol. 64, No. 2, 295-304

Copyright 1996 by the American Psychological Association, Inc.
0022-006X/96/$3.00

SPECIAL FEATURE

A Component Analysis of Cognitive–Behavioral Treatment for Depression

Neil S. Jacobson
University of Washington

Keith S. Dobson
University of Calgary

Paula A. Truax, Michael E. Addis, Kelly Koerner,
Jackie K. Gollan, Eric Gortner, and Stacey E. Prince
University of Washington

The purpose of this study was to provide an experimental test of the theory of change put forth by A. T. Beck, A. J. Rush, B. F. Shaw, and G. Emery (1979) to explain the efficacy of cognitive–behavioral therapy (CT) for depression. The comparison involved randomly assigning 150 outpatients with major depression to a treatment focused exclusively on the behavioral activation (BA) component of CT, a treatment that included both BA and the teaching of skills to modify automatic thoughts (AT), but excluding the components of CT focused on core schema, or the full CT treatment. Four experienced cognitive therapists conducted all treatments. Despite excellent adherence to treatment protocols by the therapists, a clear bias favoring CT, and the competent performance of CT, there was no evidence that the complete treatment produced better outcomes, at either the termination of acute treatment or the 6-month follow-up, than either component treatment. Furthermore, both BA and AT treatments were just as effective as CT at altering negative thinking as well as dysfunctional attributional styles. Finally, attributional style was highly predictive of both short- and long-term outcomes in the BA condition, but not in the CT condition.

The cognitive model of depression (Beck, Rush, Shaw, & Emery, 1979) states that depressed individuals have stable cognitive schemas (also referred to as underlying assumptions or core beliefs) that develop as a consequence of early learning. These schemas predispose people toward negative interpretations of life events (i.e., cognitive distortions or automatic thoughts [ATs]), which in turn, lead the depressed person to engage in depressive behavior. Cognitive–behavioral therapy (CT) for depression includes interventions that focus on publicly observable behavior, dysfunctional ATs, and inferred underlying cognitive structures or schemas. The treatment is conducted in a progressive manner so that the therapist first focuses on overt behavior change; teaches the client to assess and, when necessary, correct situation-specific distortions in thinking; and finally moves to the identification and modification of more stable depressive schemas and presumed cognitive structures.

Neil S. Jacobson, Paula A. Truax, Michael E. Addis, Kelly Koerner, Jackie K. Gollan, Eric Gortner, and Stacey E. Prince, Department of Psychology, University of Washington; Keith S. Dobson, Department of Psychology, University of Calgary, Calgary, Alberta, Canada.

Preparation of this article was supported by Grants 2R01 MH44063-06 and 5K02 MH00868-05 from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to Neil S. Jacobson, Department of Psychology, University of Washington, 1107 N.E. 45th Street, #310, Seattle, Washington 98105-4631.

A number of investigators have documented the clinical usefulness of CT for depression. In a meta-analysis of this approach, Dobson (1989) suggested that CT is at least as powerful and perhaps more effective than behavior therapy, pharmacotherapy, and other psychotherapies or waiting-list control conditions. Some have questioned the state of this evidence (Hollon, Shelton, & Loosen, 1991), in part based on the results of the Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989). However, even in the TDCRP, CT showed long-term effects that were at least as durable, if not more durable, than pharmacotherapy or interpersonal psychotherapy (Shea et al., 1992).

Beck and his associates are quite specific about the hypothesized active ingredients of CT, stating throughout their treatment manual (Beck et al., 1979) that interventions aimed at cognitive structures or core schema are the active change mechanisms. Despite this conceptual clarity, the treatment is so multifaceted that a number of alternative accounts for its efficacy are possible. We label two primary competing hypotheses the "activation hypothesis" and the "coping skills" hypothesis.

According to the activation hypothesis, CT effects change through the activation of clients; that is, by instigating them to become active again and to put themselves in contact with available sources of reinforcement. Instigative interventions play a major role particularly in the early stages of CT and may be largely responsible for its effectiveness. It has been noted that much of the change during CT occurs within the first few weeks

(Rush, Beck, Kovacs, & Hollon, 1977), when instigations toward activation play a prominent role in the treatment. Previous studies that found CT more effective than behavioral activation ([BA] e.g., Shaw, 1977) may not have used activation strategies that work as well as those used in CT. If an entire treatment based on activation interventions proved to be as effective as CT, the cognitive model of change in CT (stipulating the necessary interventions for the efficacy of CT) would be called into question. Moreover, in addition to these important theoretical questions, there are important practical considerations: Are the elaborate cognitive interventions directly designed to modify core schema necessary? It may be that a much more parsimonious set of treatment procedures would have comparable effects.

A second hypothesis that could explain the efficacy of CT is the coping skills hypothesis. According to this hypothesis, clients learn to cope with depressing events and depressogenic thinking during CT, and it is this new set of skills that, along with activation, accounts for the alleviation of depressive behavior. In other words, it is not that core cognitive structures are altered, but that people learn effective coping strategies for dealing with life stress and the ATs associated with these events. If structural changes in core schema are really necessary for changes in clients' depression, then CT should be significantly more effective than a treatment that stops with the training in modifying automatic dysfunctional thinking in specific situations.

To test these competing hypotheses, we conducted an experiment comparing three treatments for major depression in adults: one that included only behavioral activation (BA treatment), another that included both BA and work on ATs (AT treatment), and a third that corresponded to the full cognitive therapy of depression (CT treatment). The full CT treatment included not only work on BA and AT, but also a direct focus on identifying and modifying core depressogenic schema. According to the cognitive theory of depression, CT should work significantly better than AT, which in turn, should work significantly better than BA.

A second purpose of the present study was to examine the correlations between changes in specific mechanisms and outcome, both within and across treatments. For example, independently of whether BA worked as well as CT, was CT more successful at modifying cognitive schema than BA? In other words, do the various treatments differentially effect the processes that they are supposed to effect? A related question concerns whether the three treatments operate by means of different mechanisms, independently of their overall efficacy. For example, BA may work as well as CT but for different reasons. Do the variables that correlate with a positive acute treatment response differ across the treatments? One may expect BA to be more highly correlated with outcome in the BA condition than in the CT condition. More generally, it would be of interest to know how treatments effect change, independently of how well they work.

Method

Sample

The sample consisted of 152[1] participants who met criteria for major depression according to the *Diagnostic and Statistical Manual of Mental Disorders* (3rd edition, revised; *DSM-III-R*; American Psychiatric Association, 1987), scored at least 20 on the Beck Depression Inventory (BDI; Beck et al., 1979), and scored 14 or greater on the 17-item Hamilton Rating Scale for Depression. (HRSD; Hamilton, 1967). *DSM-III-R* diagnoses were based on the Structured Clinical Interview for *DSM-III-R* (SCID; Spitzer, Williams, & Gibbon, 1987). Originally, training was provided by Michael First from the biometrics research department at the New York State Psychiatric Institute, where the SCID was developed. Further training and supervision was provided by Donna Miller, an experienced psychiatrist and expert SCID interviewer who was on site. The interviews themselves were conducted by clinical psychology graduate students, carefully trained and supervised by Miller. Raters were not informed of treatment condition. Interrater reliability between Miller and SCID raters was .90, on the basis of the percentage of times Miller and the rater agreed on the primary diagnosis. Raters were also not informed of which tapes were being rated by Miller.

Similarly, the HRSD was administered as an adjunct to the SCID. For a previous study, our research group had rewritten the HRSD so that it could be inserted into a structured diagnostic interview (Whisman et al., 1989). This version of the HRSD has excellent psychometric properties and is highly reliable. Moreover, although it is a clinical interview, it can be administered by technicians without loss of reliability. Raters were not informed of the treatment condition of the participant or of which tapes were being assessed for reliability.

Eighty percent of the participants were referred directly from Group Health Cooperative, the largest health maintenance organization (HMO) in the state of Washington. The remainder were recruited from public service announcements. Of the original 152 participants accepted into the study, 110 were women and 42 were men. One hundred thirty-seven of this group completed therapy (defined as receiving at least 12 sessions of treatment). Thus, in all there were 15 dropouts (an 8% attrition rate): Three of these dropouts refused random assignment and never had a treatment session. The rates of attrition during acute treatment were comparable in the three conditions.

Exclusion criteria included a number of concurrent psychiatric disorders (bipolar or psychotic subtypes of depression, panic disorder, current alcohol or other substance abuse, past or present schizophrenia or schizophreniform disorder, organic brain syndrome, and mental retardation). We also excluded participants who were in some concurrent form of psychotherapy, who were receiving psychotropic medication, or who needed to be hospitalized because of imminent suicide potential or psychosis.

After qualifying for the study, participants were randomly assigned to one of the three treatment conditions after matching to ensure group equivalence on the following variables: number of previous episodes of depression, presence or absence of dysthymia, severity of depression, gender, and marital status (married, divorced, single, or widowed).

Table 1 shows means and standard deviations for the demographic variables. We first correlated these variables with primary measures of treatment outcome to determine whether they should be used as covariates in the primary analyses. None of them (gender, years of education, marital status, or percent Caucasian) were significantly correlated with either posttreatment BDI and HRSD scores, or changes from pre- to posttreatment on these two measures of depression severity. There were also no significant differences between treatment conditions on any of these variables, nor did the treatments differ in their pretreatment BDI scores. However, clients receiving CT and AT treatment had significantly higher pretreatment HRSD scores than did those receiving BA treatment, $F(2, 148) = 3.52, p < .05$.

[1] Sample sizes differ for some analyses because of missing data.

Table 1
Demographics and Pretreatment Variables

Variable	BA (n = 57)	AT (n = 44)	CT (n = 50)
Gender n (and %)			
Male	16 (28.1)	10 (22.7)	12 (24)
Female	41 (71.9)	34 (77.3)	38 (76)
Mean age (in years)	36.6	38.3	39.2
Education n (and %)			
High school graduate	57 (100)	44 (100)	49 (98)
College graduate	36 (63.1)	22 (50)	27 (54)
Postcollege	18 (31.6)	8 (18.2)	12 (24)
Marital status n (and %)			
Never married	20 (35.1)	18 (20.9)	11 (22)
Married once	19 (33.3)	13 (29.5)	15 (30)
Divorced	12 (21.1)	8 (18.2)	13 (26)
Widowed	2 (3.5)	1 (2.3)	7 (14)
Ethnicity n (and %)			
African American	1 (1.9)	0 (0)	3 (6)
Hispanic	2 (3.7)	1 (2.3)	0 (0)
Caucasian	50 (92.6)	40 (90.9)	38 (76)
Native American	1 (1.9)	2 (4.5)	3 (6)
Asian	0 (0)	1 (2.3)	2 (4)
Mean (and *SD*) no. of previous episodes	4.6 (5.3)	5.4 (5.7)	5.9 (6.3)
Mean (and *SD*) with pretreatment depression			
BDI	29.3 (6.9)	29.2 (6.6)	29.8 (6.3)
HRSD[a]	17.4 (3.8)	19.3 (4.0)	19.1 (4.4)

Note. $N = 151$. Percentages are within-group. Unequal *ns* for variables reflect missing data. BA = behavioral activation; AT = automatic thoughts; CT = cognitive–behavioral therapy.
[a] $F(2, 148) = 3.52, p < .05$.

Therapists

Four experienced cognitive therapists provided treatment in all three conditions. Their average age was 43.5 years (range, 37 to 49 years), and they averaged 14.8 years of postdegree clinical experience (range, 7 to 20 years). They had been practicing CT treatment for an average of 9.5 years since their formal training, with a range of 8 to 12 years.

All four therapists had participated in at least one previous clinical trial in which they served as research therapists for CT treatment. Despite their having previous experience, a year was devoted to training the therapists in and piloting the component (BA and AT) treatments. Three manuals were created for this study, one for each treatment condition.[2] All were based on the original CT manual (Beck et al., 1979) but included specific guidelines for prescribed and proscribed interventions in each treatment condition.

We developed a system for monitoring and calibrating for protocol adherence. One of the coauthors (Keith S. Dobson) listened to a randomly determined 20% of all audiotaped therapy sessions. Therapists were immediately contacted if a protocol violation occurred. In addition, monthly meetings were held involving authors Neil S. Jacobson and Keith S. Dobson, the project coordinator, and all therapists to discuss any ambiguities regarding protocol or treatment integrity issues. Therapists were also encouraged to "flag" any ambiguities in past sessions or any concerns they had about adherence in upcoming sessions. Finally, as we describe later, adherence was systematically evaluated independently of these calibration procedures.

Treatments

BA condition. In the Beck et al. (1979) CT manual, a common early strategy is to identify behavior problems and to invoke a series of interventions designed to activate people in their natural environment. These early strategies consist mostly of semistructured activities. Included in the list of interventions considered to have an activation focus are (a) monitoring of daily activities, (b) assessment of the pleasure and mastery that is achieved by engaging in a variety of activities, (c) the assignment of increasingly more difficult tasks that have the prospect of engendering a sense of pleasure or mastery, (d) cognitive rehearsal of scheduled activities, in which participants imagine themselves engaging in various activities with the intent of finding obstacles to the imagined pleasure or mastery expected from those events, (e) discussion of specific problems (e.g., difficulty in falling asleep) and the prescription of behavior therapy techniques for dealing with them; and (f) interventions to ameliorate deficits in social skills (e.g., assertiveness, communication skills). In the BA condition, activation is the exclusive focus for 20 sessions.

Activation and the modification of dysfunctional thoughts (AT condition). From a CT perspective, activation is only the first area that requires assessment and modification in depression. Thus, although it was possible to structure an entire protocol around this behavioral intervention, Beck et al. (1979) advocate that the therapist move relatively quickly into cognitive interventions. A number of techniques have been developed to identify and modify automatic dysfunctional thoughts. In particular, cognitive therapists listen for "cognitive distortions," which are negative construals of different events that precipitate sad feelings and depressed behavior. The CT therapist typically moves past activation interventions to begin to assess these negative patterns of thinking and to teach clients to be aware of them so that they can then be modified. Within this general framework, a number of different techniques have been developed to assess and modify dysfunctional thoughts. These include the following: (a) noticing mood shifts in therapy sessions and asking for the thoughts that preceded the mood shift, (b) using a daily record of dysfunctional thoughts as a form of personal diary in which the clients note particularly problematic events and the types of thoughts that surrounded those events, (c) reexamining thoughts in specific situations and determining whether the event warranted the types of conclusions that the patient had drawn about it, (d) helping clients learn how to respond in a more functional manner to negative thinking, (e) examining the possibility of attributional biases or mistakes in the way the clients see the causes of various successes and failures in their lives, and (f) the development of homework assignments in which the clients assess the validity of their negative interpretations. In this study, the AT condition permitted the use of all interventions from the BA condition and those listed earlier. The only proscription in this condition was the opportunity to work on underlying core beliefs or schemas.

CT condition. CT, in its complete form, includes the identification and modification of more general patterns of thought that are stable and presumably the causes of cognitive distortions and negative feelings. There are a number of specific interventions that are typical of therapists when they attempt to modify schema. They include the following: (a) use of the "downward arrow," a technique in which the therapist asks the client for their explanations about why certain problems have emerged, which then leads to the therapist hypothesizing various types of general concerns and eventually to the identification of core beliefs; (b) the explicit identification of underlying assumptions and core beliefs, either by direct report of the client or by inference on the part of the therapist; (c) the identification of alternative assumptions or core beliefs; (d) the discussion of the advantages and disadvantages of holding various assumptions or core beliefs; (e) the discussion of the short-term versus the long-term advantages of various assumptions or beliefs; (f) the assignment of homework that allows patients to determine whether they actually use certain assumptions or core beliefs in the way

[2] These manuals are available from Neil S. Jacobson.

they deal with their life circumstances and to explore the application of other assumptions to those circumstances; and (g) the use of the same techniques involved in modifying dysfunctional thinking, except in this case they are applied to core beliefs rather to situation-specific dysfunctional thinking.

In this study, the CT condition allowed the use of the full range of BA, AT, and CT interventions. To ensure a fair test of the core schema hypothesis, however, we required that a minimum of eight sessions have a primary focus on assumptive work.

Outcome Measures

All participants were evaluated before therapy, at the time of termination, and at 6-, 12-, 18-, and 24-month follow-ups. In this article, we focus on the immediate effects of treatment and on those at the 6-month follow-up. We include measures of depressive symptoms and the presence or absence of major depression, which were based on reports from clients and from clinical evaluators.

To assess the presence or absence of major depression at posttest, clinical evaluators gave participants a modified version of the Longitudinal Interval Follow-Up Evaluation II (LIFE; Keller et al., 1987), developed to assess the longitudinal course of psychiatric disorders. The LIFE includes a semistructured interview that allows one to assess psychopathology over the previous 6 months. In our modified version, criteria for the diagnosis of depression were changed from the Research Diagnostic Criteria used on the original LIFE to those used in the *DSM–III–R*. To determine presence or absence of major depression, we used weekly psychiatric ratings on a scale ranging from 1 (*absence*) to 6 (*presence*). We used the LIFE measure to determine whether participants continued to meet *DSM–III–R* criteria for major depression at posttest.

Participants were also given the 17-item version of the HRSD, administered by a clinical evaluator. This is a widely used interviewer-based measure of depression severity.

As a second self-report measure of depression severity, the BDI (Beck et al., 1979) was administered to participants before and after treatment. This is another widely used measure of depression severity that correlates highly with the HRSD, has excellent psychometric properties (Beck, Steer, & Garbin, 1988), and is sensitive to clinical change (Edwards et al., 1984; Lambert, Shapiro, & Bergin, 1986).

Data Analysis

All analyses of outcome were conducted on those participants who completed at least 12 sessions of treatment ("completers"), those who completed the maximum allotment of 20 treatment sessions ("maximum completers"), and those who had at least one session of therapy but dropped out before completing 12 treatment sessions ("dropouts"). For dropouts, the last available score on each outcome measure served as the termination score. Posttest HRSD and BDI scores served as the primary measures of depression severity. Analyses of covariance, with pretreatment scores on the dependent measures used as covariates, were applied to compare the efficacy of three treatments.

Treatment response was also analyzed categorically. To assess the percentage of participants in each treatment condition who either recovered or improved but failed to recover, we looked at the percentage who scored 8 or less on the BDI. These criteria, although arbitrary, were recommended by Frank et al. (1991) in an effort to standardize measures of recovery in depression research.[3] Participants were categorized as improved but not recovered if they no longer met *DSM–III–R* criteria for major depression at posttest but continued to report BDI scores greater than 8. Contingency table analyses were used to compare treatments in improvement and recovery rates.

Results

Adherence to Treatment Protocols

The measure of treatment integrity used in the present study was a modified version of the National Institute of Mental Health Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon, Evans, Elkin, Lowery, 1984). Items included both techniques designated by the treatment manual and those prohibited or proscribed by it. Ideally, the three treatment conditions should have been most different on items reflecting interventions addressing the modification of dysfunctional ATs and core schema, as all three conditions included BA. Moreover, protocol violations should have been kept to a minimum. Our scale had 7 items measuring the use of interventions focused on BA, 12 measuring work on ATs, and 7 measuring work on underlying assumptions (UA) or core schema, as well as 3 items reflecting interventions that are proscribed in all three conditions. We were also interested in "potency," that is, the ratio of interventions that are essential to the treatment to those that are compatible with it but neither unique nor essential to CT (Waltz, Addis, Koerner, & Jacobson, 1993). Five items were added in the category ENU (essential but not unique to BA, AT, or CT; e.g., setting an agenda, assigning homework); also, 11 items were added in the category COMPAT, which reflected nonessential interventions that are compatible with all conditions (e.g., skills training, assessing general functioning) but essential to none.

Thus, the total scale had 45 items, grouped into the aforementioned six scales. Raters listened to a tape of the therapy session, taking notes as they listened, and then rated each item on a scale ranging from 0 (*not at all*) to 6 (*extensively or thoroughly*). Nine clients were randomly selected from each condition for adherence ratings, for a total of 27 clients. For each of these clients, one early, one middle, and one late session were randomly selected, with sessions 1 and 20 excluded. Thus, a total of 81 tapes were rated.

Treatment condition was kept masked to trained coders. Intraclass correlation coefficients were used to determine interrater reliability. The mean intraclass correlations were .81, ranging from .73 to .89 across the six scales.

As Table 2 indicates, therapists were successful at keeping the treatments distinct. Therapists confined themselves to BA interventions in that condition and to BA and AT interventions in the AT condition and used all three types of interventions in the CT condition. The average ratings were exactly as we had expected: BA items were common in all three conditions but most common in the BA condition; AT interventions were common in both CT and AT conditions; and work on core schema was common only in the CT condition. On an absolute basis, almost no protocol violations were detected.

Another test of adherence involved asking the following question: Did the rate of occurrence of BA, AT, and UA interventions exceed the random fluctuations expected by chance? This question addressed whether the little bit of AT and UA work

[3] Results were virtually unchanged when alternative criteria recommended by Frank et al. (1991) were adopted: HRSD scores less than 7 or at least 8 weeks of not meeting criteria for major depression.

Table 2
Adherence Ratings by Treatment Condition

Type of item and condition	Rating for phase of therapy			Overall rating
	Early	Middle	Late	
Items measuring BA				
BA	83	86	67	79
AT	67	52	52	57
CT	112	53	60	75
Items measuring AT work				
BA	2	7	10	6
AT	119	133	62	105
CT	139	100	124	121
Items measuring core schema work				
BA	0	4	1	2
AT	0	9	2	4
CT	34	60	82	59

Note. In each condition, $n = 27$. The higher the score, the more frequently or thoroughly these interventions were made. Ratings were made on a Likert-type scale ranging from *not at all* (0) to *extensively or frequently* (6). The means here are multiplied by 10 to illuminate differences, and analyses were based on raw scores. BA = behavioral activation; AT = automatic thoughts; CT = cognitive–behavioral therapy.

that occurred in BA and AT conditions, respectively, differed significantly from the "noise level" that one would expect by chance. Simultaneously, it allows one to be assured that the mean ratings of BA, AT, and UA interventions differed significantly from zero when they were supposed to differ. There were no statistically significant deviations from treatment protocols in any condition.

In the BA condition, only BA interventions significantly differed from zero, $t(26) = 4.95, p < .05$. In the AT condition, both the AT intervention, $t(26) = 3.89, p < .05$, and the BA intervention, $t(26) = 3.63, p < .05$, occurred to a significant degree, but UA interventions did not, $t(26) = 1.03$, *ns*. However, in the CT condition, all three types of interventions occurred to a statistically significant degree: For BA, $t(26) = 4.61, p < .05$; for AT, $t(26) = 4.47, p < .05$; for CT, $t(26) = 3.1, p < .05$.

Finally, to compare the treatment conditions for potency, we examined the scale totals for ENU, COMPAT, and PROSCR (proscribed), interventions. No significant between-group differences emerged for any of these three scales.

Keith S. Dobson, who provided supervision for all therapists in the project, randomly selected tapes in the CT condition and rated them for competence on the Cognitive Therapy Scale (CTS), the accepted instrument for assessing competence in CT. The convention, albeit arbitrary, is to use a score of 40 as the cutoff for competence on the CTS. The overall means were above 40, as were the means for each therapist: For Therapists 1–4, *M*s = 45.16, 44.01, 47.91, and 46.17, respectively.

Treatment Outcome

Table 3 presents the means, standard deviations, and results of our primary outcome analyses. Results are presented first for the total sample (including dropouts) and then for each of three subsamples: maximum completers, completers, and dropouts. Pretreatment group differences were assessed through one-way analyses of variance (ANOVAs). With the exception of the HRSD on the total sample, there were no significant pretreatment differences between conditions.

The primary treatment outcome analyses consisted of 3 (Treatment Group) × 4 (Therapist) multivariate analyses of covariance (MANCOVAs), with posttest BDI and HRSD scores serving as dependent variables and pretest scores on the respective pretest measures serving as covariates. The MANCOVAs for the total sample failed to uncover statistically significant differences among treatments, $F(4, 252) = 1$, *ns*; therapists, $F(6, 252) = 1$, *ns*; or Therapist × Treatment interactions, $F(12, 252) < 1$, *ns*. Similar MANCOVAs for the completers showed group equivalence for treatments, $F(4, 238) = 1$, *ns*; therapists, $F(6, 238) < 1$, *ns*; and Therapist × Treatment interactions, $F(12, 238) < 1$, *ns*. Finally, for the subsample of maximum completers, there were no differences among treatments, $F(4, 226) = 1$, *ns*; therapists, $F(6, 226) < 1$, *ns*; or Treatment × Therapist interactions, $F(12, 226) < 1$, *ns*. To protect familywise error rates, we chose .01 as our level of significance. However, none of the MANCOVAs were significant at even the .05 level. Table 3 indicates the results of ANCOVAs for each subsample and each measure. Because none of the therapist or Therapist × Treatment interaction effects were significant, only the main effects for treatment are presented in Table 3. As indicated, there were no significant differences between the treatments on either the BDI or the HRSD. When we looked at the results separately for each therapist, we similarly found no differences among treatment conditions on either measure.

We also looked at the proportion of clients in each condition who improved and recovered to assess the clinical significance of each treatment condition (Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991). Table 4 presents the improvement and recovery rates for each of the treatments in each of the four samples. Chi-square analyses revealed no significant differences between treatments on improvement or recovery in any of the four samples. The mean improvement rate was 62.3% for the complete sample, 66% for maximum completers, 58.3% for partial completers, and 16.7% for dropouts. The mean recovery rate was 51.5% for the complete sample, 54.5% for maximum completers, 58.3% for partial completers, and 5.6% for dropouts. Dropouts had significantly lower rates of improvement and recovery than maximum completers ($\chi^2[1, N = 141] = 7.8, p < .01$; and $\chi^2[1, N = 141] = 9.5, p < .01$, respectively) and completers ($\chi^2[1, N = 149] = 9.51, p < .01$; and $\chi^2[1, N = 149] = 7.8, p < .01$, respectively).

Six-Month Follow-Up Results

Table 3 includes follow-up scores on the BDI and the HRSD for all participants. We were able to obtain follow-up data on all but one participant (a 99% retention rate). We conducted ANCOVAs on both measures, with follow-up scores as the dependent variables and posttest scores as covariates. This analysis provides a parametric test for changes during the follow-up period on depressive symptoms as a function of treatment condition. The analyses found that there were no significant differ-

Table 3
Mean Pretreatment, Postreatment, and 6-Month Follow-Up Scores for BDI and HRSD for Four Samples of Participants in Each Treatment Condition

Depression and measure	BA		AT		CT		
	n	*M*(*SD*)	*n*	*M*(*SD*)	*n*	*M*(*SD*)	*F*(*dfs*) and *p*
Total sample (*n* = 149)							
BDI							
Pre	56	29.3 (6.6)	43	29.1 (6.6)	50	29.8 (6.3)	$F(2, 148) < 1$, *ns*
Post	56	9.1 (7.9)	43	10.6 (9.3)	50	10.1 (9.6)	$F(2, 145) < 1$, *ns*
6 months	50	8.5 (7.6)	39	9.3 (8.2)	47	10.3 (8.6)	$F(2, 132) < 1$, *ns*
HRSD							
Pre	56	17.4 (3.9)	43	19.1 (3.9)	50	19.1 (4.4)	$F(2, 148) = 3.5$, $p < .05$
Post[a]	53	6.4 (4.6)	40	6.9 (5.8)	48	7.2 (6.7)	$F(2, 138) < 1$, *ns*
6 months	50	6.6 (4.8)	39	7.7 (6.1)	47	6.4 (5.1)	$F(2, 132) < 1$, *ns*
Maximum completers (*n* = 129)							
BDI							
Pre	48	29.3 (7.2)	37	29.2 (7.0)	44	28.9 (5.9)	$F(2, 126) < 1$, *ns*
Post	48	8.5 (7.9)	37	9.1 (8.7)	44	9.7 (9.2)	$F(2, 128) < 1$, *ns*
6 months	44	8.3 (7.8)	37	9.1 (8.4)	43	10.4 (8.7)	$F(2, 120) < 1$, *ns*
HRSD							
Pre	48	17.3 (3.8)	37	19.1 (4.1)	44	18.6 (3.3)	$F(2, 126) = 2.72$, *ns*
Post[b]	47	6.5 (4.8)	37	6.4 (5.6)	44	6.8 (5.7)	$F(2, 125) < 1$, *ns*
6 months	44	6.7 (4.9)	37	7.6 (6.3)	43	6.2 (4.8)	$F(2, 120) < 1$, *ns*
Completers (*n* = 137)							
BDI							
Pre	50	29.2 (7.1)	39	29.0 (6.9)	48	29.5 (6.2)	$F(2, 134) < 1$, *ns*
Post	50	8.4 (7.8)	39	9.3 (8.6)	48	9.3 (9.0)	$F(2, 134) < 1$, *ns*
6-month follow-up	46	8.2 (7.6)	38	9.2 (8.3)	47	10.3 (8.6)	$F(2, 127) < 1$, *ns*
HRSD							
Pre	50	17.4 (3.7)	39	19.1 (4.0)	48	18.9 (4.2)	$F(2, 134) = 3.2$, $p < .05$
Post[c]	49	6.5 (4.7)	38	6.4 (5.5)	47	6.7 (5.7)	$F(2, 131) < 1$, *ns*
6-month follow-up	46	6.5 (4.9)	38	7.7 (6.2)	47	6.4 (5.1)	$F(2, 127) < 1$, *ns*
Dropouts (*n* = 12)							
BDI							
Pre	6	30.3 (6.2)	4	29.5 (3.4)	2	36.0 (7.1)	$F(2, 9) = 1.04$, *ns*
Post	6	14.5 (7.9)	4	23.8 (2.2)	2	27.5 (7.8)	$F(2, 9) = 3.48$, *ns*
HRSD							
Pre	6	18.7 (5.3)	4	19.3 (3.6)	2	24.0 (8.5)	$F(2, 9) = .79$, *ns*
Post[d]	4	6.2 (4.5)	2	15.0 (8.5)	1	32.0 (.0)	$F(2, 4) = 3.6$, *ns*

Note. BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; BA = behavioral activation; AT = automatic thoughts; CT = cognitive–behavioral therapy; Pre = pretreatment; Post = posttreatment.
[a] Eight participants of the complete sample were unavailable for posttest and are therefore missing HRSD scores. BDI scores were taken from the final therapy session. [b] One participant in the completer sample was unavailable for posttest and, therefore, did not have an HRSD score. [c] Two participants in the partial completer sample were unavailable for posttest and therefore did not have HRSD scores. [d] Five participants of the complete sample were unavailable for posttest and are therefore missing HRSD scores.

ences between treatment conditions, $F(2, 132) < 1$, *ns*. As Table 5 indicates, the treatments were also equivalent in the ultimate impact of therapy: this conclusion is derived from ANCOVAs in which follow-up scores served as dependent variables and pretest scores served as covariates. Thus, the three treatments did not differ either in the overall impact of therapy through the 6-month follow-up or in changes in depressive symptoms over the first 6 months after posttest.

Table 5 shows the percentage of participants in each condition who had recovered during the course of therapy and relapsed by the time of the 6-month follow-up, based on the LIFE interview. Relapse was defined as meeting criteria for major depression, and we used three different definitions of recovery: 8 consecutive weeks of not meeting criteria for major depression, ending therapy with a BDI score of 8 or less, and ending therapy with an HRSD score of 7 or less. Contingency table analyses indicated that, regardless of how recovery was defined, groups did not differ significantly in relapse rates.

We also compared the recovered participants in all three treatment conditions on the number of "well weeks" during the follow-up period, again using three criteria for recovery. A well week was defined as a week when there were no or minimal

Table 4
Percentage Improved and Recovered at Posttreatment in Each Treatment Condition

Status	BA		AT		CT	
	n	%	*n*	%	*n*	%
Total sample (n = 149)						
Unimproved	22	39.3	18	41.9	16	32.0
Improved	34	60.7	25	58.1	34	68.0
Recovered	26	46.4	22	51.2	28	56.0
Maximum completers (n = 129)						
Unimproved	18	37.5	13	35.1	13	29.5
Improved	30	62.5	24	64.9	31	70.5
Recovered	24	50.0	21	56.8	25	56.8
Completers (n = 137)						
Unimproved	19	38.0	14	35.9	14	29.2
Improved	31	62.0	25	64.1	34	70.8
Recovered	25	50.0	22	56.4	28	58.3
Dropouts (n = 12)						
Unimproved	3	50.0	4	100.0	2	100.0
Improved	3	50.0	0	0.0	0	0.0
Recovered	1	16.7	0	0.0	0	0.0

Note. *Improved* is defined as no longer qualifying for major depressive disorder at posttest; *recovered* is defined as no major depressive disorder at posttest and Beck Depression Inventory scores less than 8. The p values (all of which were not significant) are from chi-square analyses. BA = behavioral activation; AT = automatic thoughts; CT = cognitive–behavioral therapy.

symptoms, based on the LIFE interview. The maximum score was 26. As Table 6 shows, there were no significant differences between treatment conditions, regardless of how recovery was defined.

To summarize, none of our follow-up analyses uncovered differences between groups. CT did not lead to decreased relapse, or better long-term functioning in terms of depressive symptoms, than did either of the component treatments.

Mechanisms of Change

In this series of analyses, we took two approaches. First, we looked at the impact of each treatment condition on the processes that it was expected to effect, as well as those allegedly outside its domain; thus, we examined the degree to which each condition resulted in increased behavioral activation, decreased negative thinking, and alterations in depressogenic cognitive structures. Second, we tried to establish a temporal relationship between changes in particular mechanisms and outcome. We used the Pleasant Events Schedule (both frequency and pleasure ratings) as our measure of behavioral activation (MacPhillamy & Lewinsohn, 1971), the Automatic Thoughts Questionnaire (Hollon & Kendall, 1980) as our measure of dysfunctional thinking, and the Expanded Attributional Style Questionnaire (EASQ; Peterson & Villanova, 1988) as our measure of cognitive structures.

To examine the degree to which the three treatments resulted in mechanism changes, we compared pre- and posttreatment scores on the three measures using paired t tests. When clients were considered as an aggregate, there were significant improvements on each of these mechanism measures. Clients in all conditions increased their frequency and enjoyability of pleasant events; decreased their negative thinking; and showed significantly lowered tendencies to attribute negative events to internal, stable, and global factors.

We examined differential outcomes based on treatment condition. We used ANCOVAs to compare between-groups changes in these measures, with the pretreatment score on each of the criterion variables serving as covariate. None of these measures changed differentially as a function of treatment condition.

It was still possible that change in a mechanism could be a cause of later depression change in one treatment and a consequence in another (Hollon, DeRubeis, & Evans, 1987). One way of evaluating this possibility was to examine the temporal relationship between change in depression and cognitive and behavioral mechanisms. Again, following DeRubeis and Feeley (1990), we calculated residual change scores from pre- to mid-treatment (early change) and from mid- to posttreatment (late change) on both the BDI and each mechanism measure. Table 7 shows the correlations between early residual change in cognitive and behavioral mechanisms and late residual change in depression in each treatment. Contrary to what was expected, early change in two subscales of the EASQ were associated with later change in the BA but not in the CT treatment. Participants in the BA treatment who made less negative attributions early in treatment became less depressed later in treatment. Also contrary to expectation, early change in frequency of pleasant events was associated with later change in depression in the CT treatment but not in the BA treatment.

We also examined the correlations between early residual change in depression and late residual change in cognitive and behavioral mechanisms. Early change in depression was not sig-

Table 5
Percentage of Participants Relapsed at 6-Month Follow-Up

Status	% (and *n*)		
	BA	AT	CT
Fully recovered[a]			
Relapsed	15.0 (6)	7.7 (2)	18.9 (7)
No relapse	85.0 (34)	92.3 (24)	81.1 (30)
Recovered (BDI < 9)[b]			
Relapsed	20.8 (5)	5.0 (1)	28.0 (7)
No relapse	79.2 (19)	95.0 (19)	72.0 (18)
Recovered (HRSD < 8)[c]			
Relapsed	16.7 (4)	13.6 (3)	22.2 (6)
No relapse	83.3 (20)	86.4 (19)	77.8 (21)

Note. *Fully recovered* included those participants who were either symptom-free or minimally symptomatic for at least 8 consecutive weeks on the Longitudinal Interval Follow-Up Evaluation II interview. *Recovered (BDI < 9)* included those participants who were either symptom-free or minimally symptomatic for at least 2 weeks directly before posttest and who also had scores of less than 9 on the BDI at posttest. *Recovered (HRSD < 8)* included those participants who were either symptom-free or minimally symptomatic for at least 2 weeks directly before posttest and who also had scores of less than 8 on the Hamilton Rating Scale for Depression (HRSD). BA = behavioral activation; AT = automatic thoughts; CT = cognitive–behavioral therapy.
[a] $\chi^2(2, N = 103) = 2.27$, *ns*. [b] $\chi^2(2, N = 69) = 3.94$, *ns*. [c] $\chi^2(2, N = 73) < 1$, *ns*.

Table 6
Mean Number of Well Weeks During 6-Month Follow-Up by Condition

Status	M (and SD) BA	AT	CT	F(dfs) and p
Fully recovered	22.2 (5.2)	20.5 (7.3)	19.8 (7.8)	$F(2, 97) = 1.2$, *ns*
Recovered (BDI < 9)	22.1 (5.6)	20.6 (7.8)	18.2 (9.6)	$F(2, 69) = 1.6$, *ns*
Recovered (HRSD < 8)	23.3 (3.4)	19.4 (8.7)	20.4 (9.4)	$F(2, 73) = 2.5$, *ns*

Note. *Fully recovered* included those participants who were either symptom-free or minimally symptomatic for at least 8 consecutive weeks on the Longitudinal Interval Follow-Up Evaluation II (LIFE) interview. *Recovered (BDI < 9)* included those participants who were either symptom-free or minimally symptomatic for at least 2 weeks directly before posttest and who also had scores of less than 9 on the Beck Depression Inventory (BDI) at posttest. *Recovered (HRSD < 8)* included those participants who were either symptom-free or minimally symptomatic for at least 2 weeks directly before posttest and who also had scores of less than 8 on the Hamilton Rating Scale for Depression (HRSD). Well weeks are defined as weeks in which a 1 (*no depressive symptoms*) or a 2 (*minimally symptomatic*) was coded on the LIFE interview. BA = behavioral activation; AT = automatic thoughts; CT = cognitive–behavioral therapy.

nificantly related to later change in the EASQ or the PES in either of the treatment conditions.

Discussion

We found no evidence in this study that CT is any more effective than either of its components. When one examines the means and standard deviations on our outcome measures, the null findings are unlikely to be attributable to inadequate power. The outcomes were quite comparable across treatment conditions and across outcome measures. Given the fact that our criteria for recovery were more stringent than in many previous studies, it is hard to compare the outcomes of this and other studies. However, our recovery rates were comparable with those of the TDCRP; despite a more severely depressed sample in this treatment study than in the TDCRP (as evidenced by higher mean BDI scores), the magnitude of change for participants in this study was comparable with those of previous CT studies.

Table 7
Correlations Between Early Mechanism Change and Late Depression Change in Each Treatment

Mechanism measure	BA	CT
EASQ		
Uncontrollable	−.01	.21
Internal	.27	.14
Stable	.45**[a]	.03
Global	.38*	.22
PES		
Frequency	.17	−.29*
Pleasure	−.26	−.25
DAS	.26	−.02

Note. BA = behavioral activation; CT = cognitive–behavioral therapy; EASQ = Expanded Attributional Style Questionnaire; PES = Pleasant Events Schedule; DAS = Dysfunctional Attitudes Scale.
[a] Probability levels are one-tailed where the relationship was predicted and two-tailed where it was unexpected.
* $p < .05$. ** $p < .01$.

The finding that BA alone is equal in efficacy to more complete versions of CT is important for both the theory and treatment of depression. We have ruled out threats to the internal validity of this study, and to the results given earlier, suggesting that these are valid findings: Our competence ratings showed that the therapists were performing CT within the range typically viewed by experts as competent; also, the absence of superiority for CT is not accounted for by unwanted overlap between treatments. The adherence ratings suggest that the treatments were quite discriminable and that the therapists did an excellent job of sticking to the treatment protocols. Thus, despite the fact that the treatments were distinct, the outcomes were indistinguishable, at least in the short term.

Furthermore, the treatments were not significantly different at follow-up. The parametric analyses included the entire sample, thus preserving random assignment. With these analyses, there were no overall differences between groups at the time of the 6-month follow-up, and groups did not change differentially during the follow-up period. All groups maintained their treatment gains for the most part during the short follow-up period. When relapse rates were examined, either parametrically in terms of the number of well weeks or nonparametrically in terms of the proportion of participants who had relapsed, CT once again failed to outperform component treatments.

Thus, participants with depression who received BA alone did as well as those who were additionally taught coping skills to counter depressive thinking. Furthermore, both component groups improved as much as those who received interventions aimed at modifying cognitive structures, specifically underlying assumptions, and core schema. These findings run contrary to hypotheses generated by the cognitive model of depression put forth by Beck and his associates (1979), who proposed that direct efforts aimed at modifying negative schema are necessary to maximize treatment outcome and prevent relapse. These results are all the more surprising, given that they run counter to the allegiance effect (Robinson, Berman, & Neimeyer, 1990), which is quite commonly related to outcome in psychotherapy research. All of the therapists expected CT to be the most effective treatment, and morale was low whenever a case was assigned to BA. Moreover,

Keith S. Dobson, one of the clinical supervisors in the TDCRP, expected CT to outperform the alternative treatments. In short, although the null hypothesis can never be accepted, especially in response to one study with negative findings, the distinctiveness of the treatments as well as the allegiance of the therapists and supervisor make the absence of a treatment effect more convincing than would otherwise be the case.

These results raise questions as to the theory of change put forth in the CT book by Beck and his associates. They also raise questions as to the necessary and sufficient conditions for change in CT. These questions are more pronounced in light of the failure to find evidence that the mechanisms addressed by the various treatments were associated with differential change in the targeted mechanisms. In fact, our analyses of moderator effects yielded the counterintuitive finding that changes in attributional style were most inclined to be followed by decreased depression in BA, not in CT, as one would expect given the cognitive theory of change. It seems as if clients who responded positively to activation were also those who altered their predictions regarding how they would respond to negative life events that might occur. Because this was not a predicted finding, it should be interpreted with caution. Nevertheless, if measures of attributional style are thought of as predictions regarding hypothetical future encounters rather than measures of cognitive structure, it may be that patients with depression who respond positively to activation instructions are also those who make more optimistic predictions once they are provided with interventions designed to place them in touch with potential sources of positive reinforcement. Of course, it is also possible that BA-focused treatments are more effective ways of changing the way people think than treatments that explicitly attempt to alter thinking. Perhaps the exposure to naturally reinforcing contingencies produces changes in thinking more effectively than the explicitly cognitive interventions do.

If BA and AT treatments are as effective as CT and also are as likely to modify the factors that are thought to be necessary for change to occur, then not only the theory but also the therapy may be in need of revision. Both BA and AT are more parsimonious treatments than CT and might be more accessible to less experienced or paraprofessional therapists. Because the intervention choices are fewer and more straightforward, these component treatments may also be more amenable to less costly alternatives to psychotherapy, such as self-administered or peer support treatments (cf. Christensen & Jacobson, 1993).

Many questions need to be answered before one can draw negative conclusions about the theory of change put forth by Beck et al. (1979). For one thing, it may be that CT will prove to be effective in preventing recurrence relative to the component treatments. If that proves to be the case, we have shown that the schema modification component of CT has a prophylactic effect, although it may not facilitate acute treatment response. As our 12-month, 18-month, and 2-year follow-up data come in, we will be able to compare the treatments in terms of their relapse–recurrence prevention.

Finally, we acknowledge current limitations in our ability to measure the constructs that were targeted for intervention by the three treatment conditions. It could be that the absence of an association between treatment condition and target mechanism has more to do with the inadequacy of currently available measuring instruments than with the absence of differential change mechanisms. This concern is especially acute for measures of negative schema, in which paper-and-pencil measures have been criticized. We recognize the limitations of these methods and acknowledge that if proper measures existed, the association between mechanism and treatment condition might indeed be stronger.

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

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

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100.

Christensen, A., & Jacobson, N. S. (1993). Who or what can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5,* 8–14.

DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14,* 469–482.

Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 57,* 414–419.

Edwards, B. C., Lambert, M. J., Moran, P. W., McCully, T., Smith, K. C., & Ellingson, A. G. (1984). A meta-analytic comparison of the Beck Depression Inventory and the Hamilton Rating Scale for Depression as measures of treatment outcome. *British Journal of Clinical Psychology, 23,* 93–99.

Elkin, I., Shea, T., Watkins, J. T., Imber, S. C., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Fiester, S. J., Docherty, J., & Parloff, M. B. (1989). NIMH Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry, 46,* 971–982.

Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., Rush, A. J., & Weissman, M. M. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry, 48,* 851–855.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6,* 276–296.

Hollon, S. D., DeRubeis, R. J., & Evans, M. D. (1987). Causal mediation of change in treatment for depression: Discriminating between nonspecificity and noncausality. *Psychological Bulletin, 102,* 139–149.

Hollon, S. D., Evans, M. D., Elkin, I., & Lowery, A. (1984, August). *System for rating therapies for depression.* Paper presented at the 92nd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an Automatic Thoughts Questionnaire. *Cognitive Therapy and Research, 4,* 383–395.

Hollon, S. D., Shelton, R. C., & Loosen, P. T. (1991). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting and Clinical Psychology, 59,* 88–99.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336–352.

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

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreason, N. C. (1987). The longitudinal interval follow-up evaluation: A comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry, 44,* 540–548.

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

MacPhillamy, D. J., & Lewinsohn, P. M. (1971). *The Pleasant Events Schedule.* Unpublished manuscript, University of Oregon.

Peterson, C., & Villanova, P. (1988). An expanded attributional style questionnaire. *Journal of Abnormal Psychology, 97,* 87–89.

Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin, 108,* 30–49.

Rush, A., Beck, A., Kovacs, M., & Hollon, S. (1977). Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research, 1,* 17–37.

Shaw, B. F. (1977). Comparisons of cognitive therapy and behavior therapy in the treatment of depression. *Journal of Consulting and Clinical Psychology, 45,* 543–554.

Shea, M. G., Elkin, I., Imber, S. D., Sotsky, S. M., Watkins, J. T., Collins, J. F., Pilkonis, P. A., Beckham, E., Glass, D. R., Dolan, R. T., & Parloff, M. B. (1992). Course of depressive symptoms over follow-up: Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry, 49,* 782–787.

Spitzer, R. L., Williams, J. B., & Gibbon, M. (1987). *Instruction manual for the Structured Clinical Interview for DSM–III–R.* (Available from the Biometrics Research Department, New York State Psychiatric Institute, 722 W. 168 St., New York, NY 10032).

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620–630.

Whisman, M. A., Strosahl, K., Fruzzetti, A. E., Schmaling, K. B., Jacobson, N. S., & Miller, D. M. (1989). A structured interview version of the Hamilton Rating Scale for Depression: Reliability and validity. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 238–241.

Received January 5, 1995
Revision received February 25, 1995
Accepted March 1, 1995 ■