Traditional Versus Integrative Behavioral Couple Therapy for Significantly and Chronically Distressed Married Couples

Andrew Christensen
University of California, Los Angeles

David C. Atkins, Sara Berns, and Jennifer Wheeler
University of Washington

Donald H. Baucom
University of North Carolina, Chapel Hill

Lorelei E. Simpson
University of California, Los Angeles

A randomized clinical trial compared the effects of traditional behavioral couple therapy (TBCT) and integrative behavioral couple therapy (IBCT) on 134 seriously and chronically distressed married couples, stratified into moderately and severely distressed groups. Couples in IBCT made steady improvements in satisfaction throughout the course of treatment, whereas TBCT couples improved more quickly than IBCT couples early in treatment but then, in contrast to the IBCT group, plateaued later in treatment. Both treatments produced similar levels of clinically significant improvement by the end of treatment (71% of IBCT couples and 59% of TBCT couples were reliably improved or recovered on the Dyadic Adjustment Scale; G. B. Spanier, 1976). Measures of communication also showed improvement for both groups. Measures of individual functioning improved as marital satisfaction improved.

Over the past 30 years, dozens of clinical trials have demonstrated the efficacy of couple therapy for improving relationship satisfaction (see Baucom, Shoham, Mueser, Daiuto, & Stickle, 1998; Christensen & Heavey, 1999, for recent reviews). The results were summarized well by Jacobson and Addis (1993): "in no published study has a tested model failed to outperform a control group. In virtually every instance in which a bona fide treatment has been tested against a control group, the treatment has shown reliable change" (p. 85). Not only does couple therapy improve relationship satisfaction, couple therapy also has been shown to be efficacious as an adjunctive treatment and as a treatment in its own right for several *Diagnostic and Statistical Manual of Mental Disorders* conditions, such as anxiety disorders, depression, alcoholism, and sexual disorders (Baucom et al., 1998).

There are several different types of couple therapies that are empirically supported. Nine studies have provided empirical support for emotionally focused couple therapy (Johnson, Hunsley, Greenberg, & Schindler, 1999) and Baucom et al. (1998) designated this treatment as "efficacious and possibly specific." They designated several additional therapies with more limited empirical support, such as cognitive and cognitive–behavioral couple therapies, as "possibly efficacious" treatments. However, behavioral marital therapy, or what we call traditional behavioral couple therapy (TBCT),[1] is the only couple treatment that meets the highest criterion of empirical support, that of an "efficacious and specific treatment" (Baucom et al., 1998; Chambless & Hollon, 1998). This designation means that it has been shown to be more effective than no treatment and to an alternative treatment (a placebo intervention or a bona fide therapy) in at least two independent research settings. Indeed, far more studies have been conducted on TBCT than on any other couple therapy.

Although many practitioners use the term *behavioral* to describe their work, our definition of TBCT is based on the procedures used in clinical trials of behavioral approaches, particularly the treatment manual of Jacobson and Margolin (1979), the most commonly used behavioral treatment manual. The focus in TBCT is on making positive changes in each partner's behavior so that they provide one another with less punishing and more rewarding interactions. The therapeutic strategy is directive and prescriptive (or "rule governed"; Skinner, 1966). TBCT therapists first assess the couple's problems and strengths, provide feedback to them about the assessment, and discuss the goals and procedures of treatment. Therapists then assist couples in identifying positive

[1] The change from "marital therapy" to "couple therapy" reflects a broadening of the emphasis from heterosexual married couples to all romantically involved couples. The addition of "traditional" is intended to distinguish this treatment from other behavioral approaches.

actions they could take toward one another and direct them to engage in these positive behaviors. Through instruction, modeling, behavior rehearsal, and feedback, TBCT therapists teach partners how to communicate and problem solve around their difficulties. As couples make positive changes in their relationship, TBCT therapists becomes less directive, with the expectation that couples will begin to generalize skills learned in therapy to their daily lives.

The fact that TBCT has garnered the most research and the highest classification of empirical support says little about its power to effect change. For that, one must look at effect sizes and rates of clinically significant change. In a meta-analysis of 17 TBCT studies, Hahlweg and Markman (1988) found an overall effect size of .95 for TBCT, which is generally considered a large effect size and certainly comparable to effect sizes found for individual therapy. However, rates of clinically significant change have been disappointing. For example, in an examination of 4 studies of TBCT that included a total sample of 148 participants, Jacobson et al. (1984) found that 54.7% of couples showed reliable improvement immediately after treatment. However, only 35.3% of couples in the 3 clinical studies were recovered (reliable improvement and movement into the nondistressed range). Only 16.7% were recovered in an analogue study.

Unfortunately, not only is the immediate success of TBCT disappointing, but there is also evidence of deterioration over the long term. Studies have generally indicated maintenance of treatment gains from 3 to 6 months following treatment with mixed results for up to 1 year after treatment (Hahlweg & Markman, 1988; Jacobson et al., 1984). However, the only two studies in the literature that have looked beyond a year after treatment find substantial deterioration. Jacobson, Schmaling, and Holtzworth-Munroe (1987) assessed 12 couples 2 years after TBCT and found that 25% had deteriorated below their pretest functioning and 9% had separated or divorced. In perhaps the most comprehensive investigation to date of marital therapy, Snyder and his colleagues (Snyder & Wills, 1989; Snyder, Wills, & Grady-Fletcher, 1991) found that couples in TBCT were superior to the control condition at termination and substantially maintained their improvement at 6-month follow-up. However, at the next assessment at 4 years posttreatment, 38% of couples in TBCT were divorced. One possible explanation for these dramatic findings is that the changes induced by behavioral marital therapy could not be maintained by the couple and led to alienation and breakup.

Because of these limitations of TBCT, Christensen and Jacobson developed an alternative treatment program, integrative behavioral couple therapy (IBCT; Christensen, Jacobson, & Babcock, 1995; Jacobson & Christensen, 1998), that builds on TBCT but includes an emphasis on emotional acceptance. IBCT assumes that there are genuine incompatibilities in all couples that are not amenable to change, that partners' emotional reactions to each other's behavior are at least as problematic as the behavior itself, and that a focus on change can often lead to a resistance to change. Therefore, emotional acceptance between partners is as much or more a goal of intervention as is active change in the partner's behavior. Not only is the therapeutic focus of IBCT different from that of TBCT, the treatment strategy is also different. Rather than having a primary reliance on prescriptive, rule-governed changes, IBCT emphasizes nondirective, "contingency-shaped" changes (Skinner, 1966). Through a series of strategies that are described later, IBCT therapists try to alter the context in the therapy session, moving the partners from adversarial confrontation to collaborative engagement.

There have been three small empirical studies of IBCT. Wimberly (1998) demonstrated that 8 couples randomly assigned to a group format of IBCT were significantly more satisfied than 9 wait-list couples at the end of therapy. In an unpublished study of 29 depressed women who were maritally distressed, Trapp, Pace, and Stoltenberg (1997, cited in Christensen & Heavey, 1999) showed that IBCT was as effective in reducing depression as cognitive therapy for depression. Finally, in a clinical trial of 21 couples that served as a basis for the current clinical trial, Jacobson, Christensen, Prince, Cordova, and Eldridge (2000) demonstrated that TBCT and IBCT could be distinguished when delivered by the same therapists according to independently coded measures of adherence. Because of the small sample size of this preliminary study, no statistical tests were conducted comparing the two treatments. However, the effect size data and clinical significance data favored IBCT. In combination, these three studies suggest that IBCT is a different therapy than TBCT and is a viable treatment for marital discord.

The primary purpose of the current study is to examine the overall and comparative efficacy of TBCT versus IBCT. To increase the chances of finding a difference between these treatments, we used the largest sample of couples in a randomized clinical trial: 134 couples total, with 66 in IBCT and 68 in TBCT. This contrasts sharply with the average of 10 couples per treatment condition that Shadish et al. (1993) found in their review of marital and family treatment studies. It is also greater than the 25–30 participants per condition that Chambless and Hollon (1998) recommended to show "a reasonably stable estimate of the effects of treatment" (p. 9).

In addition, we selected only seriously and stably distressed couples. Previous research has found that the greater the relationship distress, the poorer the outcome in treatment (Halford, 2001; Jacobson & Addis; 1993; Snyder, Mangrum, & Wills, 1993). We reasoned that treatments would show their relative power most clearly when faced with difficult cases. To be included in the study, couples had to *repeatedly* report substantial relationship distress. We wanted to exclude not only those couples who were mildly distressed but also those couples whose scores in the moderately to severely distressed range were unstable (e.g., a couple who was especially distressed after a recent argument but who would soon return to a higher and more typical level of satisfaction). To examine the impact of severity on outcome, we stratified these couples into severe and moderate levels of distress and randomly assigned them to treatment within these levels.

Because most outcome studies of couple therapy assess couples prior to treatment and at the end of treatment, little information is available on change during the course of couple therapy. We do not know, for example, whether improvement occurs gradually and steadily over the course of treatment or whether there is a more variable course, with perhaps surges early or late in treatment. Halford (2001) has argued that change during TBCT often takes place early in treatment, soon after assessment, feedback, and goal setting. Some of his studies have included intrasubject analyses and have shown that change occurs early in treatment with little or no change later in treatment (e.g., Kelly & Halford, 1995). On the basis of his work, we might predict that change in both types of couple therapy, but particularly TBCT, would come early in treat-

ment with lesser change later in treatment. In the current study, we examine the impact of treatment both during and after treatment delivery.

Not only do most studies assess participants only at pre- and at posttreatment, they examine average dyad scores rather than individual partner or spouse scores (Dunn & Schwebel, 1995). Therefore, we know little about the relative impact of treatment on men versus women. However, we do know that women are more likely to seek couple therapy than men (Doss, Atkins, & Christensen, 2003) and that men and women show many other differences on relationship variables (Eldridge & Christensen, 2002). Furthermore, in a reexamination of the data from the outcome study by Snyder and his colleagues, Kashy and Snyder (1995) found that behavioral couple therapy was more effective for husbands but insight-oriented couple therapy was more effective for wives. Therefore, in the current study, we examine differential treatment effects for husbands and wives.

Couples typically seek treatment to improve their happiness in the relationship and to prevent separation and divorce. Similarly, the most common outcome variables in research on couple therapy, and possibly in research on couples in general, are satisfaction and stability. Couple therapies, particularly behavioral therapies, try to achieve these goals by improving communication, in part because communication is a common presenting complaint. Therefore, another common relationship outcome measure is communication. Some studies also assess individual functioning because improvements in relationship satisfaction may also stimulate general adjustment. In the current study, we examine all of these outcome measures, including relationship satisfaction, stability, communication, and individual adjustment.

On the basis of these considerations from the previous literature, the current study has the following five hypotheses: (a) Both TBCT and IBCT will lead to improvement in relationship and individual outcomes; (b) these treatments will have their greatest impact early in treatment with lesser impact later in treatment; (c) IBCT will have a greater impact than TBCT on relationship and individual outcomes; (d) these treatments will show the greatest impact on moderately rather than severely distressed couples; and (e) husbands and wives will respond differently to couple treatment, with husbands possibly benefiting more from TBCT than IBCT.

Method

Participants

One hundred thirty-four seriously and chronically distressed married couples were recruited for a therapy program in Los Angeles (71 couples) and Seattle (63 couples). Newspaper advertisements, radio and TV announcements, and letters and brochures sent to clinics and practitioners described a free therapy program for unhappy couples who wished to improve their relationship.

The first 18 couples in Los Angeles and the first 8 couples in Seattle were part of a pilot study to train therapists, to evaluate assessment procedures, and to determine an optimal cutpoint for separating moderately versus severely distressed groups of couples. Later, we describe procedures for the study proper and note in a special section how a few procedures for the pilot phase were different.

To be included in the study, all couples had to be legally married and living together, had to request couple therapy, and had to meet criteria for

serious and stable marital distress. These distress criteria (described later) are based on three measures of marital satisfaction that were completed at three different time points prior to random assignment and treatment. As evidence of the chronicity of the marital distress of these couples, approximately half reported previously attending marital therapy with their current spouse.[2]

To be included in the study, both partners had to have a high school education or its equivalent, both had to be between the ages of 18 and 65, and both had to be fluent in English. Each spouse was given a diagnostic interview, but only diagnoses that might directly interfere with treatment were excluded: current Axis I disorders of schizophrenia, bipolar disorder, or alcohol/drug abuse or dependence or current Axis II disorders of borderline, schizotypal, or antisocial personality disorder. To avoid confounding therapy results with the presence of alternative treatments, neither partner could be in psychotherapy or, if he or she was, treatment needed to be stopped for the duration of the marital therapy. Partners could be on psychotropic medication if they had been taking the medication for a minimum of 12 weeks, were on at stable dose for a minimum of 6 weeks prior to the pretreatment assessment, and their physician, when contacted by the project, did not anticipant changing medication or dosage. To ensure that our sample did not include battering men, wife reports of violence were used to eliminate couples in which husbands had engaged in dangerous levels of violence

The mean age of wives was 41.62 years ($SD = 8.59$), and the mean age of husbands was 43.49 years ($SD = 8.74$). The mean number of years of education (counting kindergarten) was 16.97 ($SD = 3.23$) for wives and 17.03 ($SD = 3.17$) for husbands. Couples had been married a mean of 10.00 years ($SD = 7.60$) and had an average of 1.10 ($SD = 1.03$) children. Most of the participants were Caucasian (husbands: 79.1%, wives: 76.1%). Other ethnicities include African American (husbands: 6.7%, wives: 8.2%), Asian or Pacific Islander (husbands: 6.0%, wives: 4.5%), Latino or Latina (husbands: 5.2%, wives: 5.2%), and Native American or Alaskan Native (husbands: 0.7%).

There were no treatment group, distress level, or site differences for age, education, income, years married, or number of children. Pilot and study proper participants were equivalent on these variables. Although there were no treatment group differences for participants' ethnic identity, there were site differences. In particular, the ethnic composition of the husband sample was not significantly different between the two sites, but the composition of the wife sample did vary by site, $\chi^2(4, N = 134) = 15.43$, $p < .01$. Wives at the University of Washington (UW) site were more likely to be Caucasian and less likely to be from a minority group than wives at the University of California, Los Angeles (UCLA) site (Caucasian wives: $n = 57$ at UW, $n = 45$ at UCLA).

Screening Measures

All couples participated in a three-stage screening process that included (a) a phone interview to assess basic demographic eligibility and marital satisfaction, (b) a mailed packet of questionnaires to assess marital satisfaction and domestic violence, and (c) an in-person intake evaluation to assess marital satisfaction and conduct individual psychiatric interviews. The length of time between the initial phone interview and the intake evaluation averaged 6 weeks because of time taken by couples to return mailed questionnaires and scheduling difficulties.

---

[2] After the first 35 cases, we began asking couples whether they had had previous marital therapy. Of these 99 cases, in 45 couples both partners agreed they had had previous marital therapy, 15 partners reported marital therapy that was not corroborated by the other (perhaps because it was not conjoint therapy), and 39 couples agreed that they had not had previous marital therapy. Thus, a majority of those assessed in our sample had tried marital therapy before, but it had not been successful in leaving them in an enduring state of satisfaction.

Marital Adjustment Test (MAT; Locke & Wallace, 1959). The MAT is a commonly used, well-validated self-report measure of marital satisfaction. Potential participants were given a phone version of the MAT at Stage 1 of the screening process. If partners scored an average of less than 100, a common cutoff for marital distress, then they were eligible to advance to the next stage of screening.

Marital Satisfaction Inventory—Revised (MSI–R; Snyder, 1997). The MSI–R is a well-normed questionnaire that provides 2 validity scales, 1 global distress scale, and 10 scales assessing specific domains of marriage. The Global Distress Scale (GDS) of the MSI–R is a measure of overall dissatisfaction in the relationships and served as both a screening measure and an outcome measure in this research. As part of the Stage 2 screening, partners completed the full MSI–R. At the Stage 3 screening (the intake session), partners completed only the GDS of the MSI–R as a final verification of the stability of their distress. To be eligible for the study, at least one partner had to attain a *T* score of 59 or higher on the GDS at both Stage 2 and Stage 3 screenings.

Dyadic Adjustment Scale (DAS; Spanier, 1976). Along with the MAT and the GDS, the DAS is a widely used self-report measure of marital satisfaction and perhaps the most widely used measure of couple treatment outcome. At Stage 3 of screening (the intake assessment), at least one spouse had to score at least one standard deviation below the population mean (< 98) for the couple to be included in the study.

Conflict Tactics Scale—Revised (CTS–2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996). Respondents reported verbal, sexual, and physical aggression and physical injury that they had both inflicted and received from their spouses on the CTS–2. On the basis of previously developed criteria for defining male battering (Jacobson & Gottman, 1998; Jacobson et al., 1994), we excluded couples with moderate to severe husband-to-wife domestic violence from our study and referred them for specific, violence-related individual treatment.

Structured Clinical Interview for DSM–IV (SCID). Specially trained graduate students conducted SCID interviews for Axis I (First, Spitzer, Gibbon, & Williams, 1994) and Axis II (Spitzer, Williams, Gibbon, & First, 1994) disorders and audio- or videotaped the interviews for interrater reliability analyses. These graduate students assessed for both present and past Axis I disorders, except for past substance abuse and dependence. To calculate the reliability of these interviews, we randomly selected 41 tapes, approximately 15% of these SCID interviews, ensuring that every interviewer had at least one tape in this sample. Each of these taped interviews was rated by an interviewer at the alternative site. Raters reached 88% overall agreement on the presence or absence of a current diagnosis ($\kappa = .66$, $p < .001$) and 85% agreement on the specific diagnosis ($\kappa = .57$, $p < .001$). Raters reached 90% overall agreement on the presence or absence of a past diagnosis ($\kappa = .75$, $p < .001$) and 90% agreement on the specific diagnosis ($\kappa = .72$, $p < .001$).

As a result of these screening criteria, 94 couples were excluded because they did not meet our criteria of marital distress (not sufficiently distressed); 101 couples were excluded because the husbands were too violent; and 3 were excluded because they met criteria for exclusionary diagnoses.

Outcome Measures

Our outcome measures assessed relationship satisfaction, relationship stability, communication, spouses' individual functioning, and client reactions to treatment. At the intake assessment, at 13 weeks after intake, and at 26 weeks after intake couples came to UW or UCLA for an evaluation session and completed all of the outcome measures described later except for the client reaction measures. At the end of the fourth therapy session, clients completed the Short Therapeutic Bond measure (described in the *Therapeutic bond* section) and sent it by mail to the project office. At the end of the final treatment session, the therapist gave each couple the relationship satisfaction and client evaluation of services measures. The couple completed these measures and mailed them to the project.

Relationship satisfaction. The DAS and the GDS, discussed previously, were the primary methods of assessing change in relationship satisfaction. Each was administered at all four assessment sessions: intake, 13 week, 26 week, and final session.

Relationship stability. The Marital Status Inventory (MSI; Weiss & Cerreto, 1980) consists of 14 true–false items that measure steps toward separation/divorce, ranging from thoughts (e.g., thinking of separation/divorce after an argument), to tentative steps (e.g., talking to a friend), to actual separation/divorce actions (e.g., moving out). Scores range from 0 to 14 depending on the number of steps the respondent has taken toward divorce. Research has shown that the MSI can identify couples at risk of divorce (Crane, Newfield, & Armstrong, 1984). The MSI was administered at the intake and at 13-week and 26-week assessments. At intake couples completed the questionnaire for their entire relationship. At 13 and 26 weeks, couples completed the measure with regard to the time period since the last assessment.

Communication. Two subscales of the MSI–R were used to assess the couples' self-report of their communication: problem solving communication (PSC) and affective communication (AFC). The 19 true–false items on the PSC reflect three domains: difficulty resolving minor differences, lack of problem-solving skills, and inability to discuss sensitive issues. The 13 true–false items on the AFC reflect two dimensions: lack of support and affection and limited disclosure of feelings or lack of understanding. Snyder (1997) describes the extensive validation conducted on the MSI–R and its subscales.

Individual functioning. The Compass Outpatient Treatment Assessment System (Sperry, Brill, Howard, & Grissom, 1996) includes three self-report scales that evaluate patient functioning: Subjective Well-Being, Current Symptoms, and Current Life Functioning. The Mental Health Index (MHI) is the combination of these three scales converted into a *T* score. The MHI has been evaluated on thousands of adult outpatients, has an internal consistency of .87, and a 3–4-week test–retest stability of .82. The scale is scored such that a higher score represents a more adaptive status. In comparisons between a community and patient sample, *T* scores of 60 or below characterized the patient sample. In addition to the MHI, we also examined the Current Symptoms (CS) subscale. Like the MHI, the CS has been evaluated on thousands of adult outpatients, has an internal consistency of .94, and a 3–4-week test–retest stability of .85. Currently it is scored such that higher *T* scores indicate more symptoms. In comparisons between an outpatient and community sample, a *T* score of 40 or above characterized the patient sample. All *T* scores were calculated on the basis of recent normative data provided by Kenneth Howard and his staff (personal communication, May 28, 1998).

Client Reactions to Treatment

Therapeutic bond. The Short Therapeutic Bond Scale consists of 6 items designed to assess working alliance, empathic resonance, and mutual affirmation (2 items for each). It is based on a 12-item Therapeutic Bond Scale (Sperry et al., 1996), which in turn was based on the original, 50-item Therapeutic Bond Scale (Saunders, Howard, & Orlinsky, 1989). The sum of the 6 items is converted into a *T* score, based on results from a large population of 1958 outpatients. Alpha on this measure from that population was .92.

Client evaluation of services. The Client Evaluation of Services Questionnaire included the eight-item Client Satisfaction Questionnaire (CSQ-8), which is a brief version of a larger Client Satisfaction Questionnaire (Nguyen, Attkisson, & Stegner, 1983). On 4-point scales, clients rate the effectiveness and their satisfaction with the services they received. The scale has an alpha of .93. In a sample of several thousand clients at various psychiatric facilities, the average total score on the CSQ-8 was 27.09 with a standard deviation of 4.01.

To optimize honest reporting on both of these measures, we had couples mail their questionnaires directly to the project, and we told couples that their therapist would not see their answers.

Treatment Procedures

After couples had successfully completed the three-stage screening procedures (the last stage of which was the intake), they were given the name of a project therapist and instructed to schedule their first appointment. After scheduling their first appointment with the therapist, couples were randomly assigned to one of the two treatment conditions. A total of 68 couples were assigned to TBCT and 66 to IBCT.

Criteria for the stratification into moderately and severely distressed groups were obtained from the pilot portion of the study on the basis of average scores of husband and wife on the DAS and GDS. The average DAS scores were first converted to a *T* score, and then the GDS and the DAS were averaged so that higher scores would indicate greater distress. A median split of the pilot sample created the criteria (*T* = 66) for the two groups. Mean scores for the moderately distressed couples were 62.7, whereas the mean scores for the severely distressed couples were 70.6. Snyder (1997) found that couples with an intake GDS *T* score of 66 or higher who received treatment had a 45% chance of eventually divorcing, suggesting that our severely distressed couples are at high risk for divorce, even with treatment.

During the study-proper phase of the study, couples were randomly assigned to TBCT or IBCT within the two strata of moderately or severely distressed couples. Combining pilot and study-proper couples, there were 66 moderately distressed couples and 68 severely distressed couples. A chi-square analysis of the number of moderately and severely distressed couples in TBCT and IBCT was nonsignificant. An analysis of the combined DAS and GDS *T* scores indicated, as expected, a major significant effect of stratification (*p* < .001) but no significant effect of therapy condition and no significant interaction of Therapy Condition × Stratification.

In both treatment conditions, couples completed a four-session sequence of evaluation and feedback. In the first session, attended by both husband and wife, the therapist inquired about the presenting problems and obtained a brief relationship history of the couple. The next two sessions were individual contacts with the husband and wife separately, in no predetermined order. In these sessions, the therapist obtained more information about the presenting problems and obtained an individual history from each partner. Finally, in the fourth session, the therapist obtained any additional needed information and provided the couple with feedback, appropriate to their treatment condition, about their problems and about their upcoming treatment. The remaining sessions were devoted to treatment procedures and, toward the end, preparing for termination. Because of the level of distress, both treatments were designed to allow for a relatively lengthy course of treatment, if needed. Couples were told that they could have a maximum of 26 sessions but could end earlier if they felt their problems were sufficiently resolved. These sessions would normally occur once a week but had to be completed within a year.

The mean number of sessions was 22.9 (*SD* = 5.35; median and mode were 26), and the median length of time to complete these sessions was 36 weeks. Mean number of sessions was not significantly different in IBCT (*M* = 23.5, *SD* = 4.7) than in TBCT (*M* = 21.7, *SD* = 6.9), nor were they significantly different in moderately (*M* = 22.4, *SD* = 6.7) and severely distressed (*M* = 22.7, *SD* = 5.3) conditions. In previous research on couple therapy, 10 has been the mean and modal number of sessions (Vu & Christensen, 2003). Therefore, given that some couples ended their treatment prior to the maximum number of sessions we allowed and given our desire to include as a treatment completer all couples who had completed a substantial number of sessions, we used 10 or more sessions as our definition of "treatment completer." Of the 134 couples, 126 (94%) were considered "completers." Of the 8 couples who did not complete therapy, 1 was in IBCT and 7 were in TBCT (Fisher's exact test probability = .062).

Couples were recruited sequentially in five cohorts of about 12 couples at each site (with more couples in the last cohort) between November 1997 and February 2001. The first cohort consisted of the pilot cases.

TBCT. The goals of TBCT are to promote positive change in couples through direct instruction and skill training. During the feedback session, the therapist emphasizes the strengths of the couple and delineates specific problem areas that could be the target for later communication and problem-solving efforts. During treatment the therapist relies on three primary treatment strategies: behavioral exchange, communication training, and problem-solving training. In behavioral exchange, the therapist structures direct efforts to increase mutual, positive behavior exchange. For example, the therapist might engage each spouse in generating a list of specific, positive, noncontroversial behaviors that they could do for the partner. This list might then be enhanced with input from the partner. When both partners have a list of positive actions, the therapist might encourage each spouse to perform activities from the list in an effort to increase mutual positive reinforcement. In communication-skills training, the therapist teaches partners both speaking and listening skills. Speaking skills might include a focus on "I" statements and teaching partners to specify their emotions and behavior (for example, learning to say "I feel disappointed when you come home late without calling" vs. "You are so selfish and inconsiderate"). Listening skills emphasize each partner learning to paraphrase or summarize the other's message. In training problem-solving skills, the therapist teaches couples how to define problems, to generate positive alternatives to current problem behavior, to evaluate the pros and cons of those alternatives, to negotiate alternatives, and to implement and evaluate planned change.

The treatment manual for TBCT is Jacobson and Margolin's (1979) classic monograph, supplemented by a shorter, updated, more succinct treatment manual (Jacobson & Christensen, 1994). Couples were also given a communication guide by Gottman, Notarius, Markman, and Gonso (1977), which they read, in part, during the communication-training segment of the therapy.

IBCT. IBCT was designed to enhance TBCT by adding a focus on emotional acceptance. This focus is based on several assumptions: All close relationships are characterized by some genuine incompatibilities, the reactions to problem behavior are often as problematic as the behavior itself, and direct change efforts are often as much a problem for couples as they are a solution. Therefore, IBCT focuses more on the emotional reactions of partners to the difficulties they encounter in their relationships and less on the active solutions they can take to resolve these difficulties, especially for what seem to be insoluble problems. During the feedback stage, the therapist focuses on broad themes in the struggle between spouses rather than on particular problematic issues. The therapist offers a formulation of the couple's difficulties in terms of the differences between them, the understandable though often ineffective or self-defeating actions that each has taken, and the natural emotional reactions that each experiences. The therapist also describes the realistic strengths the couple has and offers the possibility that, through an examination of their daily emotional hurts and struggles, they may come to a greater understanding and appreciation of each other's emotional reactions and to a greater closeness.

During the treatment phase, the focus of the sessions is usually on a salient incident that recently occurred (e.g., an argument last night), will soon occur (e.g., a trip to her mother's this weekend), or is now occurring (e.g., a spouse feels invalidated by the other's reaction in the session). The therapist uses three major strategies to promote emotional acceptance: empathic joining around the problem, unified detachment from the problem, and building tolerance to some of the responses that the problem can trigger. To facilitate empathic joining around the problem, the therapist attempts to elicit vulnerable feelings from each spouse that may underlie emotional reactions to the problem. The therapist encourages partners to express and elaborate these feelings, and he or she communicates empathy for having these understandable reactions. By taking this stance toward both partners, the therapist may also elicit empathy between the partners for each other. To facilitate unified detachment from the problem, the therapist helps the couple to step back from the problem and take a descriptive rather than evaluative stance toward the problem. The therapist

may engage the couple in an effort to describe the sequence of actions they take during their problematic pattern, to specify the triggers that activate and escalate their emotions, to consider variations of their patterned behavior and what might account for these variations (e.g., a typical struggle over their child was less intense because they had felt close to each other earlier), and to generate a name for their problematic pattern. To facilitate tolerance building, the therapist engages the couple in an analysis of the positive functions as well as the negative functions of their differences and their problematic behavioral patterns. The therapist might encourage the couple to deliberately engage in the problem behavior during the session or at home, so that each partner can become more aware of the pattern and take it less personally. The IBCT therapist is also free to use the direct change efforts of TBCT described previously. Some of the specific strategies of IBCT are similar to strategies described by other approaches, such as those of Wile (2002) and Johnson and Denton (2002).

The treatment manual for IBCT was a book by Jacobson and Christensen (1998), which was supplemented by a chapter by Christensen et al. (1995). Couples were also given a self-help book about IBCT by Christensen and Jacobson (2000) to read during treatment.

Ensuring Treatment Quality and Integrity

Therapist selection, training, and supervision. We sought well-trained, licensed, and experienced therapists practicing in the community. Four doctoral-level clinical psychologists in Los Angeles and three in Seattle, all with between 7 and 15 years experience postlicensure, were selected on the basis of their expertise and reputation.[3] Therapists were trained in the protocol by (a) reading the treatment manuals and (b) attending a workshop by either Christensen or Jacobson. They then began treating cases. The first four cases that each therapist saw were deemed training cases and are part of the pilot sample (described later).

It was important to ensure that therapists received expert supervision in both TBCT and IBCT. Two of the supervisors, Christensen and Jacobson, were experts in TBCT and had published outcome research and reviews on TBCT. Jacobson in particular had a long record of clinical research in TBCT, having written perhaps the its most widely used treatment manual (Jacobson & Margolin, 1979). Christensen and Jacobson had also developed IBCT together and written the treatment manuals for it (Christensen et al., 1995; Jacobson & Christensen, 1998). A third supervisor, Peter Fehrenbach, was a therapist on the initial study of TBCT and IBCT (Jacobson et al., 2000) and was considered an expert on both. When Jacobson died, Don Baucom, an expert on TBCT who had published extensively on behavioral approaches with couples, was brought in to supervise many of the TBCT cases.

Intense supervision of therapists was maintained throughout the project to ensure adherence to treatment protocol and competence in treatment delivery. Throughout the study, therapists sent audio- and/or videotapes of their sessions to the supervisors each week. The supervisors observed the sessions and provided feedback to therapists prior to their next meeting. For the first half of the study, but particularly during the pilot phase, supervisors listened to almost all of every tape from every session and then talked on the telephone with the therapist prior to the next session. During the second half of the project, supervisors often listened to only parts of sessions or skipped occasional entire sessions on the basis of feedback from the therapists. Commentary was provided by e-mail as well as phone. However, even during this phase, the supervisor observed a majority of the therapy and provided commentary between most sessions.

Measures of therapy adherence. Our adherence scale, developed for the preliminary study of TBCT and IBCT (Jacobson et al., 2000), includes eight items reflecting change-oriented interventions and nine items reflecting acceptance-oriented interventions. One case from each of the five cohorts was randomly selected from each of the seven therapists with alternation across cohorts between TBCT and IBCT cases. This procedure yielded 35 cases (26% of all cases). Within each case at least one early, one middle, and one late therapy session was randomly selected from the sessions following the first four sessions (assessment and feedback), generating a total of 115 sessions (16% of the possible sessions). Three senior graduate students independently coded each of these sessions for adherence to treatment protocol.

Our experience in our initial study (Jacobson et al., 2000), as well as in the current study, demonstrated that graduate students cannot remain uninformed as to treatment assignment. Knowing IBCT and TBCT, they can immediately determine from observing the tapes which condition the couple has been assigned. Therefore, we also used undergraduate observers who had no knowledge of IBCT or TBCT as adherence raters. Over the course of the study, 11 undergraduate observers rated the 35 cases described above, but coded two early, two middle, and two late therapy sessions, a total of 208 or 30% of possible sessions that were randomly selected. Prior to rating these tapes, these undergraduate observers were extensively trained in a coding system that was based on the system described but only included four 9-point scales: the extent to which therapists (a) set and followed an agenda, (b) engaged in change oriented strategies, (c) engaged in acceptance based strategies, and (d) assigned and checked homework (Jacobson et al., 2000).

Measures of therapist competence. To ensure that TBCT received a fair test in the current study, an outside consultant who had no other connection with the project, Gayla Margolin, was hired to assess competence. As the coauthor of our treatment manual for TBCT (Jacobson & Margolin, 1979), she was the best available judge of competence. She used the Behavioral Couple Therapy Competence Rating Scale (Jacobson et al., 2000) to evaluate one randomly selected case from each of the seven therapists. This manual describes 10 skills essential to good TBCT (9 are rated on 6-point scales, 1 is rated on a 12-point scale). Competence cannot be judged on isolated sessions; it can only be judged with an understanding of the case. Therefore, for each case, Margolin observed the first four sessions (the first joint session and the two individual sessions, which constituted the evaluation sessions, and the 4th session, which was the feedback session) to get an understanding of the case. Then she observed and rated a randomly selected sequence of three consecutive sessions from the first half of treatment and a randomly selected sequence of three consecutive sessions from the second half of treatment. Her ratings were on consecutive sessions so that she could see the continuity between sessions. Thus, she observed 70 sessions of TBCT but rated only 42 sessions for therapist competence.

Procedures for Pilot Cases

Therapists were trained and procedures were developed on the first 26 couples, who were designated as pilot cases. Each therapist saw two couples in each condition and was supervised by all three supervisors. However, in Los Angeles, four therapists were trained on 18 cases because two couples had to leave treatment early and therefore two cases were added. In Seattle only two of the therapists needed training, which required 8 cases (two therapists had worked on the initial study so did not need training). During the pilot study, we did not have an upper age limit and included one couple over 65. Also, we excluded potential participants for substance dependence but not substance abuse and had one couple in which the husband met criteria for substance abuse. We developed our criteria for moderate and severe distress on the basis of a median split of the combined GDS and DAS scores of pilot couples. Therefore, in the pilot study, classification into severe and moderate levels of distress was done after randomization to treatment conditions. Finally, some assessment procedures were done differently in the pilot study than in the study proper. For example, we did not have a separate GDS for the pilot study but only the

---

[3] At the Seattle site, one master's-level family therapist was selected on the basis of expertise and reputation but left for personal reasons after treating only two cases, one in IBCT and one in TBCT.

GDS from the MSI–R. Also, the DAS was sent out at the Level 2 screening (the mailed packet) rather than at intake. Because of these differences between pilot and study proper cases, we examine whether these differences led to changes in treatment effectiveness.

Results

Treatment Integrity

Adherence. Using our detailed adherence coding system, three graduate students coded 58 sessions of TBCT and 57 sessions of IBCT. On two major summary scores of IBCT interventions and TBCT interventions, alpha reliabilities computed across coders were .93 and .97, respectively. Mann–Whitney's nonparametric tests indicated that, as expected, TBCT interventions were much more likely in TBCT sessions ($p < .001$) and IBCT interventions were much more likely in IBCT sessions ($p < .001$). Using a global coding system, 11 undergraduates rated 101 sessions of TBCT and 107 sessions of IBCT. Alpha reliabilities for the four global ratings were .75 for agenda, .93 for change, .92 for acceptance, and .83 for homework. Mann–Whitney's nonparametric tests indicated that TBCT and IBCT differed in expected ways on each dimension ($p < .001$). In fact, therapists in IBCT engaged in about three times more acceptance strategies than therapists in TBCT, who engaged in about three times more change strategies than therapists in IBCT. Clearly, therapists were performing different treatments, in accord with each treatment protocol, in IBCT and TBCT sessions.

Competence. Across 42 sessions from a total of seven cases, Gayla Margolin's ratings on the Behavioral Couple Therapy Competence Rating Scale averaged 52.1 (or a rating between *good* and *excellent*). Average scores for each case ranged from a low of 33.3 (a score midway between *mediocre* and *good*) to a high of 60.9 (a score just under the maximum possible of 66).

Outcome Data

To examine the impact of our treatments across our repeated assessments (intake, 13-week, 26-week, and final session), we used hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), which has also been referred to as multilevel modeling (Snijders & Bosker, 1999) and mixed-effects modeling (Pinheiro & Bates, 2000). There are several important advantages of HLM over other common approaches to repeated measures data, such as repeated measures analysis of variance and analysis of covariance. Most other approaches assume that all participants are assessed at identical and equally spaced intervals, an assumption that is rarely true. In the current study, participants were not assessed at exactly 13 and 26 weeks after treatment began and the final session assessment varied notably between couples. The maximum likelihood estimation used in HLM allows the data to be unbalanced. Thus, participants need not be assessed at the same time or provide data at every time point. Other statistical approaches often drop participants who have missing data whereas HLM uses all available data and generates unbiased empirical Bayes's estimates for each individual in the analysis. Finally, HLM is uniquely suited for the dependencies in couple data in which spouses' data are often correlated.

HLM proceeds in two broad stages. It first creates trajectories of each individual based on the available data from that individual and describes a linear or curvilinear model of change that fits these data. For the present data, each individual's repeated measures were fit with an intercept, slope, and quadratic. The intercept estimates the initial level of distress of an individual when he or she began treatment. The slope estimates the linear change over time, either constant improvement or deterioration, and the quadratic component captures any acceleration or deceleration in the change that may occur. Then, each of these Level 1 parameters are treated as outcome variables to be explained by predictors at higher levels.[4] To test our first hypothesis, that change occurs over the course of treatment, we examine the slope of the trajectories. To test our second hypothesis, that change slows later in treatment, we examine the quadratic component of the trajectories to see if there is a flattening out of change over time. To test our third hypothesis of differential treatment effects, we examine whether our two treatments have different effects on the slope and quadratic shape of couples' trajectories. To test our fourth hypothesis, that change occurs differently for husbands and wives, we compare the slope and quadratic shape of husband and wife trajectories. To test our fifth hypothesis, that change occurs differently for severely distressed versus moderately distressed couples, we examine whether the trajectories of these two groups of couples demonstrate different slopes and quadratics. In addition to these hypotheses, we compare the trajectories of pilot participants versus study-proper participants (because both are included in the analyses) and of our two research sites. Univariate analyses were conducted with SPSS v11.0.1 (SPSS, 2001); HLM analyses were conducted using the nlme library of functions (Pinheiro & Bates, 2000) in S-Plus 2000 Professional Release 2 (Mathsoft, 1999), and generalized linear mixed model (GLMM) analyses used the glmmPQL function in the MASS library (Venables & Ripley, 1997) in *R* v1.4 (Ihaka & Gentleman, 1996).

Relationship satisfaction. We examined the impact of our treatments on our two measures of marital satisfaction, the DAS and the GDS. Because previous research by Whisman and Jacobson (1992) has shown that the DAS is the more sensitive measure of change, we examined it first.

Table 1 presents the means and standard deviations for DAS scores for husbands and wives in TBCT and IBCT. The slope of

---

[4] Most HLM analyses with couple data have followed the methods of Raudenbush, Brennan, and Barnett (1995), in which husbands' and wives' trajectories are represented in a two-level multivariate model. For the present analyses, we used a three-level model with repeated measures, individuals, and couples as the three levels. For the majority of analyses, the data were represented well by a random intercept at the individual level and random effects for intercept, slope, and quadratic at the couple level. The reason for modeling the present data in this manner is twofold. The current analyses pertain to a relatively brief period of time in which both partners are actively engaged in therapy. Thus, partners' data are extremely correlated, especially change over time. This led to extreme correlations among the random effects using Raudenbush et al.'s (1995) approach with some models. In addition, the three-level model is more parsimonious in its representation of the random effects. Raudenbush et al.'s (1995) model uses 22 parameters in estimating the random effects, whereas the three-level model uses only 8 parameters. Comparisons between the two-level and three-level representations using Akaike's information criterion (Akaike, 1974) and Bayes's information criterion (Schwarz, 1978) favored the three-level model in the present analyses.

Table 1
Dyadic Adjustment Scale and Global Distress Scale Scores for Wives and Husbands in Traditional Behavioral Couple Therapy (TBCT) and Integrative Behavioral Couple Therapy (IBCT)

Measure	Pretreatment		Week 13		Week 26		Final session	
	M	SD	M	SD	M	SD	M	SD
Dyadic Adjustment Scale								
TBCT								
Wife	83.71	14.22	88.34	12.83	92.17	17.38	93.84	18.64
Husband	82.69	16.46	90.66	14.91	95.15	18.60	96.72	18.96
IBCT								
Wife	85.72	13.76	88.96	17.64	91.14	19.09	96.06	20.67
Husband	86.36	13.11	90.97	17.00	92.78	18.82	99.41	16.77
Global Distress Scale								
TBCT								
Wife	64.64	5.76	64.16	7.11	61.74	6.89	61.07	8.31
Husband	69.29	5.98	67.18	6.96	64.17	8.32	62.55	8.89
IBCT								
Wife	64.81	3.82	63.25	6.47	62.45	7.72	59.71	8.88
Husband	68.06	4.75	65.53	8.15	64.60	8.50	62.34	8.80

the trajectory of the DAS across intake, 13-week, 26-week, and final session assessments was highly significant, $B = 0.365$, $SE = 0.078$, $t(723) = 4.71$, $p < .0001$. Overall, couples were improving in their marital satisfaction over treatment at the rate of 0.37 DAS points per week.[5] The pre- to posttreatment effect size of this overall change during treatment was $d = .86$. The quadratic effect (a deceleration or flattening out of the rate of change over time) was not significant, $B = -0.002$, $SE = 0.002$, $t(723) = -1.126$, *ns*. Thus, considering both therapies together, there was no evidence that couple therapy had its primary impact early in treatment.

There was a significant effect of therapy on the slope,[6] $B = 0.170$, $SE = 0.078$, $t(723) = 2.19$, $p < .03$, $d = .58$,[7] and a significant effect of therapy on the quadratic, $B = -0.004$, $SE = 0.002$, $t(723) = -2.053$, $p < .05$, $d = .62$. These interactions can be seen in Figure 1, which plots the predicted DAS values for each of the two treatments over time. TBCT couples improved more quickly than IBCT couples but then plateaued while IBCT couples showed slow but steady improvement across treatment with no flattening out or deterioration.

There was no effect of gender on the intercept. However, there was a significant gender effect on the slope, $B = 0.105$, $SE = 0.050$, $t(717) = 2.11$, $p < .05$, $d = .37$, but not on the quadratic, $B = -0.002$, $SE = 0.001$, $t(717) = -1.711$, $p < .10$. Husbands progressed more quickly in treatment than wives.

As expected, there was a significant effect of stratification[8] on the intercept of DAS scores, $B = -8.536$, $SE = 0.727$, $t(130) = -11.742$, $p < .0001$. There was no significant impact of stratification on the slope, but there was a significant impact of stratification on the quadratic, $B = -0.005$, $SE = 0.002$, $t(717) = -2.72$, $p < .01$, $d = .88$. As compared with moderately distressed couples, severely distressed couples showed a greater slowing or flattening out of satisfaction over time.

There was no significant effect of pilot versus study proper cases on the intercept of DAS scores. However, there was a significant pilot effect on the slope of DAS scores, $B = 0.25$, $SE = 0.103$, $t(717) = 2.43$, $p < .05$, as well as a significant pilot effect

on the quadratic, $B = -0.004$, $SE = 0.002$, $t(717) = -2.151$, $p < .05$. Compared with study proper cases, pilot cases showed greater improvement early on in treatment but then showed a greater deceleration or flattening out later in treatment. There were no significant effects for site. Nor did we find significant effects for therapists.

Next, we computed clinical significance statistics using average couple DAS scores by treatment condition. We computed clinical significance using the method described in Jacobson and Truax (1991). Because there is normative information for the DAS (Spanier, 1976), we used their "cutoff c" for our calculations, which is the midpoint between the normative mean and the pretherapy mean. With the present data, this yielded a cutoff score of 96.8, which is slightly lower than one standard deviation below the normative mean. Our categories of clinical significance are deteriorated (reliable change in a negative direction; separation, or dropout of treatment because of doing poorly), no change (no reliable improvement in either direction), improvement (reliable

---

[5] This interpretation of the slope is correct initially (i.e., at the intercept). However, if there are significant higher order terms in the model, such as a significant quadratic, the rate of change is affected accordingly.

[6] Categorical variables were coded using effect contrasts. For therapy, TBCT was coded as 1 and IBCT was coded as –1. Thus, the coefficient represents the deviation of each therapy from the overall slope (i.e., IBCT slope coefficient is 0.195; TBCT slope coefficient is 0.535).

[7] At the present time, there is not a universally accepted approach to calculating effect sizes in HLM analyses. We have followed the recommendation of Raudenbush and Liu (2001) for binary predictors by dividing the fixed effect estimate by the square root of the corresponding random effect (see those authors' Footnote 4 on p. 391 for differences between their approach and Cohen's *d*).

[8] Including stratification as a fixed effect predictor of the intercept led to a huge decrease in the random variation at the intercept. This is to be expected because the stratification variable is based on pretreatment marital satisfaction scores. Thus, models with stratification as a predictor have random effects at the couple level for slope and quadratic but not intercept.
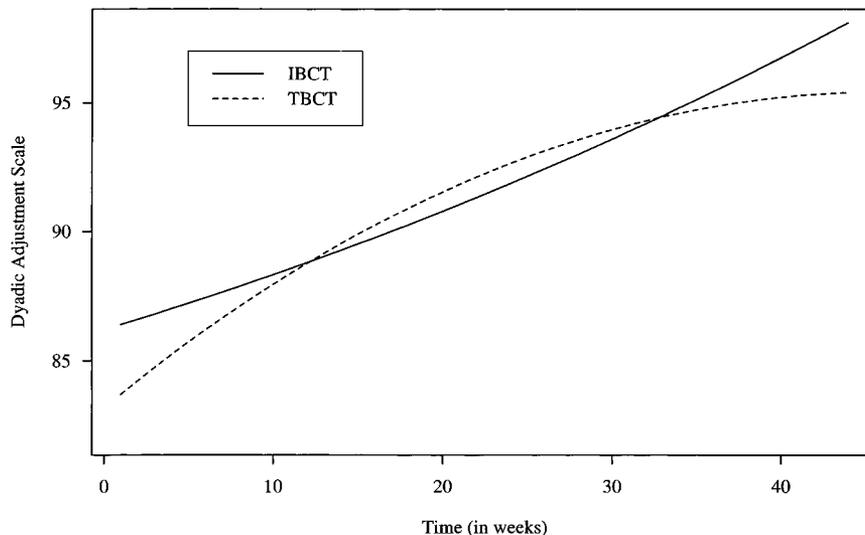
Figure 1. Predicted Dyadic Adjustment Scale (DAS) scores over treatment for Traditional (TBCT) and Integrative (IBCT) Behavioral Couple Therapy.

change in a positive direction but not reaching the normal range), and recovery (reliable change in a positive direction and reaching the normal range—i.e., DAS > 96.8). Table 2 provides these data on the DAS. In IBCT, 71% of couples showed reliable improvement or recovery. In TBCT, 59% of couples showed reliable improvement or recovery. A chi-square test examining whether therapy condition was independent of outcome was nonsignificant, $\chi^2(3, N = 130) = 3.32$, $p = .34$. Not surprisingly, there was a significant difference in clinical outcome between moderately and severely distressed couples, $\chi^2(3, N = 130) = 13.90$, $p = .0030$. For DAS, 73% of the moderately distressed were improved or recovered, whereas 54% of the severely distressed were improved or recovered.

Next, we examined the GDS. Table 1 presents the means and standard deviations for GDS scores for husbands and wives in TBCT and IBCT. The slope of the trajectory of the GDS across intake, 13-week, 26-week, and final session assessments was highly significant, $B = -.125$, $SE = 0.037$, $t(675) = -3.38$, $p < .001$.[9] Overall, couples were decreasing in their global distress over treatment at the rate of 0.13 GDS T-score points per week. The pre- to posttreatment effect size of this change was $d = .85$. There was no statistically significant quadratic effect (a deceleration or flattening out of the rate of change over time), $B = 0.000$, $SE = 0.001$, $t(675) = -0.33$, *ns*. Thus, there was no evidence that couple therapy has its primary impact early in treatment. There were also no main effects of therapy, no significant Therapy × Time effects or Therapy × Quadratic effects. Couples were improving on the GDS but not differentially by treatment condition.

As with the DAS, we examined the impact of gender of spouse, stratification into high and moderate distress, pilot versus study proper cases, and the two sites. There was a significant effect of gender on the intercept, $B = 1.948$, $SE = 0.330$, $t(131) = 5.89$, $p < .0001$, with men evidencing more distress at the start of treatment than women. Because this measure is gender normed, it means that men were more distressed, relative to other men, than women were, relative to other women. There was also a significant

gender effect on the slope, $B = -0.061$, $SE = 0.027$, $t(669) = -2.28$, $p < .05$, $d = .22$, with husbands showing greater change over time than wives, but not showing a significant gender effect on the quadratic, $B = 0.001$, $SE = 0.001$, $t(669) = 1.74$, $p < .10$.

As expected, there was a significant effect of our stratification at the intercept, $B = 2.75$, $SE = 0.340$, $t(128) = 8.08$, $p < .0001$. However, there were no interactions between stratification and time or stratification and the quadratic. Furthermore, there were no other significant effects.

In comparing the findings of the DAS and GDS, there are notably fewer significant findings with the GDS. Moreover, there were some strong differences between the DAS and GDS HLM models in the reliabilities of the random effects (not shown). With both measures, the reliability of the quadratic component was somewhat lower than the reliabilities for the intercepts and slopes, which would be expected with only four total data points. However, the reliability of the random effect of the slope for the GDS was also quite low. The empirical Bayes's estimates of the random effects are "shrunken" toward the mean effects based on the reliabilities. Thus, the GDS slope and quadratic random effects were largely determined by the group means leading to very high correlations among the GDS random effects. In addition, Raudenbush and Liu (2001) have shown that the power of a fixed effect depends on the reliability of the random effect. Thus, the lower reliabilities for the GDS affect the power to find effects and may partly explain the differences between the DAS and GDS.

One possible explanation for the low reliability of the GDS change components pertains to the historical nature of some GDS items. Some GDS items are historical (e.g., "My partner and I have never come close to ending our relationship") and would thus not be expected to change during therapy. However, it is possible that

---

[9] Unlike the DAS, the GDS was not given at all time points to the pilot couples, which accounts for the slightly lower degrees of freedom for the GDS models.

Table 2
Clinically Significant Outcome on the Dyadic Adjustment Scale and the Global Distress Scale for Couples in Traditional Behavioral Couple Therapy (TBCT) and Integrative Behavioral Couple Therapy (IBCT)

Measure	Deteriorated	Unchanged	Improved	Recovered
Dyadic Adjustment Scale				
TBCT	12 (18%)	15 (23%)	10 (15%)	29 (44%)
IBCT	5 (8%)	14 (22%)	12 (19%)	33 (52%)
Global Distress Scale				
TBCT	13 (20%)	15 (23%)	12 (18%)	26 (39%)
IBCT	5 (8%)	18 (28%)	8 (13%)	33 (52%)

Note. Clinical significance could not be computed for two TBCT and two IBCT couples, who did not complete the Dyadic Adjustment Scale or Global Distress Scale at or near treatment end.

some participants answered these questions with respect to their last assessment point. To the extent that some individuals may have answered historical items in a literal sense, whereas others answered them relative to their last assessment, this would increase the measurement error of the GDS, particularly with respect to change over time, and may have contributed to low reliabilities.

As we did with the DAS, we computed percentages of couples who achieved clinically significant change on the GDS. Table 2 presents these data. In IBCT, 65% of couples showed reliable change or recovery; the comparable figure for TBCT was 57%. Chi-square analyses revealed no significant differences in outcome between therapy conditions, $\chi^2(3, N = 130) = 5.43, p = .14$. However, as with the DAS, there was a significant difference in clinical outcome between moderately and severely distressed couples, $\chi^2(3, N = 130) = 19.81, p = .0001$. On the GDS, 67% of the moderately distressed were improved or recovered while 53% of the severely distressed were improved or recovered.

Relationship stability. Table 3 presents the means, medians, and standard deviations of the MSI for husbands and wives in TBCT and IBCT. The MSI was highly skewed; the median MSI at 13 weeks and 26 weeks is 1 in most cases. Thus, most couples are reducing the number of steps and bottoming out at 0 or 1 step, which is the normal level, but a few couples are remaining unstable or becoming more so. This nonnormal distribution of the MSI is best represented by the Poisson distribution, which is often used for rate or count variables (Long, 1997). Therefore, we modeled the changes in MSI using a GLMM with a Poisson distribution and log link (Breslow & Clayton, 1993; Raudenbush & Bryk, 2002). In

examining the data below, it is important to remember that the coefficients reported are on the log scale.

Looking at the basic model, we find that the slope of the trajectory is significantly decreasing over time, $B = -0.069, SE = 0.009, t(455) = -8.08, p < .0001$. However, there is also a significant quadratic effect, $B = 0.001, SE = 0.000, t(455) = 4.77, p < .0001$. Because GLMM models are inherently nonlinear, it is usually recommended that significant effects be portrayed through plotting the estimated regression line or providing estimates at meaningful values of the covariates (Long, 1997). The estimated regression values at pretreatment and 26-week assessment are 3.6 and 1.6, respectively. A plot of the regression line (not shown) demonstrates that marital instability is decreasing over time, but the primary decrease occurs during the early stage of treatment and then "bottoms out." There were no significant Therapy × Time or Therapy × Quadratic interactions.

We also looked at the impact of gender of spouse, stratification into high and moderate distress, pilot versus study proper cases, and the two sites. There was no impact of site, pilot, or gender. As expected, there was a main effect of stratification, $B = -0.303, SE = 0.055, t(131) = -5.52, p < .0001$. Severely distressed couples were more unstable than moderately distressed couples. There was a significant Gender × Stratification interaction on the quadratic, $B = 0.001, SE = .000, t(445) = 2.03, p < .05$. In severely distressed couples, husbands' stability flattened out more so than wives'. This pattern was somewhat reversed in moderately distressed couples. There were also significant Stratification × Therapy interactions on the slope, $B = -0.207, SE = 0.009$,

Table 3
The Marital Status Inventory for Wives and Husbands in Traditional Behavioral Couple Therapy (TBCT) and Integrative Behavioral Couple Therapy (IBCT)

Condition	Pretreatment			Week 13			Week 26		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
TBCT									
Wife	4.10	3.50	2.93	2.75	1.50	3.35	2.68	1.00	3.08
Husband	4.53	5.00	2.91	2.67	2.00	2.64	2.71	1.50	3.47
IBCT									
Wife	3.35	3.00	2.40	2.08	1.00	2.50	2.51	1.00	3.26
Husband	4.11	4.00	3.03	2.33	1.00	2.85	2.78	1.00	3.52

$t(445) = -2.326$, $p < .02$, and on the quadratic, $B = 0.001$, $SE = 0.000$, $t(445) = 2.39$, $p < .05$. TBCT and IBCT performed similarly for severely distressed couples but TBCT performed somewhat better for moderately distressed couples. However, this difference seemed to be driven primarily by a tendency toward pretreatment differences in moderately distressed TBCT and IBCT cases.

Communication.   We examined both AFC and PSC subscales from the MSI-R. Table 4 presents the means and standard deviations for husbands and wives in TBCT and IBCT over the three assessment points for both of these variables. AFC showed improvement over time, averaging about 0.11 T-score points per week, $B = -0.112$, $SE = 0.050$, $t(490) = -2.26$, $p < .05$. There was no significant quadratic effect. Furthermore, there were no main effects of therapy condition or interactions with time or the quadratic.

Looking at the other predictors, there were no significant differences in AFC for site or pilot cases. However, women began therapy more distressed in AFC than men, $B = -0.936$, $SE = 0.373$, $t(130) = -2.51$, $p < .05$. There were no differences between spouses in the amount of change over the course of therapy. The analyses also revealed that severely distressed couples began treatment more distressed in their AFC, $B = 2.428$, $SE = 0.421$, $t(130) = 5.76$, $p < .0001$, and changed less during therapy, $B = 0.040$, $SE = 0.020$, $t(486) = 1.99$, $p < .05$.

PSC changed over time, improving on the average of 0.1 T-score points per week, $B = -0.106$, $SE = 0.052$, $t(491) = -2.05$, $p < .05$. There was no significant quadratic effect, nor was there an effect of therapy or any interactions between therapy and time or therapy and the quadratic. There were no significant effects of gender, site, or pilot. There was a significant difference based on stratification such that severely distressed couples began treatment with poorer PSC than moderately distressed couples, $B = 1.898$, $SE = 0.422$, $t(491) = 4.49$, $p < .0001$.

Individual variables.   We examined the overall MHI and CS from the Compass. Table 5 presents the means and standard deviations for husbands and wives in TBCT and IBCT over the three assessment points. On average, both husbands and wives

start out slightly better than typical outpatients in individual therapy. HLM analyses revealed no significant effects of time, the quadratic, or therapy. Overall, MHI changed very little over the course of treatment.

Because our treatment was directed at the marital relationship, we expected improvements in individual functioning only to the extent that the marital relationship improved. To test this hypothesis, we included DAS scores as a time varying covariate. Following the suggestions of Diggle, Liang, and Zeger (1994), two components of DAS scores were included to reflect the level of satisfaction and change in satisfaction for each individual. Time 1 DAS scores and the individuals' deviations from their Time 1 scores were calculated and entered into the HLM models. Pretreatment DAS was not a significant predictor of the MHI, indicating that DAS scores and MHI scores were unrelated at pretreatment. However, changes in DAS scores over time were highly associated with changes in MHI scores, $B = 0.116$, $SE = 0.025$, $t(486) = 4.70$, $p < .0001$. For each point increase in DAS, the MHI increased by 0.116 points. Therefore, to the extent that MHI changed, it changed only as DAS changed. The only other significant effect was an effect of stratification on the intercept: Severely distressed couples scored more poorly on the MHI than did moderately distressed couples, $B = 1.481$, $SE = 0.621$, $t(130) = -2.39$, $p < .05$.

Current symptoms data were highly skewed, showing that the majority of individuals reported few psychological symptoms with a smaller number of individuals reporting clinical levels of psychological symptoms. We used Box–Cox's method to identify an appropriate transformation (Draper & Smith, 1998; Venables & Ripley, 1997). The analysis revealed that raising the symptoms variable to the –3 power (i.e., one over symptoms cubed) yielded a reasonably normal distribution. As with the analyses of the MHI, there were no significant effects of time, the quadratic, or therapy. Overall, current symptoms changed very little over the course of treatment.

Following the same logic as used with the MHI, we only expected improvements in symptom scores to the extent that the marital relationship improved. To test this hypothesis, we again

Table 4

Affective and Problem-Solving Communication for Wives and Husbands in Traditional Behavioral Couple Therapy (TBCT) and Integrative Behavioral Couple Therapy (IBCT)

Communication	Pretreatment		Week 13		Week 26	
	M	SD	M	SD	M	SD
Affective						
TBCT						
Wife	65.43	6.82	63.21	7.62	60.84	9.11
Husband	62.35	7.51	60.66	7.74	58.97	8.24
IBCT						
Wife	63.21	6.99	61.59	8.05	60.59	8.79
Husband	62.32	5.62	61.38	8.01	59.98	8.64
Problem solving						
TBCT						
Wife	64.29	5.76	62.29	6.12	59.18	7.77
Husband	63.32	5.96	61.43	6.58	58.83	7.68
IBCT						
Wife	62.52	6.36	60.77	7.68	58.81	8.30
Husband	62.82	7.19	61.75	7.03	60.11	8.36

Table 5
Mental Health Index and Current Symptoms for Wives and Husbands in Traditional Behavioral Couple Therapy (TBCT) and Integrative Behavioral Couple Therapy (IBCT)

Variable	Pretreatment		Week 13		Week 26	
	M	SD	M	SD	M	SD
Mental Health Index						
TBCT						
Wife	60.32	7.18	61.51	7.98	62.83	8.54
Husband	62.70	8.60	61.53	8.13	63.70	7.31
IBCT						
Wife	61.02	8.23	61.60	7.85	61.07	8.40
Husband	61.79	7.16	61.41	8.07	62.79	8.26
Current symptoms						
TBCT						
Wife	38.50	5.68	38.95	6.26	38.57	6.92
Husband	37.74	6.80	38.57	6.45	37.63	5.43
IBCT						
Wife	38.55	6.41	38.51	5.92	38.85	6.18
Husband	38.33	5.36	38.94	5.89	38.10	5.33

included DAS scores as a time-varying covariate. The pretreatment DAS was not statistically significant, indicating that DAS scores and symptom scores are unrelated at pretreatment. However, improvement in DAS scores over time was highly associated with improvement in symptom scores, $t(490) = 3.00$, $p < .01$ (because these analyses were done with transformed scores, the coefficients and standard errors are on the inverse-cube scale and thus are not directly interpretable). Furthermore, after controlling for DAS scores, there was a significant effect of time, $t(490) = -2.08$, $p < .05$, indicating that symptom scores were improving. There was also a significant therapy by stratification by time interaction, $t(484) = 2.37$, $p < .05$. Graphs of the predicted values indicated that husbands and wives in severely distressed couples were both improving in both TBCT and IBCT, but moderately distressed wives were improving more in TBCT. Because these effects are found on transformed data after controlling for DAS scores, they should be interpreted very cautiously.

Client Reactions to Treatment

We obtained two measures of client reactions to treatment: a measure of the therapeutic bond at the end of the fourth session (the feedback session) and a measure of client evaluation of services at the end of the final session. Mean scores on the therapeutic bond measure in TBCT were 52.73 ($SD = 6.88$) for wives and 51.53 ($SD = 7.38$) for husbands. In IBCT, mean scores were 52.85 ($SD = 6.57$) for wives and 51.01 ($SD = 8.07$) for husbands. Mean scores on the client evaluation of services in TBCT were 27.54 ($SD = 4.80$) for wives and 27.65 ($SD = 4.96$) for husbands. Mean scores on the client evaluation of services in IBCT were 27.85 ($SD = 4.76$) for wives and 27.57 ($SD = 4.74$) for husbands. On both measures, clients scored slightly higher than the normative means for outpatient samples on these measures, indicating that, in general, clients had a good bond with their therapist and were satisfied with treatment.

Because each of these measures was given at only one occasion, analysis of variance was used to examine the results. We considered three 2-level between-groups factors (therapy condition, site, and cohort [pilot vs. proper]) and one 2-level repeated measures factor (husband vs. wife). Spouse was considered a repeated measures factor because of the linkage between husband and wife in the same marriage. On the therapy bond measure, wives rated their bond slightly higher than husbands, $F(1, 122) = 4.53$, $p < .05$. There were no other significant main effects or interactions. On the evaluation of services, there were no significant main effects or interactions.

Discussion

Results must always be considered in the context of the sample from which they are derived. We deliberately sought a sample of significantly and chronically distressed couples. We excluded 94 couples who wanted treatment but were not sufficiently dissatisfied on one of our three measures of marital satisfaction given at three time points. A brief phone follow-up on these excluded couples indicated that one half sought therapy elsewhere (Frousakis, Simpson, & Christensen, 2003). Of the couples who participated in our therapy study, more than half had received couple therapy before. Also, about half scored 66 or higher on the GDS of the MSI–R; Snyder (1997) found that couples with an intake GDS *T* score of 66 or higher who received treatment had a 45% chance of eventually divorcing. Thus, these were clearly seriously distressed couples. We believed that a stably and seriously distressed population would provide the most rigorous test of marital therapy in general and of IBCT and TBCT in particular.

Heyman and Neidig (1997) noted that "All marital therapists treat violent couples, whether they know it or not." Our data support the truth in that statement. Seventy-three percent of all the couples that we screened reported at least one incident of physical violence at some point in their relationship, and 44% reported at least one incident of severe violence. Using conservative criteria for male-to-female battering, we eliminated 101 couples and referred them for treatment targeted at violence. We were concerned about safety issues for these women and did not believe that standard marital therapy was the appropriate treatment for battering (Bograd & Mederos, 1999).

Our first hypothesis addressed the overall impact of marital therapy on this sample of couples. We found statistically significant effects indicating improved relationship satisfaction, stability, and communication. Not surprisingly, improvements in individual functioning only occurred to the extent that marital satisfaction improved. Even though there was no control group for comparison to these treatment effects, a number of studies have shown that no-treatment control groups of distressed couples show no improvement without treatment and even some deterioration. For example, across 17 studies, Baucom, Hahlweg, and Kuschel (2003) found an average effect size for control groups of $M = -0.06$ ($SE = 0.10$). In contrast, the effect size of the improvement in marital satisfaction in this study was large and comparable to or greater than other studies of marital therapy. For example, the largest meta-analysis of couple therapy (Shadish et al., 1993) found an overall effect size of $d = .60$ ($SE = .09$) for 27 studies of couple therapies. Our effect sizes of .86 for the DAS and .85 for the GDS are significantly larger. Also, our rates of clinically significant change compare favorably with other data. In their review of several behavioral studies, Jacobson et al. (1984) found that 54.7% of couples showed reliable improvement. However, only 35.3% showed reliable improvement and recovery (reaching a level of satisfaction more similar to nondistressed than distressed couples). In their meta-analysis, Shadish et al. (1993) calculated that 41% of 19 studies reviewed brought the average client in the study to recovery. Even though their strategy calculated percentages across studies rather than across couples, they obtained a similar percentage of recovery as Jacobson et al. (1984). In the current study, across both treatments, 65% of couples showed reliable improvement on the DAS (61% on the GDS), whereas 48% demonstrated recovery on the DAS (45% on the GDS). The lower the client's initial satisfaction, the more improvement they must evidence to reach recovery. In general, our study had more severe cases than those in the studies reviewed but gave them more sessions of treatment.

We did not find evidence for our second hypothesis, that satisfaction improves more rapidly early in treatment than later in treatment. These results suggest that more treatment may be better treatment, a finding that may be particularly important for severely distressed couples, who need to improve substantially.

In contrast to our satisfaction data, our measure of stability showed greater improvement early on in treatment. However, this finding is something of an artifact. By the Week 13 assessment, most couples had reached the normative level of stability on the MSI, so there was little room for further improvement. Also, no other measures showed differential change over treatment. Thus, there was little evidence to suggest that, in general, the impact of marital therapy is greatest early on in treatment. However, Halford's (2001) argument that change takes place early in treatment was made specifically in regard to traditional behavioral approaches rather than couple therapy in general. To address that more specific notion, we need to examine our third hypothesis, about the relative impact of TBCT and IBCT.

In conducting a comparison between two different psychological treatments, one must first ask: Was it a fair comparison? Several safeguards were put in place in the current study to ensure a fair comparison. First, both investigators were extensively trained in TBCT and both, particularly Jacobson, had conducted research and scholarship on that approach. In fact, Jacobson was the primary author of the primary treatment manual for TBCT. Second, therapists were provided extensive training and supervision in both approaches. When Jacobson died, an outside expert on TBCT, Don Baucom, was brought in to assist with the supervision of TBCT cases. Third, a detailed and a global coding system were used to evaluate adherence to treatment protocol. Finally, another outside expert on TBCT, Gayla Margolin, who coauthored the treatment manual for TBCT, was brought in to provide competency ratings for TBCT. In addition to these safeguards, considerable data suggest that the comparison was a fair one. First, the adherence ratings indicated that the two approaches differed in ways consistent with the treatment protocols for each. Second, the ratings provided by Margolin indicated that TBCT was competently delivered. Third, clients' ratings of the therapeutic bond indicated that couples in both conditions evaluated their therapists highly and comparably. Fourth, clients' ratings at the end of treatment indicated that couples in both conditions rated their entire treatment highly and comparably. Finally, outcome in TBCT was comparable to or exceeded that in other studies of TBCT. For example, our rates of 59% reliable improvement and 44% recovery in TBCT on the DAS compare favorably to the review by Jacobson et al. (1984) with a rate of 55% reliable improvement and 35% recovery (based primarily on the MAT and the DAS, which are quite similar).

Assuming a fair test of the two treatments, we then can examine the results of our third hypothesis about the relative outcome of IBCT versus TBCT. For the most part, TBCT and IBCT performed similarly across measures, despite being demonstrably different treatments. The results are consistent with the idea that IBCT is at least as efficacious as TBCT in treating relationship distress at the end of their respective treatments. On the DAS, however, there were significant effects of therapy on both the slope and quadratic of the trajectory of change, both of which were moderate effect sizes. As indicated in Figure 1, TBCT couples improved more quickly early on in treatment than IBCT couples but tended to flatten out over the course of therapy, whereas IBCT couples made steady improvement over the course of therapy. This pattern of more rapid improvement early in treatment, which is just what Halford (2001) predicted, is likely due to the early emphasis on behavioral exchange in TBCT. Early in therapy, prior to training in communication and problem-solving skills and to a focus on long-standing problems, TBCT therapists typically focus couples on improving their positive actions toward each other (e.g., making lists of positive actions each could do, agreeing to do more of those actions). These positive actions may create an immediate boost in satisfaction, but as couples later focus on their enduring problems, their increase in satisfaction may level off. In contrast, IBCT provides no quick boost for couples but focuses immediately on the central themes and issues that trouble the couple and leads to steady improvement. The important question raised by these intriguing differences between TBCT and IBCT is whether the different trajectories have any implications for follow-up. Is the flattening out in TBCT a predictor of poorer long-term outcome? Or is early improvement an indicator of longer maintenance? Only follow-up data can provide information about the significance of these different trajectories of change.

Our fourth hypothesis concerns differences between husbands and wives in their response to treatment. On the gender-normed GDS T scores, but not on the DAS scores, husbands were more

dissatisfied at the beginning of treatment than were wives. However, on both measures, husbands tended to improve more quickly than wives early in treatment. We know that husbands in general are more reluctant to enter therapy than are wives and that this finding is true in the current sample (Doss et al., 2003). It may be that husbands fear they will be criticized by both their wives and the therapist for their misdeeds and are reluctant to put themselves in such a situation by coming to therapy. When husbands enter treatment and find that, contrary to their fears, the therapist takes an even-handed stance toward both and that therapy may benefit them as well as their wives, husbands may show a greater rise in satisfaction than their wives. The only other differences between spouses occurred with AFC, in which wives started therapy more distressed than husbands, and with the therapeutic bond, in which wives rated their therapists more highly than did husbands. Both of these differences fit gender-role stereotypes, with wives being more attuned to relationships and to emotional communication than husbands.

Our final hypothesis concerned the impact of our stratification of couples into moderately and severely distressed groups. As expected, there was a significant impact of stratification on the intercepts for all our variables, but stratification did not affect the slopes of any variables and only the quadratic of the DAS. Although severely distressed couples evidenced a similar rate of change as moderately distressed couples, they demonstrated a greater flattening out on the DAS. Although there were a couple of inconsistent interactions with other variables, there was, overall, very little impact of stratification on the rate of change in therapy. Thus, although severely distressed couples start at a significantly lower point than moderately distressed couples on most of our outcome measures, they improve at a comparable rate.

This finding of comparable rates of change in severely and moderately distressed couples is encouraging. It means that IBCT and TBCT can be applied to even very severely distressed couples with a reasonable hope of improvement. However, it is important to remember that the intercepts of the two groups are still significantly different. Even though they improve similarly, the severely distressed couples start out and end up in worse condition. This fact can be seen when we examine rates of recovery for moderately and severely distressed couples. Because recovery depends not just on how much a couple has improved but also on the absolute level of satisfaction achieved, a moderately distressed couple is more likely to achieve recovery than a severely distressed couple. Whether or not a couple achieves recovery (a level of "normal" satisfaction) may have important implications for maintenance.

There are a number of important strengths of this study. First, with 134 couples, it is the largest clinical trial of couple therapy to date. To our knowledge, the next largest is a study by Hahlweg, Revenstorf, and Schindler (1982) that involved 85 couples. A large clinical trial gives the researcher the necessary statistical power to provide a robust test of hypotheses. Second, this is the first study to require that couples meet repeated criteria for dissatisfaction. Thus, we generated a group of stably dissatisfied couples and reduced the impact of regression toward the mean on results. Because of our recruitment strategies, we have one of the most, if not the most, severely distressed samples of couples to participate in a randomized clinical trial. Third, this is probably the most diverse sample of couples to participate in a clinical trial of couple therapy. This diversity comes about in part because couples were recruited from two geographical locations, Los Angeles and Seattle. Also, specific efforts were made to recruit minority couples (e.g., advertisements in minority outlets, use of minority therapists). In almost one third of the couples in the current study (42 couples), one or both partners were Asian, African American, or Latino. Most clinical trials of couple therapy have not reported the percentage of minority couples, but of those that have, this, to our knowledge, is the highest level of minority participation. Fourth, we believe our study provides the most rigorous test to date of different treatments. Because of the use of experienced therapists, the high level of training and supervision, the measures of adherence and competence, and the measures of therapeutic bond and evaluation of services, we could provide a high level of comparable service delivery and measure that level of delivery. Fifth, this is the first clinical trial of couple therapy to use the statistical strategies of HLM. Because these strategies enable the researcher to separate (a) initial status (intercept), (b) rate of change (slope), and (c) change in the rate of change (quadratic), they provide a more differentiated view of outcome. In addition, these strategies allow the researcher to consider simultaneously the impact of treatment on husbands and wives, nested within a marriage. Finally, this is the first study we know of that systematically examines the level of initial distress by stratifying couples into severely and moderately distressed conditions.

There are also limitations of the current study. Because it was an efficacy rather than an effectiveness study, the emphasis in this study was on internal rather than external validity. There are substantial limitations in generalizing our results to the practice of couple therapy. Although our therapists were licensed practitioners in the community who saw couples in their private offices, we selected these therapists on the basis of the quality of their training, experience, and reputation. They are hardly a randomly selected group of therapists. They also practiced treatments that are probably not widely used in the community. IBCT is too new to be widely adopted. While TBCT has had considerable influence, some of its strategies, such as communication training, are probably used with an eclectic mix of other strategies. In addition, we provided our therapists with extensive training and supervision, a provision that is not ordinarily a part of usual practice. Our couples were given free treatment and paid to participate in assessments. It is unclear how these factors affect outcome. Also, couples were given a maximum of 26 treatment sessions, which probably exceeds the amount usually provided by managed care organizations. In general, our approach was to provide a test of couple therapy under optimal treatment conditions but with highly distressed couples. There is little research on the effectiveness of couple therapy in the community to serve as a comparison to our treatment effects. However, a recent study by Hahlweg and Klann (1997) evaluated the outcome of couple therapy as implemented by practicing clinicians in Germany. Their findings indicated an overall effect size of only .28, which is small compared with our effect size and with effect sizes for efficacy studies. The comparison suggests that we may have been successful in providing an optimal test of our treatments. Later research can examine the extent to which these findings can generalize to less trained therapists, to fewer sessions, and to more mildly distressed couples.

Perhaps the most serious limitation of the current study is that it provides data only on the immediate outcome of treatment. We cannot fully evaluate the success of treatment until we look at the

long-term maintenance of treatment gains. Traditional behavioral approaches to couple therapy have demonstrated disappointing results at long-term follow-up (Jacobson et al., 1987; Snyder et al., 1991). IBCT was developed in part to address long-term follow-up. Specifically, its focus on broad relationship themes rather than target behaviors, its emphasis on emotional acceptance as well as change, and its emphasis on contingency-shaped rather than deliberate, rule-governed change were designed to lead to greater maintenance of gains in satisfaction. We are currently assessing the couples in this study at 6-, 12-, 18-, and 24-month follow-up periods. A later article will provide results on these follow-up data.

Serious marital distress is all too common. Current estimates are that half of first marriages will end in divorce and second marriages will fare even worse (Bramlett & Mosher, 2001). Marital distress and divorce are associated with a number of negative consequences for spouses and their children (Amato, 2000). Therefore, any effective treatment for marital distress will likely prevent a host of damaging effects for both spouses and their offspring. The current study suggests that couple therapy can be effective, at least in the short term, for even very seriously distressed couples.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

Amato, P. R. (2000). The consequences of divorce for adults and children. *Journal of Marriage and the Family, 62,* 1269–1287.

Baucom, D. H., Hahlweg, K., & Kuschel, A. (2003). Are waiting-list control groups needed in future marital therapy outcome research? *Behavior Therapy, 34,* 179–188.

Baucom, D. H., Shoham, V., Mueser, K. T., Daiuto, A. D., & Stickle, T. R. (1998). Empirically supported couple and family interventions for marital distress and adult mental health problems. *Journal of Consulting and Clinical Psychology, 66,* 53–88.

Bograd, M., & Mederos, F. (1999). Battering and couples therapy: Universal screening and selection of treatment modality. *Journal of Marital and Family Therapy, 25,* 291–312.

Bramlett, M. D., & Mosher, W. D. (2001). *First marriage dissolution, divorce, and remarriage: United States: Advance data from Vital and Health Statistics* [Report No. 323]. Hyattsville, MD: National Center for Health Statistics.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88,* 9–25.

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

Christensen, A., & Heavey, C. L. (1999). Interventions for couples. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), *Annual review of psychology* (pp. 165–190). Palo Alto, CA: Annual Reviews.

Christensen, A., & Jacobson, N. S. (2000). *Reconcilable differences.* New York: Guilford Press.

Christensen, A., Jacobson, N. S., & Babcock, J. C. (1995). Integrative behavioral couple therapy. In N. S. Jacobson & A. S. Gurman (Eds.), *Clinical handbook of marital therapy* (2nd ed., pp. 31–64). New York: Guilford Press.

Crane, D. R., Newfield, N., & Armstrong, D. (1984). Predicting divorce at marital therapy intake: Wives' distress and the Marital Status Inventory. *Journal of Marital and Family Therapy, 10,* 305–312.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data.* New York: Oxford University Press.

Doss, B. D., Atkins, D. C., & Christensen, A. (2003). Who's dragging their feet?: Husbands and wives seeking marital therapy. *Journal of Marital and Family Therapy, 29,* 165–177.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis.* New York: Wiley.

Dunn, R. L., & Schwebel, A. I. (1995). Meta-analytic review of marital therapy outcome research. *Journal of Family Psychology, 9,* 58–68.

Eldridge, K. A., & Christensen, A. (2002). Demand–withdraw communication during couple conflict: A review and analysis. In P. Noller & J. A. Feeney (Eds.), *Understanding marriage: Developments in the study of couple interaction* (pp. 289–322). Cambridge University Press.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1994). *Structured Clinical Interview for Axis I DSM-IV Disorders—Patient Edition.* Washington, DC: American Psychiatric Press.

Frousakis, N. N., Simpson, L. E., & Christensen, A. (2003, November). *Changing on their own: Characteristics and outcomes of couples excluded from clinical trials.* Paper presented at the annual convention of the Association for the Advancement of Behavior Therapy in Boston.

Gottman, J. M., Notarius, C., Markman, H., & Gonso, J. (1977). *A couple's guide to communication.* Champaign, IL: Research Press.

Hahlweg, K., & Klann, N. (1997). The effectiveness of marital counseling in Germany: A contribution to health services research. *Journal of Family Psychology, 11,* 410–421.

Hahlweg, K., & Markman, H. J. (1988). Effectiveness of behavioral marital therapy: Empirical status of behavioral techniques in preventing and alleviating marital distress. *Journal of Consulting and Clinical Psychology, 56,* 440–477.

Hahlweg, K., Revenstorf, D., & Schindler, L. (1982). Treatment of marital distress: Comparing formats and modalities. *Advances in Behavioral Research and Therapy, 4,* 57–74.

Halford, W. K. (2001). *Brief therapy for couples: Helping partners help themselves.* New York: Guilford Press.

Heyman, R. E., & Neidig, P. H. (1997). Physical aggression couples treatment. In W. K. Halford & H. Markman (Eds.), *Clinical handbook of marriage and couples' interventions* (pp. 589–617). New York: Wiley.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5,* 299–314.

Jacobson, N. S., & Addis, M. E. (1993). Research on couples and couple therapy: What do we know? *Journal of Consulting and Clinical Psychology, 61,* 85–93.

Jacobson, N. S., & Christensen, A. (1994). *Traditional behavioral couple therapy manual.* Unpublished manuscript, University of Washington.

Jacobson, N. S., & Christensen, A. (1998). *Acceptance and change in couple therapy: A therapist's guide to transforming relationships.* New York: Norton.

Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: An acceptance-based, promising new treatment for couple discord. *Journal of Consulting and Clinical Psychology, 68,* 351–355.

Jacobson, N. S., Follette, W. C., Revenstorf, D., Baucom, D. H., Hahlweg, K., & Margolin, G. (1984). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. *Journal of Consulting and Clinical Psychology, 52,* 497–504.

Jacobson, N. S., & Gottman, J. M. (1998). *When men batter women: New insights into ending abusive relationships.* New York: Simon & Schuster.

Jacobson, N. S., Gottman, J. M., Waltz, J., Rushe, R., Babcock, J., & Holtzworth-Munroe, A. (1994). Affect, verbal content, and psychophysiology in the arguments of couples with a violent husband. *Journal of Consulting and Clinical Psychology, 62,* 982–988.

Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles.* New York: Brunner/Mazel.

Jacobson, N. S., Schmaling, K. B., & Holtzworth-Munroe, A. (1987). Component analysis of behavioral marital therapy: 2-year follow-up and

prediction of relapse. *Journal of Marital and Family Therapy, 13,* 187–195.

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.

Johnson, S. M., & Denton, W. (2002). Emotionally focused couple therapy: Creating secure connections. In A. S. Gurman & N. S. Jacobson (Eds.), *Clinical handbook of couple therapy* (pp. 221–250). New York: Guilford Press.

Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (1999). Emotionally focused couple therapy: Status and challenges. *Clinical Psychology: Science and Practice, 6,* 67–79.

Kashy, D. A., & Snyder, D. K. (1995). Measurement and data analytic issues in couples research. *Psychological Assessment, 7,* 338–348.

Kelly, A. B., & Halford, W. K. (1995). The generalisation of cognitive behavioural marital therapy in behavioural, cognitive and physiological domains. *Behavioural & Cognitive Psychotherapy, 23,* 381–398.

Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living, 21,* 251–255.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

Mathsoft. (1999). S-Plus 2000 [Computer software]. Seattle, WA: Data Analysis Products Division, Author.

Nguyen, T. D., Attkisson, C. C., & Stegner, B. L. (1983). Assessment of patient satisfaction: Development and refinement of a service evaluation questionnaire. *Evaluation and Program Planning, 6,* 299–314.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus.* New York: Springer.

Raudenbush, S. W., Brennan, R. T., & Barnett, R. C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology, 9,* 161–174.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Raudenbush, S.W. & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6,* 387–401

Saunders, S. M., Howard, K. I., & Orlinsky, D. E. (1989). The Therapeutic Bond Scales: Psychometric characteristics and relationship to treatment effectiveness. *Psychological Assessment. 1,* 323–330.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Shadish, W. R., Montomery, L. M., Wilson, P., Wilson, M. R., Bright, I., & Okwumabua, T. (1993). Effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology, 61,* 992–1002.

Skinner, B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), *Problem solving: Research method teaching* (pp. 225–257). New York: Wiley.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage Publications.

Snyder, D. K. (1997). *Marital Satisfaction Inventory—Revised (MSI–R) Manual.* Los Angeles: Western Psychological Services.

Snyder, D. K., Mangrum, L. F., & Wills, R. M. (1993). Predicting couples' response to marital therapy: A comparison of short- and long-term predictors. *Journal of Consulting and Clinical Psychology, 61,* 61–69.

Snyder, D. K., & Wills, R. M. (1989). Behavioral versus insight-oriented marital therapy: Effects on individual and interspousal functioning. *Journal of Consulting and Clinical Psychology, 57,* 39–46.

Snyder, D. K., Wills, R. M., & Grady-Fletcher, A. (1991). Long-term effectiveness of behavioral versus insight-oriented marital therapy: A 4-year follow-up study. *Journal of Consulting and Clinical Psychology, 59,* 138–141.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38,* 15–28.

Sperry, L., Brill, P. L., Howard, K. I., & Grissom, G. R. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions.* New York: Brunner/Mazel.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1994). *Structured Clinical Interview for DSM–IV Personality Disorders (SCID-II).* Washington, DC: American Psychiatric Press.

SPSS. (2001). Statistical package for the social sciences 11.0.1 for Windows [Computer software]. Chicago: Author.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The Revised Conflict Tactics Scales: Development and preliminary psychometric data. *Journal of Family Issues, 17,* 283–316.

Venables, W. N., & Ripley, B. D. (1997). *Modern applied statistics with S-Plus.* New York: Springer.

Vu, C. T., & Christensen, A. (2003, April). *Methodology of outcome studies in couple therapy: Analysis and recommendations.* Paper presented at the annual meeting of the Western Psychological Association, Vancouver, British Columbia, Canada.

Weiss, R. L., & Cerreto, M. (1980). The Marital Status Inventory: Development of a measure of dissolution potential. *The American Journal of Family Therapy, 8,* 80–86.

Whisman, M. A., & Jacobson, N. S. (1992). Change in marital adjustment following marital therapy: A comparison of two outcome measures. *Psychological Assessment, 4,* 219–223.

Wile, D. B. (2002). Collaborative couple therapy. In A. S. Gurman & N. S. Jacobson (Eds.), *Clinical handbook of couple therapy* (pp. 281–307). New York: Guilford Press.

Wimberly, J. D. (1998). An outcome study of integrative couples therapy delivered in a group format (Doctoral dissertation, University of Montana, 1997). *Dissertation Abstracts International: Section B: The Sciences & Engineering, 58(12-B),* 6832.