

# Dyadic predictors of outcome in a cognitive-behavioral program for patients with generalized anxiety disorder in committed relationships: A “spoonful of sugar” and a dose of non-hostile criticism may help

Richard E. Zinbarg<sup>a,b,\*</sup>, Jeong Eun Lee<sup>a</sup>, K. Lira Yoon<sup>a</sup>

<sup>a</sup>Psychology Department, Northwestern University, 102 Swift Hall, Evanston, IL 60208-2710, USA

<sup>b</sup>The Family Institute at Northwestern University, 618 Library Place, Evanston, IL 60201, USA

Received 23 February 2006; received in revised form 27 May 2006; accepted 16 June 2006

---

## Abstract

The present study tested whether pre-treatment levels of partner hostility and non-hostile criticism predicted outcome in an individual cognitive-behavioral therapy package for generalized anxiety disorder (GAD). Eighteen patients with a principal or co-principal diagnosis of GAD were randomly allocated to a treatment condition ( $n = 8$ ) or a delayed treatment condition ( $n = 10$ ). In addition, the patients and their partners were videotaped discussing the patients' worries. These videotapes were later coded for levels of partner hostility and non-hostile criticism directed at the patients. Treatment resulted in statistically and clinically significant change at post-test. Finally, partner hostility predicted worse end-state functioning whereas partner non-hostile criticism predicted better end-state functioning.

© 2006 Elsevier Ltd. All rights reserved.

*Keywords:* Cognitive-behavioral therapy; Generalized anxiety disorder; Exposure therapy; Couple functioning; Treatment outcome

---

## Introduction

Standard cognitive-behavioral therapy (CBT) for generalized anxiety disorder (GAD) typically consists of relaxation training (RT) and cognitive restructuring (CR), perhaps with some in-vivo situational exposure for patients with overt behavioral avoidance (e.g., Barlow, Rapee, & Brown, 1992; Borkovec & Costello, 1993; Butler, Fennell, Robson, & Gelder, 1991). The outcome literature clearly shows that standard CBT for GAD works as meta-analyses reveal that CBT is significantly more effective than wait-list and placebo control conditions (Borkovec & Whisman, 1996; Chambless & Gillis, 1993; Gould, Otto, Pollack, & Yap, 1997). Looking at the clinical significance of standard CBT packages for GAD, however, paints a more sobering picture. The most widely used strategy for assessing clinically significant change in this literature has been to

---

\*Corresponding author. Psychology Department, Northwestern University, 102 Swift Hall, Evanston, IL 60208-2710, USA.  
Tel.: +1 8474672290; fax: +1 8474917859.

E-mail address: [rzinbarg@northwestern.edu](mailto:rzinbarg@northwestern.edu) (R.E. Zinbarg).

classify patients according to whether they have achieved high end-state functioning (HES), which essentially relates to whether scores on outcome measures fall within the non-clinical range. Borkovec and Whisman (1996) obtained an average HES figure of 50% from their meta-analysis of standard CBT trials. Whereas this figure should not be interpreted precisely as the majority of these studies used different HES criteria, it seems clear that many patients with GAD do not experience clinically significant change from standard CBT packages.

Borkovec, Newman, Pincus, and Lytle (2002) increased therapy time by 50% and made other refinements of their CBT package. Despite these modifications, similar results were obtained, leading Borkovec et al. (2002) to conclude that “Our clinical research program has now spent 16 years attempting to refine, develop, and evaluate behavioral and CT [cognitive therapy] methods for treating GAD. Outcomes from the present study suggest that we need to look elsewhere for ways of incrementing the effectiveness of psychological treatment for this disorder. (p. 296)”. As they also found that both pre-treatment and post-treatment administrations of the Inventory of Interpersonal Problems—Circumplex Scales (Alden, Wiggins, & Pincus, 1990; Horowitz, Alden, Wiggins, & Pincus, 2000) were associated with poorer outcome at 6-month follow-up, these authors also concluded “there thus may be potential therapeutic value in adding some form of interpersonal therapy to the CBT package”.

That interpersonal difficulties should predict therapy outcome as found by Borkovec et al. (2002) is not surprising from the perspective of interpersonal theorists, who postulate that behavior often takes place in a relational context in which interactants exert mutual influence (e.g., Benjamin, 1993; Kiesler, 1996; Leary, 1957; Sullivan, 1953). Indeed, there is a vast literature documenting that the probability of a given behavior for a particular interactant is often conditional on the antecedent behavior of the other interactant (e.g., Gottman, 1980; Gurtman, 2001; Kenny & La Voie, 1984; Tracey, 1994). Both theory (e.g., Kiesler, 1996) and evidence (Kenny & La Voie, 1984) suggest, however, that such interactional contingencies are most likely to be observed in long-term relationships. Thus, it seems reasonable to hypothesize that problematic patterns of interpersonal transactions in long-term relationships are among the factors maintaining GAD symptoms for at least some patients and will therefore constrain the impact of therapy if they are not addressed in the therapy.

Not surprisingly, there is also evidence documenting that marriage or a marriage-like relationship is often the relationship that is the greatest source of both social support (e.g., Argyle, 1999; Argyle & Furnham, 1983; Denoff, 1982) and conflict (e.g., Argyle, 1999; Argyle & Furnham, 1983; Whisman, Sheldon, & Goering, 2000). Thus, the study of couple functioning would seem to be an important domain of inquiry in the quest to understand the potential interpersonal constraints that might impede the effectiveness of individual CBT for GAD.

The primary aim of this study was to test the validity of couple functioning variables as predictors of clinically meaningful individual treatment outcome in GAD patients in committed relationships.

### *Couple functioning and response to CBT for GAD and other disorders*

McLeod (1994) and Whisman (1999) found that wives with GAD reported significantly higher levels of marital distress than wives without GAD. Whisman et al. (2000) extended this finding by examining nine diagnoses and found that the strongest diagnostic correlate of marital dissatisfaction was GAD and this effect was not moderated by gender and was not attenuated by the use of dissatisfaction with other relationships as a covariate. Moreover, Kessler, Walters, and Forthofer (1998) found that GAD was associated with a significantly elevated risk of divorce.

There is also evidence suggesting that couple functioning variables predict response to individual CBT for GAD. Durham, Allan, and Hackett (1997) randomly assigned GAD patients to cognitive therapy, analytic therapy or behavior therapy and found that increased pre-treatment levels of self-reported marital tension significantly predicted a reduced likelihood of improvement across all three treatments.

Indirect evidence that couple functioning might predict response to individual CBT for GAD comes from studies of couple functioning and CBT for other anxiety disorders. For example, Craske, Burton, and Barlow (1989) found that couples' specific communication concerning panic disorder with agoraphobia (PDA) symptoms predicted outcome even though all of the couples in the study were largely satisfied with their

relationship. In reviewing this and similar studies of CBT for PDA, Daiuto, Baucom, Epstein, and Dutton (1998) concluded that many pre-treatment measures of general interactional styles and patterns (in areas such as communication, problem-solving, and cohesion) “have predicted outcome even when the patients were satisfied with their relationship and level of satisfaction did not predict outcome”.

Other studies have focused on hostility and criticism expressed toward patients by their spouses and other key relatives as predictors of response to CBT for PDA, OCD, and PTSD. Peter and Hand (1988) found that higher criticism by spouses predicted better long-term outcome in CBT for PDA. Tarrier, Sommerfield, and Pilgrim (1999) found that greater levels of hostility expressed toward the patient by relatives (72% of whom were spouses) predicted poorer treatment outcome in PTSD patients treated with either cognitive therapy or imaginal exposure therapy. Though the Tarrier et al. and Peter and Hand results appear to conflict, Chambless and Steketee (1999) obtained results among patients with either OCD or PDA that may resolve this conflict. On the one hand, Chambless and Steketee found that greater levels of hostility expressed toward the patient by relatives (73% of whom were spouses) predicted higher rates of dropout and poorer treatment outcome. On the other hand, they also found that higher rates of non-hostile criticism predicted better outcome. These results highlight that criticism need not be hostile and that hostility and non-hostile criticism seem to exert opposite effects.

Given that family members' criticism and hostility toward patients are two of the core facets of the construct of expressed emotion (EE), it might be hoped that the relatively mature literature on EE might offer some solid theoretical links between partner interactions and treatment response. Unfortunately, though there is a relatively large literature demonstrating that EE predicts long-term response in schizophrenia, depression and eating disorders (e.g., Butzlaff & Hooley, 1998), several authors have concluded that theoretical development has not kept pace with empirical advances in this area (e.g., Hooley, 1985; Tarrier et al., 1999). Some tentative suggestions have been outlined however and a few more may be offered.

Regarding links between hostility and outcome, one suggested mechanism is that hostility might activate and reinforce negative self-evaluations and negative core beliefs about the self and these negative evaluations and beliefs maintain/generate symptoms (e.g., Barrowclough, Tarrier, Humphreys, Ward, Gregg, & Andrews, 2003; Kuipers, 2006; Tarrier et al., 1999). Second, hostile interactions may be conceptualized as psychosocial stressors that interact with the diathesis underlying the patient's symptoms to produce continued symptom generation (e.g., Hooley, 2004; Tarrier et al., 1999). Third, we speculate that to the extent to which the desire to improve for loved ones can serve as one positive source of motivation for change in therapy, partner hostility would have obvious potential for undermining this source of motivation for change. Fourth, to the extent that the expression of symptoms tends to be followed by a temporary reductions in the probability of hostility as found by Biglan, Rothlind, Hops, and Sherman (1989) for the case of expressions of distress typical of depression, symptoms may even be maintained via negative reinforcement.

Turning to the potentially beneficial effects of non-hostile criticism, criticisms of avoidant tendencies that do not communicate rejection of the patients themselves may facilitate compliance with exposure to anxiety-provoking stimuli and activities (e.g., Chambless & Steketee, 1999; Peter & Hand, 1988). Finally, we speculate that a partner who is willing to disagree with the patient also might assist in CR by providing alternative, more balanced perspectives to negative thoughts and beliefs and by doing so without rejecting the patient also increases the likelihood that the patient will consider the partner's perspective.

### *Study hypotheses and overview*

This study sought to test two hypotheses that are based on the results of Chambless and Steketee (1999) and posit that high pre-treatment levels of partner hostility directed at the patient will predict worse end-state functioning, whereas high pre-treatment levels of non-hostile criticism will predict better end-state functioning. To test these hypotheses, we used an observational measure of couple interaction obtained at the time of the patient's initial assessment. As a manipulation check, we included a wait-list condition to confirm that our CBT program was efficacious. To increase statistical power, our hypotheses regarding hostility and non-hostile criticism as predictors of end-state functioning were tested by combining patients who received the treatment immediately with those who received treatment after completing the wait-list condition.

## Methods

### *Participants and procedure*

To recruit participants, we ran advertisements in local newspapers over the course of 4 weeks in May 2003 and June 2003 and on a local radio station over the course of 1 week in May 2003. We had also included information about the study on The Family Institute's website beginning in March 2003. To make our advertisements as specific to GAD as possible, we not only included information about the symptoms of GAD but also about some common GAD worry themes (e.g., money, work, kids, "minor, everyday activities"). The ads further specified that we were seeking patients with GAD and their partners to participate in a study of couple functioning and GAD in exchange for free individual GAD treatment. Phone screens were conducted by the first author and the Assistant Director of the Anxiety and Panic Treatment Program of The Family Institute at Northwestern University (Dr. Paula Young) to ensure that the individual was in a committed relationship and was likely to have GAD. Those callers meeting these preliminary criteria were scheduled for a videotaped diagnostic interview using the Structured Clinical Interview for DSM-IV-Clinical Version (SCID-CV; First, Spitzer, Gibbon, & Williams, 2002), with an advanced graduate student (described below). We conducted 41 phone screens and 28 diagnostic interviews. Twenty-one patients received a principal or co-principal diagnosis of GAD. Two patients dropped out of the study prior to randomization (one because she and her partner terminated their relationship, the other dropped out for unknown reasons as he never responded to any of our contact attempts following his diagnostic interview) leaving a total of 19 patients who were randomized to conditions by the first author. One patient in the treatment condition dropped out of the study after five sessions due to pregnancy complications that resulted in her being confined to bed rest. As the reason she dropped out of the study had nothing to do with her response to treatment, her data were eliminated from analysis. Thus, our analyses are based on a total of 18 patients, with eight in the treatment group and 10 who first went through the wait-list condition.

The 18 patients who completed the study averaged 41.94 years of age ( $SD = 12.23$ ), duration of the GAD problem averaged 23.41 years ( $SD = 13.87$ ) and the duration of their current relationship averaged 160.87 months ( $SD = 138.72$ , range = 6–444). Eleven patients were married, four were not currently married but were cohabiting with their partner, one was engaged to their partner but not cohabiting, and the remaining two were neither married to nor cohabiting with their partner but did consider the relationship to be a committed one. Ethnicity was represented by 15 White patients, 2 Hispanic patients, and 1 Asian patient. In terms of gross family incomes, two patients reported between \$21,000 and \$40,000, two reported between \$41,000 and \$60,000, eight reported between \$61,000 and \$100,000 and six reported more than \$100,000. Twelve patients were women. The two treatment conditions did not significantly differ on any of these demographic variables with the only variable even approaching significance being ethnicity (the treatment group included 62.5% White participants whereas 100% of the participants in the wait-list group were White; Fisher's exact test  $p = .069$ ).

In terms of psychiatric comorbidity, one patient had three comorbid current diagnoses and six patients had one comorbid current diagnosis. For three of the patients, their comorbid diagnosis was a co-principal diagnosis with GAD meaning that it was deemed to be causing equivalent levels of distress and impairment as the GAD diagnosis. Co-principal diagnoses were major depressive disorder (MDD;  $n = 1$ ), PDA ( $n = 1$ ) and specific phobia ( $n = 1$ ). Comorbid additional diagnoses were current MDD ( $n = 2$ ), PDA ( $n = 1$ ), social phobia ( $n = 2$ ), and OCD ( $n = 1$ ). In addition, three patients had a lifetime history of MDD but were not currently in a depressive episode and one patient had a lifetime history of alcohol abuse but had not met criteria for alcohol abuse for several years. There were also two patients who met the symptom criteria for additional current diagnoses including panic disorder ( $n = 1$ ) and social phobia ( $n = 1$ ), but the distress and impairment associated with these symptoms were rated as being questionable in terms of clinical significance. Seven patients were taking psychotropic drugs at the initial assessment; three were taking SSRIs, two were taking benzodiazepines, one was taking a tricyclic antidepressant, and one was taking a SSRI and Buspar. All of these characteristics were nearly equally distributed between the two conditions and were not significantly different.

After completing the SCID-CV (First et al., 2002) and pre-test versions of the questionnaires used as outcome measures (described below), the patient and their partner were mailed a packet of questionnaires about their relationship and were invited to come to the clinic for a videotaped couple interaction task. Each member of the couple received \$25 for completing the couple interaction task and \$10 for the questionnaires. Informed consent was obtained at the SCID-CV for patients and at the couple interaction task for the partners. All procedures were approved by the Northwestern University Internal Review Board.

The couple interaction task included a warm-up discussion (planning their next vacation together), two conflict resolution discussions, two worry discussions and a cool-down discussion (discussion about the three best things about their relationship or each partner). Only the discussions of the patients' worries were included in the analyses based on the assumption that interactions about the patients' worries would be most related to treatment response. For the worry discussion, the experimenter asked the couple to have a 10 min discussion of something coming up in the future that makes the GAD patient worried, anxious or nervous and how he or she might cope with it including what role, if any, the partner would play.

After completing the couple interaction task and relationship questionnaires, the patient was randomly allocated to either the treatment ( $n = 8$ ) or wait-list ( $n = 10$ ) condition. There were no dropouts in either group other than the one patient described above in the treatment condition. The outcome assessment battery administered at the end of the wait-list period and the treatment program was identical to that administered at the initial assessment except that only the GAD module of the diagnostic interview was administered at the post-test.

## Measures

### *Structured clinical interview for DSM-IV-Clinical version*

(SCID-CV; First et al., 2002). At the initial assessment, we used a version of the SCID designed to make lifetime and current diagnoses. At post-treatment (and post-wait list) mini-SCIDs were administered which included the GAD module only and interviewers were kept blind to group assignment. All SCIDs were administered by advanced clinical psychology graduate students and were reviewed by the first author to confirm the GAD diagnosis (the first author agreed with the principal diagnoses assigned by the interviewers on 26 of the 28 initial assessments conducted for this study).

### *Clinician Severity Rating*

(CSR; Di Nardo & Barlow, 1988). After completing each SCID interview, the interviewers rated the severity of each current diagnosis in the past month taking into account the number and frequency of symptoms, distress and impairment using the 0–8 CSR developed by Di Nardo and Barlow (1988). Scores of 1 and 2 indicate that some symptoms have been present in the past month but are clearly sub-clinical. A score of 3 indicates that symptoms have not only been present but may be clinically significant. Scores of 4 or above indicate symptoms associated with clinically significant distress or impairment.

### *Penn State Worry Questionnaire*

(PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990). The PSWQ consists of 16 items that assess the generality, excessiveness, and uncontrollability of worry without focusing on particular domains of worry and distress (e.g., “When I’m under pressure, I worry a lot,” “I worry all the time,” “Once I start worrying, I cannot stop”, “My worries overwhelm me”). Participants respond using a 4-point scale. The reliability and validity of the PSWQ have been widely researched, and the instrument has sound psychometric properties (Fresco, Frankel, Mennin, Turk, & Heimberg, 2002; Molina & Borkovec, 1994). At the initial assessment, participants were instructed to rate how much each item was true of them in general. At post-treatment (and post-wait list), participants were asked to rate how much each item was true of them in the past week.

### *Beck Anxiety Inventory*

(BAI; Beck, Epstein, Brown, & Steer, 1988). The BAI consists of 21 items, each describing a common symptom of anxiety. The participant is asked to rate how much he or she has been bothered by each symptom over the past week on a 0–3 scale. The measure has high internal consistency (Fydrich, Dowdall, &

Chambless, 1992) and high concurrent validity with other measures of anxiety (Steer, Ranieri, Beck, & Clark, 1993).

#### *Depression–Anxiety–Stress Scales–Stress and –Depression subscales*

(DASS-Stress and DASS-Depression) (Lovibond & Lovibond, 1995). The DASS-Stress scale assesses symptoms such as tension, irritability, difficulty relaxing, and a tendency to overreact to stressful events. The DASS-Depression scale assesses symptoms associated with depressed mood including sadness, self-deprecation, and anhedonia. Participants were asked to use 4-point severity/frequency scales to rate the extent to which they experienced each symptom over the past week. The internal consistencies of the DASS for GAD patients were .95 for DASS-Depression and .94 for DASS-Stress (Brown, Chorpita, Korotitsch, & Barlow, 1997). Brown et al. reported that GAD patients obtained significantly higher DASS-Stress scores than PDA, social phobia, OCD, and simple phobia. Thus, the DASS-Stress scale appears to be an index of symptoms that are relatively specific to GAD (in contrast Brown et al. did not find the DASS-Anxiety scale to be specific to GAD and was not included to reduce the burden on the participants). The DASS-Depression scale has demonstrated an acceptable pattern of convergent and discriminant validity (Antony, Bieling, Cox, Enns, & Swinson, 1998; Brown et al., 1997).

#### *Global satisfaction with relationship with partner*

All patients and their partners completed a relationship with partner questionnaire that included a global satisfaction item quite similar to the marital satisfaction item used by Whisman and his colleagues (Whisman, 1999; Whisman et al., 2000). Specifically, the global satisfaction item asked “Overall, how satisfied are you with your relationship with your partner?” This question was answered on a 5-point scale similar to the one used by Whisman et al. (2000). Response options and frequency of responses for patients and partners, respectively, were: “Very” (44.4%, 35.3%); “Quite” (28%, 33.3%), “Fairly” (11.1%, 5.9%); “Not Too” (0%, 5.9%); and “Not At All” (11.1%, 5.9%).

By way of comparison, the response options and frequency of responses in response to the question “During the past 6 months, how well have you gotten along with your spouse” in an epidemiological probability sample of 4933 participants reported by Whisman et al. (2000) were as follows: “Very well” (57%); “Quite well” (28%); “Fairly well” (13%); “Not too well” (2%); and “Not well at all” (0.4%). Thus, both the patients and their partners in the current sample appeared to report greater dissatisfaction with their relationship compared with Whisman et al.’s epidemiological sample. These differences were confirmed by two  $\chi^2$ -tests (each with d.f. = 4) comparing the distribution obtained by Whisman et al. with the distributions of the patients ( $\chi^2 = 14.13$ ,  $p \leq .01$ ) and their partners ( $\chi^2 = 19.35$ ,  $p \leq .001$ ) from the current sample. In addition, the percent of the GAD patients in the current sample reporting the lowest level of satisfaction with their relationship on our 5-point scale (11.1%) is quite similar to the percent of patients meeting current criteria for GAD reporting the lowest rating on a 4-point scale quality of their marriage (10.4%) in Whisman’s (personal communication, 1/6/2004) analysis of the National Comorbidity Study (Kessler et al., 1994) database.

#### *Hostility and non-hostile criticism from the Kategoriensystem für Partnerschaftliche Interaktion*

(KPI; Hahlweg & Conrad, 1983). The KPI is a coding system developed to gain an intimate understanding of the verbal and non-verbal behaviors that occur during couple’s dyadic interactions and designed to capture aspects of couples’ communication that are considered to be functionally important by marital therapy researchers and theorists. This coding system has shown good inter-rater reliability and convergent and criterion-related validity in previous research (Halford, Hahlweg, & Dunne, 1990; Hooley, 1986). The KPI has 10 verbal codes that are assigned to an individual’s response. In this study, we used only one of the negative verbal codes, Criticism (CR), which is divided into two sub-categories: Devaluation of the partner (CRD) and specific critique (CRS). CRD is noted when remarks contain negative judgments about the partner in the form of global accusations, insults, charges, as well as nasty personal remarks. This code also includes negative comments about the partner’s attributes in general or character traits. Thus, CRD appears to be tapping virtually the identical construct as the measure of hostility in the Camberwell Family Interview (CFI; Vaughn & Leff, 1976) used in the studies by Chambless & Steketee (1999), Peter and Hand (1988), and Tarrier et al.

(1999). CRS is noted when a remark clearly expresses disapproval of or disagreement with the partner's specific behaviors. Furthermore, we wanted to specify that 'non-hostile criticism' would be criticism without negative tone or gestures. Thus, we made a slight modification of the KPI coding system in which we further divided the CRS based on the non-verbal aspect of the critical comments. That is, the code for CRS was divided into CRS+/0 and CRS- based on how it was delivered. CRS+/0 was used to designate non-hostile criticism that is a specific criticism that was delivered in a positive or neutral way. In contrast, CRS- was used to designate specific criticism that was delivered in a negative or hostile manner.

The original KPI was designed to separate the interaction into speaking turns so that sequential analysis could be used. The present study, however, was not designed with the intent of using sequential analysis; rather, it was meant to examine global ratings of hostility and non-hostile criticism. Therefore, the 10-min discussions of each patient's worry were divided into 2-min units. During each 2-min block of time, ratings of CRS+, CRS- and CRD are made on scales ranging from 0 (absent) to 2 (marked). Raters were kept blind to the status of patients. The rating for each code was summed over five 2-min segments to determine the degree of the CRS+/0, CRS-, and CRD. Due to recording equipment failures, tapes were not available for two of the couples, leaving a final sample size of 16 for the analyses of these variables.

Coding of the interaction was completed by a group of four raters who were unaware of the diagnostic status of the members of the couples. The raters practiced rating tapes using the conflict discussions under the supervision of the second author until the team obtained an acceptable level of reliability as measured by an intraclass correlation (ICC) of .70 or greater. ICCs for the ratings of the patients' worry discussions used in the analyses reported equaled .83 for CRS+/0, .82 for CRS-, and .74 for CRD. As CRS- and CRD were highly correlated ( $r = .85$ ), we standardized them and then used their average as our measure of hostility to prevent multi-collinearity problems. CRS+/0 provided our measure of non-hostile criticism and was more modestly correlated with CRD ( $r = .32$ ), CRS- ( $r = .54$ ) and the hostility composite ( $r = .45$ ).

### *Therapists*

Three doctoral-level therapists conducted the therapy. Carol Donnelly, a clinical post-doctoral fellow with 2 years of training in and experience with CBT for anxiety, saw five of the patients in the therapy condition and seven wait-list patients. Paula Young, a licensed psychologist with 8 years of CBT experience, saw one patient in the therapy condition and one wait-list patient. The first author, a licensed psychologist with 15 years of CBT experience saw two patients in the therapy condition and two wait-list patients and listened to audiotapes of sessions and provided weekly individual supervision to Dr. Donnelly to maximize therapy adherence and quality.

### *Therapy conditions*

Patients in the wait-list condition were told that treatment would begin 16 weeks after randomization. At the end of the 16-week waiting period, they completed the post-test assessment. After completing the post-test assessment, these patients received the therapy after which they completed a second post-test assessment.

The treatment consisted of 12 60–75 min sessions over 16 weeks (the first 9 sessions were weekly, sessions 10 and 11 were bi-weekly and session 12 was scheduled 3 weeks after session 11). The treatment closely followed the Mastery of your Anxiety and Worry (MAW) treatment package that includes CR, RT and imagery exposure (IE) as major components (Craske & Barlow, 2006; Craske, Barlow, & O'Leary, 1992; Zinbarg, Craske, & Barlow, 1994, 2006). The only deviations from the MAW program were that we eliminated the time management and problem-solving components and modified the RT component (by starting with a 8 muscle-group rather than 12 muscle-group procedure to save some session time). Thus, the treatment consisted of (a) psychoeducation regarding the nature of GAD and the rationale for the treatment program (sessions 1 and 2), (b) RT (sessions 3 and 4 and portions of sessions thereafter), (c) CR (sessions 5, 6 and 7 and portions of sessions thereafter), (d) IE (sessions 8, 9, 10 and 11), and (e) plans for maintenance, relapse prevention and termination (sessions 11 and 12).

### Treatment integrity

Audiotapes or videotapes from three sessions (randomly selected from sessions 1–4, 5–8, and 9–12) for 16 patients (including 6 of the 8 patients in the treatment condition and all 10 of the patients in the wait-list condition who received the treatment after completing the waiting period) were checked for protocol adherence by two trained graduate students, who listened to the entire session and rated the therapist's interventions against a standardized checklist of allowed and not-allowed interventions (due to equipment failures, we did not have audible tapes for two of the patients). Our integrity checklist was adapted from one obtained from T. D. Borkovec (personal communication, August 30, 2002). Seven randomly selected cases were rated by both coders and the remaining cases were divided between the two. The coders agreed on 92% of their codes for the seven cases rated in common ( $\kappa = .74$ ). Of the 48 sessions that were rated, only two sessions contained violations (i.e., the therapists adhered to the protocol 95.85% of the time).

## Results

### Preliminary analyses

One patient in the treatment condition began taking medications in the last month of her therapy program and she stated that this was because she felt the therapy was not helping her. Her pre-test scores were therefore carried forward to her post-test scores in an endpoint score analysis. Table 1 presents means and standard deviations on the five outcome measures at pre-test and post-test and the mean improvement scores and standard deviations for the treatment and wait-list conditions. It is worth noting that the pre-test scores reported here are quite comparable with those reported in previous samples of GAD patients (e.g., the pre-test BAI scores are somewhat lower than in Butler et al. (1991) but slightly higher than those reported by Ladouceur et al. (2000) and Dugas et al. (2003); and the pre-test PSWQ scores are slightly higher than those reported in Ladouceur et al. (2000), Dugas et al. (2003), and Borkovec et al. (2002).

Given the importance of insuring that the groups were equivalent at the pre-test (i.e., that the treatment group was not less symptomatic to start with) for proper interpretation of group differences in post-test improvement, when testing for pre-test differences we decided to err on the side of possibly inflating our type I

Table 1

Means and standard deviations on outcome measures at pre-test and post-test for the wait-list ( $n = 10$ ) and treatment ( $n = 8$ ) groups

|                     | CSR           | PSWQ          | DASS-Stress   | BAI           | DASS-Depression |
|---------------------|---------------|---------------|---------------|---------------|-----------------|
| <i>Wait-list</i>    |               |               |               |               |                 |
| Pre-test            | 5.00 (1.05)   | 71.00 (5.58)  | 16.50 (6.04)  | 17.40 (7.32)  | 9.00 (6.62)     |
| Post-test           | 5.10 (1.05)   | 68.70 (6.82)  | 16.70 (6.99)  | 15.60 (5.76)  | 12.90 (8.25)    |
| Improvement         | -.10 (.99)    | 2.30 (5.23)   | -.20 (6.88)   | 1.80 (5.98)   | -3.90 (5.20)    |
| HES                 | .0%           | .0%           | 30.0%         | 50.0%         | 60.0%           |
| Reliable change     | 20.0% (40.0%) | 30.0% (0.0%)  | 20.0% (20.0%) | 10.0% (10.0%) | .0% (40.0%)     |
| <i>Treatment</i>    |               |               |               |               |                 |
| Pre-test            | 5.25 (.89)    | 69.83 (10.55) | 16.83 (8.13)  | 25.33 (14.81) | 11.83 (13.56)   |
| Post-test           | 2.38 (1.77)   | 52.60 (9.71)  | 9.48 (5.44)   | 12.92 (8.06)  | 4.98 (3.85)     |
| Improvement         | 2.87 (1.25)   | 17.23 (16.14) | 7.35 (8.29)   | 12.41 (9.07)  | 6.85 (11.18)    |
| HES                 | 62.5%         | 50.0%         | 75.0%         | 50.0%         | 87.5%           |
| Reliable change     | 87.5% (.0%)   | 57.1% (14.3%) | 57.1% (.0%)   | 57.1% (.0%)   | 28.6% (.0%)     |
| <i>d</i> -Post-test | 2.59          | 2.36          | 1.03          | .47           | .96             |

Note: CSR = Clinician Severity Rating of GAD based on SCID; PSWQ = Penn State Worry Questionnaire; DASS-Stress = Depression–Anxiety–Stress Scales Stress subscale; BAI = Beck Anxiety Inventory; DASS-Depression = Depression–Anxiety–Stress Scales Depression subscale. Reliable change values in parentheses show the percentage of participants showing reliable deterioration. *d*-Post-test was computed by subtracting the mean post-test score in the treatment group from the mean post-test score in the wait-list group and dividing the mean difference by the standard deviation of the post-test score in the wait-list group.

error rate for these analyses to minimize the likelihood of missing an effect. Thus, one-way analyses of variance (ANOVA) were conducted on each of the five pre-test measures separately, which indicated no significant differences between-group differences (all  $p$ 's  $\geq .17$ ). Moreover, if anything, the treatment group reported higher levels of distress on four of the five measures at the pre-test. Thus, there was clearly no evidence that the treatment group was less symptomatic than the wait-list group at the pre-test.

### Post-test improvement

#### Treatment versus wait-list

Improvement scores were created for each participant on each of the five outcome measures by subtracting post-test scores from pre-test scores. Thus, larger scores indicated greater improvement. To control our experiment-wise type I error rate and as the mean correlation among our improvement scores on the five outcome measures equaled .61 ( $M_d = .60$ ), the first step of our two-step main analyses began with an independent sample  $t$ -test using the average, standardized GAD improvement score as the dependent variable (DV). If the effect of treatment condition on this first step had not been significant, we would not have conducted individual independent sample  $t$ -tests on each of the five improvement scores separately.

Whereas a more conventional first step to this analysis would have been a MANOVA rather than a univariate test on a composite of the measures (e.g., Borkovec et al., 2002; Ladouceur et al., 2000), the power of MANOVA declines as the number of DVs and the across-groups associations increase (Stevens, 1996). Our composite approach retains the MANOVA virtue of protecting against the inflation of type I error that would otherwise accrue from the multiple tests of the individual measures. Rather than losing power as the correlations among the outcome measures increase, however, our composite approach gains power (while holding type I error constant) as the composite becomes a more reliable measure of the common latent variable.<sup>1</sup>

The average, standardized GAD improvement scores revealed that the treatment group showed more improvement ( $M = .70$ ,  $SD = .78$ ) than the wait-list group ( $M = -.48$ ,  $SD = .39$ ) and this difference was reliable,  $t(16) = 4.20$ ,  $p \leq .001$ .<sup>2</sup> Follow-up independent samples  $t$ -tests conducted to see which, if any, of the individual components of the composite measure were making a significant contribution to this effect showed significantly more improvement in the treatment group on the CSR,  $t(16) = 5.64$ ,  $p \leq .001$ ; PSWQ,  $t(16) = 2.43$ ,  $p \leq .05$ ; BAI,  $t(16) = 2.69$ ,  $p \leq .05$ ; and DASS-Depression,  $t(16) = 2.59$ ,  $p \leq .05$ . The groups did not significantly differ in improvement on the DASS-Stress,  $t(16) = 1.96$ ,  $p = .07$ . Using the Smith and Glass ES formula, an ES based on post-test scores was computed by subtracting the mean treatment group post-test score from the mean wait-list group post-test score and dividing this difference by the standard deviation of the wait-list group post-test score. These ES estimates are shown in the bottom row of Table 1 and the mean across all five measures was large ( $M = 1.48$ ).

To measure the degree of clinically meaningful gains, end-state functioning was calculated on each of the five outcome measures at post-test according to whether the patient either fell within one standard deviation of the mean of normative samples (PSWQ, BAI, DASS-Stress, and DASS-Depression) or a score that exceeded a face-valid level of meaningful change on the CSR for which norms are not available (i.e., a score of 2 which is associated with the anchor of "some symptoms have been present in the past month but clearly not clinically

<sup>1</sup>Perhaps the most conventional way to analyze these data would have been a 5 (measure)  $\times$  2 (time: pre versus post)  $\times$  2 (group) repeated measures (RM) MANOVA followed by 2 (time)  $\times$  2 (group) RM ANOVAs. However, such analyses would have yielded identical conclusions as those reached here given the equivalence of comparisons of differences scores across groups with RM analyses as clearly demonstrated by Brogan and Kutner (1980). We report the analyses of the improvement scores to keep our emphasis on change.

<sup>2</sup>Given our small sample size, we re-ran this  $t$ -test as a regression analysis with treatment group as a dummy coded predictor variable and examined the influence diagnostic statistics. No cases had Cook's  $D$  values above the cutoff based on the  $F$  distribution value for  $\alpha = .50$  suggested by Cohen, Cohen, West, and Aiken (2003) which in this case was a value of .72 (the largest Cook's  $D$  equaled .42); nor did any cases have DFBETA values exceeding the cutoff of  $\pm 1$  suggested by Cohen et al. with the largest DFBETAs equaling .42 for the intercept, and .83 for group. Similarly, there were not any cases with highly discrepant leverage values compared with the other cases (10 cases had leverage values of .044, the other 8 cases had leverage values of .069). However, there were two cases—both in the treatment group—that had externally studentized residual values exceeding the cutoff of  $\pm 2$  suggested by Cohen et al. When these two cases were deleted, the results were not drastically altered. The effect of treatment group was still significant,  $t(14) = 5.31$ ,  $p \leq .001$ , with the treatment group continuing to show more improvement ( $M = .67$ ,  $SD = .48$ ) than the wait-list group.

significant”). As seen in Table 1, this analysis revealed that the percentage of the treatment group meeting this criterion ranged from 50% (on the PSWQ and BAI) to 87.5% (on the DASS-Depression) and these outcomes are particularly impressive in comparison to the wait-list group on the CSR, PSWQ and DASS-Tension subscale. To measure the percentage of patients showing reliable change (RC), the Jacobson and Truax (1991) RC index was calculated on each of the five outcome measures with values greater than or equal to 1.96 indicating reliable improvement and values less than or equal to  $-1.96$  indicating reliable deterioration. Table 1 shows that the percentage of the treatment group showing reliable improvement on the four anxiety measures (i.e., CSR, PSWQ, DASS-Tension, BAI) ranged from 57.1% to 87.5%. It is also notable that whereas the percentage showing reliable improvement on the DASS-Depression was modest (28.6%) in the treatment group, none of them showed reliable deterioration on this measure whereas no one in the wait-list showed reliable improvement on this measure with a sizable percentage (40%) showing reliable deterioration.

### Total sample

The 10 participants in the wait-list condition were offered the same treatment following the 4-month delay period. Post wait-list scores were used as their pre-test scores. One patient in the wait-list condition completed his post-test assessment after completing the waiting period, and completed the mini-SCID when he subsequently completed the therapy program but never returned his questionnaire battery after completing therapy despite several phone calls and mailings from the research team over the course of 3 months. His scores were carried forward from the questionnaires he turned in at his final therapy session to his post-test scores in an endpoint score analysis. Thus, the following analyses are based on the total sample of 18 participants including those who received the treatment after the 4-month delay period.

A 2 (group)  $\times$  2 (time)  $\times$  5 (measure) repeated measures (RM) MANVOA with group as a between-subjects factor and time and measure as within-subjects factors revealed that the main effect of group and all interactions involving group were not significant. As there were no significant effects of group, Table 2 presents means and standard deviations on the five outcome measures at pre-test and post-test for all 18 treated individuals aggregated across the two groups. The main effect of time,  $F(1, 17) = 18.06$ ,  $p \leq .001$ , and the time  $\times$  measure interaction,  $F(4, 14) = 4.09$ ,  $p \leq .05$ , were significant. Follow-up RM ANOVAs indicated significant improvement on all five measures: CSR,  $F(17) = 54.50$ ,  $p \leq .001$ ; PSWQ,  $F(17) = 14.94$ ,  $p \leq .001$ ; DASS-Stress,  $F(17) = 8.38$ ,  $p \leq .01$ ; BAI,  $F(17) = 15.85$ ,  $p \leq .001$ ; and DASS-Depression,  $F(17) = 5.69$ ,  $p \leq .05$ . An ES was computed for each measure by subtracting the mean post-test score from the mean pre-test score and dividing the mean difference by the standard deviation of the pre-test scores. These ES estimates are shown in the bottom row of Table 2 and the mean ES across all five measures was 1.71. End-state and RC results on each measure for the total sample are shown in Table 2 and are largely congruent with those shown in Table 1 for the treatment group.

Table 2

Means and standard deviations on outcome measures at pre-test and post-test for all participants ( $n = 18$ )

|                 | CSR         | PSWQ          | DASS-Stress  | BAI           | DASS-Depression |
|-----------------|-------------|---------------|--------------|---------------|-----------------|
| Pre-test        | 5.17 (.79)  | 69.17 (7.83)  | 16.75 (6.94) | 19.61 (10.56) | 12.46 (9.80)    |
| Post-test       | 2.56 (1.82) | 54.82 (9.71)  | 10.38 (5.44) | 11.52 (6.60)  | 5.64 (3.46)     |
| Improvement     | 2.61 (1.50) | 14.35 (14.37) | 6.37 (8.80)  | 8.09 (8.10)   | 6.82 (9.64)     |
| HES             | 61.1%       | 44.4%         | 61.1%        | 66.7%         | 88.9%           |
| Reliable change | 83.3% (.0%) | 52.9% (5.9%)  | 52.9% (5.9%) | 41.2% (.0%)   | 31.3% (6.3%)    |
| <i>d</i>        | 3.30        | 1.83          | .92          | .77           | .70             |

Note: CSR = Clinician Severity Rating of GAD based on SCID; PSWQ = Penn State Worry Questionnaire; DASS-Stress = Depression–Anxiety–Stress Scales Stress subscale; BAI = Beck Anxiety Inventory; DASS-Depression = Depression–Anxiety–Stress Scales Depression subscale. Reliable change values in parentheses show the percentage of participants showing reliable deterioration. *d* was computed by subtracting the mean post-test score from the mean pre-test score and dividing the mean difference by the standard deviation of the pre-test scores.

### End-state functioning in the total sample

An overall end-state functioning measure was used to measure the breadth of clinically meaningful gains. Thus, overall end-state functioning was calculated by summing the number of the five outcome measures on which the patient achieved the end-state criteria described above. Twelve (66.7%) patients achieved an end-state functioning score of 3 or higher whereas nine (50%) patients achieved a score of 4 or higher.

To test whether hostility and non-hostile criticism predicted treatment response, we entered these two variables as predictors a simultaneous multiple regression among the 16 patients for whom we had audible tapes of their couple interactions. To identify predictors of clinically meaningful gains, we used overall end-state functioning scores as the DV in this analysis. Together hostility and non-hostile criticism accounted for 41% of the variance in overall end-state functioning in this sample (with an estimate of the proportion of variance accounted for adjusted for shrinkage equal to 32%), which was statistically significant ( $p \leq .05$ ). The unique effects of both hostility,  $\beta = -.65$ , semi-partial  $r = -.58$ ,  $\leq .05$ , and non-hostile criticism,  $\beta = .56$ , semi-partial  $r = .50$ ,  $p \leq .05$ , were significant.<sup>3</sup> (the partial regression plots for hostility and non-hostile criticism with overall end-state functioning are available upon request from the first author).

We also calculated a pre-treatment functioning score by summing the number of the pre-treatment measures on which the patient fell within one standard deviation of the mean of normative samples on the PSWQ, BAI, DASS-Stress, and DASS-Depression (a pre-treatment CSR score of at least 4 was required for inclusion in the study). Next, we repeated the regression analysis with the pre-treatment functioning score as a third predictor in addition to hostility and non-hostile criticism.<sup>4</sup> The results for hostility and non-hostile criticism (hostility:  $\beta = -.62$ , semi-partial:  $r = -.50$ ,  $p \leq .05$ ; non-hostile criticism:  $\beta = .62$ , semi-partial  $r = .51$ ,  $p \leq .05$ ) were virtually identical to the analysis that did not include pre-treatment functioning in the model. In contrast, the unique effect of pre-treatment functioning was small and did not even approach significance,  $\beta = .11$ , semi-partial  $r = .11$ ,  $p = .64$ .

### Discussion

Our manipulation check confirmed that the CBT package consisting of the RT, CR and IE components of the MAW led to statistically and clinically significant reductions in GAD symptoms. However, just as in previous end-state functioning analyses in this area, whereas our end-state functioning and reliable change results showed that the treatment was clearly effective for many patients, they just as clearly showed that many patients did not experience clinically significant change. This variability enabled a meaningful test of the hypotheses that pre-treatment levels of partner hostility and non-hostile criticism directed at the patient would predict variability in end-state functioning and both hypotheses were supported. More specifically, as hypothesized higher pre-treatment levels of partner hostility directed at the patient predicted worse end-state functioning whereas higher levels of non-hostile criticism predicted better end-state functioning.

The associations of pre-treatment partner hostility and non-hostile criticism with end-state functioning constitute a conceptual replication of [Chambless and Steketee \(1999\)](#) and suggest an additional direction to pursue in efforts to improve upon the clinical significance of psychotherapy for GAD. Several studies have already shown that involving the partner in treatment for PDA can lead to superior outcomes (e.g., [Barlow,](#)

<sup>3</sup>For similar reasons as in the analyses of the effects of treatment group on outcome, influence diagnostic statistics were closely examined. No cases had a significant externally studentized residual. Similarly, no cases had Cook's  $D$  values above the cutoff based on the  $F$  distribution value for  $\alpha = .50$  suggested by [Cohen et al. \(2003\)](#) which in this case was a value of .83 (the largest Cook's  $D$  equaled .35); nor did any cases have DFBETA values exceeding the cutoff of  $\pm 1$  suggested by Cohen et al. with the largest DFBETAs equaling .68 for the intercept, .27 for hostility, and .46 for non-hostile criticism. Two cases did have much higher leverage value than the other cases identifying them as potentially influential cases. When these two cases were deleted, the results were not drastically altered. Together hostility and non-hostile criticism still accounted for 35% of the variance in end-state functioning, with unique ES estimates of  $\beta = -.39$  and semi-partial  $r = -.39$  for hostility and  $\beta = .40$  and semi-partial  $r = .40$  for non-hostile criticism.

<sup>4</sup>Multicollinearity diagnostic statistics indicated that multicollinearity did not appear to unduly influence the results of either this regression or the previous one that did not include pre-test functioning (all variance inflation factors  $\leq 1.47$ ; all tolerances  $\geq .68$ ; all condition indices  $\leq 18.86$ ). Regarding the correlations among the variables: end-state functioning correlated .24 with pre-test functioning, .27 with non-hostile criticism, and  $-.40$  with hostility; pre-test functioning correlated  $-.09$  with non-hostile criticism and  $-.35$  with hostility, and non-hostile criticism correlated .45 with hostility.

O'Brien, & Last, 1984; Cerny, Barlow, Craske, & Himadi, 1987), with some findings suggesting that the beneficial effects of doing so may result from interventions designed to produce changes in the partners' interaction patterns (Daiuto et al., 1998). The findings reported here add to a growing body of evidence suggesting that couple therapy may also have potential for incrementing the effectiveness of CBT for GAD (e.g., Durham et al., 1997; McLeod, 1994; Whisman, 1999; Whisman et al., 2000).

The results reported here also begin to suggest what the targets of a couple therapy module for GAD might include. In particular, the finding that non-hostile criticism predicted better outcome suggests a target for couple intervention that may not be obvious. That is, it may be beneficial to not only decrease hostility and devaluation when present but to also increase non-hostile criticism for couples who are low on that variable. Of course, there might be a third variable that accounts for both high hostility and low non-hostile criticism on the one hand and poor response to individual CBT on the other hand. Thus, a treatment study in which GAD patients receiving individual CBT whose partners are either high on hostility or low on non-hostile criticism or both are randomized to either receive a couple intervention targeting these dyadic variables or to an appropriate control condition would not only test the intervention's clinical utility but would also be helpful for making causal inferences in this area.

The present study has several limitations. Perhaps the major limitation of this study is its relatively small sample size. Though we would argue that two (i.e., inadequate power, vulnerability to influential cases) of the three potentially deleterious effects associated with small sample sizes were not problematic, the third cannot be dismissed. That is, our small sample size precluded inclusion of other potential predictors of outcome in our regression analyses. The literature suggests that other potentially potent interpersonal predictors of the outcome of CBT for GAD include the patient's perception of partner criticism (Chambless & Steketee, 1999) and patient self-report of interpersonal problems (Borkovec et al., 2002). Thus, questions such as whether observer ratings of partner hostility and non-hostile criticism predict outcome above and beyond these other interpersonal predictors or sociodemographic variables (such as age and gender) or clinical variables (such as the presence/severity of comorbid conditions and medication status) will have to be addressed in future studies with larger sample sizes.

Another major limitation is that we did not collect follow-up data and a limitation that also applied to the study by Ladoceur et al. (2000) is that inter-rater agreement of diagnosis was obtained by having a second clinician review the original diagnostic interview. As noted by Ladoceur et al., whereas a review of the original diagnostic interview is better than not having any check on the diagnosis, having each clinician conduct an independent diagnostic interview would have been better.

Another potential limitation is that the recruitment methods may have introduced selection bias. To increase the sample size of couples in which at least one of the partners had GAD, the advertisements for this study specifically mentioned that couples were being recruited. Individuals with GAD in the most troubled relationships either might have been under-represented because their partners were not willing to participate or they might have been over-represented because they expected the study would help reduce some of their relationship problems and so may have been more motivated than other individuals with GAD to participate. Evidence bearing on this question comes from the comparisons of the global rating of marital satisfaction data from the present study with similar data from the samples studied by Whisman (1999) and Whisman et al. (2000). These comparisons suggested that relationship dissatisfaction in our sample was quite commensurate with that reported by patients with GAD in epidemiological studies. On the other hand, these single-item global ratings of marital satisfaction are very coarse measures and it might be that a more fine-grained measurement of relationship functioning would reveal that the relationships of our sample of patients with GAD differ in important ways from relationships in the population of patients with GAD who are in committed relationships. It may well be, however, that the population of patients and their partners willing to respond to advertisements similar to those used in this study would be precisely the population who would also be willing to participate in treatments involving a couple intervention. If so, the associations reported here should be helpful for informing the design of interventions for patients with GAD who are willing to participate in a couple intervention.

In summary, though this study has limitations, it provided evidence supporting the conclusion that, at least among couples willing to participate in research, partner's hostility and non-hostile criticism are significant predictors of end-state functioning. Thus, studies of these two variables together with other potential outcome

predictors and efforts to develop and test a couple intervention for GAD might also be fruitful directions aimed at increasing the likelihood that patients with GAD will experience clinically meaningful change.

## Acknowledgments

Preparation of this article was supported by the Nielsen Research Chair Endowment from The Family Institute at Northwestern University to Richard E. Zinbarg. We thank Emily Durbin, Eli Finkel, Greg Friedman, Richard Heyman, Lynne Knobloch-Fedders, Jay Lebow, William Pincus, and Paula Young who contributed to this article through many conversations about this study with the first author. We also thank T. D. Borkovec for providing a copy of his adherence rating checklist, Carol Donnelly for treating many of the patients at very low cost to the study, and all of the undergraduate and graduate student research assistants who helped with all phases of this study. Portions of this article are based on a thesis submitted by Jeong Eun Lee in partial fulfillment of the Master of Arts degree at Northwestern University.

## References

- Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, *55*, 521–536.
- Argyle, M. (1999). Causes and correlates of happiness. In: Kahneman, D., Diener, E. & Schwarz, N., (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 353–373). New York, NY: Russell Sage Foundation.
- Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychological Assessment*, *10*, 176–181.
- Argyle, M., & Furnham, A. (1983). Sources of satisfaction and conflict in long-term relationships. *Journal of Marriage and The Family*, *45*, 481–493.
- Barlow, D. H., O'Brien, G. T., & Last, C. G. (1984). Couples treatment of agoraphobia. *Behavior Therapy*, *15*, 41–58.
- Barlow, D. H., Rapee, R. M., & Brown, T. A. (1992). Behavioral treatment of generalized anxiety disorder. *Behavior Therapy*, *23*, 551–570.
- Barrowclough, C., Tarrier, N., Humphreys, L., Ward, J., Gregg, L., & Andrews, B. (2003). Self-esteem in schizophrenia: Relationships between self-evaluation, family attitudes, and symptomatology. *Journal of Abnormal Psychology*, *112*, 92–99.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*, 893–897.
- Benjamin, L. S. (1993). *Interpersonal diagnosis and treatment of personality disorders*. New York, NY: Guilford.
- Biglan, A., Rothlind, J., Hops, H., & Sherman, L. (1989). Impact of distressed and aggressive behavior. *Journal of Abnormal Psychology*, *98*, 218–228.
- Borkovec, T. D., & Costello, E. (1993). Efficacy of applied relaxation and cognitive-behavioral therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology*, *61*, 611–619.
- Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology*, *70*, 288–298.
- Borkovec, T. D., & Whisman, M. A. (1996). Psychosocial treatment for generalized anxiety disorder. In M. R. Mavissakalian, & R. F. Prien (Eds.), *Long-term treatments of anxiety disorders* (pp. 171–199). Washington, DC, USA: American Psychiatric Association.
- Brogan, D. R., & Kutner, M. H. (1980). Comparative analyses of pretest–posttest research designs. *American Statistician*, *34*, 229–232.
- Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, *35*, 79–89.
- Butler, G., Fennell, M., Robson, P., & Gelder, M. (1991). Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology*, *59*, 167–175.
- Butzlaff, R., & Hooley, J. (1998). Expressed emotion and psychiatric relapse: A meta-analysis. *Archives of General Psychiatry*, *55*, 547–552.
- Cerny, J. A., Barlow, D. H., Craske, M. G., & Himadi, W. G. (1987). Couples treatment of agoraphobia: A two-year follow-up. *Behavior Therapy*, *18*, 401–415.
- Chambless, D. L., & Gillis, M. M. (1993). Cognitive therapy of anxiety disorders. *Journal of Consulting and Clinical Psychology*, *61*, 248–260.
- Chambless, D. L., & Steketee, G. (1999). Expressed emotion and behavior therapy outcome: A prospective study with obsessive-compulsive and agoraphobic outpatients. *Journal of Consulting and Clinical Psychology*, *67*, 658–665.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: L. Erlbaum Associates.
- Craske, M. G., & Barlow, D. H. (2006). *Mastery of your anxiety and worry* (2nd ed.). York, NY: Oxford University Press.
- Craske, M. G., Barlow, D. H., & O'Leary, T. A. (1992). *Mastery of your anxiety and worry*. San Antonio, TX: The Psychological Corporation.

- Craske, M. G., Burton, T., & Barlow, D. H. (1989). Relationships among measures of marital satisfaction, and exposure during couples treatment of agoraphobia. *Behaviour Research and Therapy*, 27, 131–140.
- Daiuto, A. D., Baucom, D. H., Epstein, N., & Dutton, S. S. (1998). The application of behavioral couples therapy to the assessment and treatment of agoraphobia: Implications of empirical research. *Clinical Psychology Review*, 18, 663–687.
- Denoff, M. S. (1982). The differentiation of supportive functions among network members: An empirical inquiry. *Journal of Social Service Research*, 5, 45–59.
- Di Nardo, P. A., & Barlow, D. H. (1988). *Anxiety Disorders Interview Schedule-Revised (ADIS-R)*. Albany, NY: Phobia and Anxiety Disorders Clinic, State University of New York.
- Dugas, M. J., Ladouceur, R., Leger, E., Freeston, M. H., Langolis, F., Provencher, M. D., et al. (2003). Group cognitive-behavioral therapy for generalized anxiety disorder: Treatment outcome and long-term follow-up. *Journal of Consulting and Clinical Psychology*, 71, 821–825.
- Durham, R. C., Allan, T., & Hackett, C. A. (1997). On predicting improvement and relapse in generalized anxiety disorder following psychotherapy. *British Journal of Clinical Psychology*, 36, 101–119.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM-IV-TR Axis I disorders, research version, patient edition with psychotic screen (SCID-I/P W/ PSY SCREEN)*. New York: Biometrics Research, New York State Psychiatric Institute.
- Fresco, D. M., Frankel, A. N., Mennin, D. S., Turk, C. L., & Heimberg, R. G. (2002). Distinct and overlapping features of rumination and worry: The relationship of cognitive production to negative affective states. *Cognitive Therapy and Research*, 26, 179–188.
- Fydrich, T., Dowdall, D., & Chambless, D. L. (1992). Reliability and validity of the beck anxiety inventory. *Journal of Anxiety Disorders*, 6, 55–61.
- Gottman, J. M. (1980). Consistency of nonverbal affect and affect reciprocity in marital interaction. *Journal of Consulting and Clinical Psychology*, 48, 711–717.
- Gould, R. A., Otto, M. W., Pollack, M. H., & Yap, L. (1997). Cognitive behavioral and pharmacological treatment of generalized anxiety disorder: A preliminary meta-analysis. *Behavior Therapy*, 28, 285–305.
- Gurtman, M. B. (2001). Interpersonal complementarity: Integrating interpersonal measurement with interpersonal models. *Journal of Counseling Psychology*, 48, 97–110.
- Hahlweg, K., Conrad, M. (1983). *Coding manual: Kategoriensystem Zur beobachtung partnerschaftlicher Interaktion [Interaction Coding System]*. Unpublished manual available from Kurt Hahlweg, Institut für Psychologie, TU Braunschweig, Spielmannstrasse 12a, 38106 Braunschweig, Germany.
- Halford, W. K., Hahlweg, K., & Dunne, M. (1990). The cross-cultural consistency of marital communication associated with marital distress. *Journal of Marriage and The Family*, 52, 487–500.
- Hooley, J. M. (1985). Expressed emotion: A review of the critical literature. *Clinical Psychology Review*, 5, 119–139.
- Hooley, J. M. (1985). Expressed emotion and depression: Interactions between patients and high-versus low-expressed emotion spouses. *Journal of Abnormal Psychology*, 95, 237–246.
- Hooley, J. M. (2004). Do psychiatric patients do better clinically if they live with certain kinds of families? *Current Directions in Psychological Science*, 13, 202–205.
- Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *IIP-64/IIP-32 professional manual*. San Antonio, TX: The Psychological Corporation.
- 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.
- Kenny, D. A., & La Voie, L. (1984). The social relations model. *Advances in Experimental Social Psychology*, 18, 141–182.
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eschleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry*, 51, 8–19.
- Kessler, R. C., Walters, E. E., & Forthofer, M. S. (1998). The social consequences of psychiatric disorders, III: Probability of marital stability. *American Journal of Psychiatry*, 155, 1092–1096.
- Kiesler, D. J. (1996). *Contemporary interpersonal theory & research: Personality, psychopathology and psychotherapy*. New York: Wiley.
- Kuipers, E. (2006). Family interventions in schizophrenia: Evidence for efficacy and proposed mechanisms of change. *Journal of Family Therapy*, 28, 73–80.
- Ladouceur, R., Dugas, M. J., Freeston, M. H., Leger, E., Gagnon, F., & Thibodeau, N. (2000). Efficacy of a cognitive-behavioral treatment for generalized anxiety disorder: Evaluation in a controlled clinical trial. *Journal of Consulting and Clinical Psychology*, 68, 957–964.
- Leary, T. F. (1957). *Interpersonal diagnosis of personality*. New York: Ronald.
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33, 335–343.
- McLeod, J. D. (1994). Anxiety disorders and marital quality. *Journal of Abnormal Psychology*, 103, 767–776.
- Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28, 487–495.
- Molina, S., & Borkovec, T. D. (1994). The Penn State Worry Questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey, & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 265–283). Oxford, England: Wiley.

- Peter, H., & Hand, I. (1988). Patterns of patient–spouse interaction in agoraphobics: Assessment by Camberwell Family Interview (CFI) and impact on outcome of self-exposure treatment. In I. Hand, & H.-U. Wittchen (Eds.), *Panic and phobias: 2. Treatments and variables affecting course and outcome* (pp. 240–251). Berlin, Germany: Springer.
- Steer, R. A., Ranieri, W. F., Beck, A. T., & Clark, D. A. (1993). Further evidence for the validity of the Beck Anxiety Inventory with psychiatric outpatients. *Journal of Anxiety Disorders*, 7, 195–205.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: Norton.
- Tarrier, N., Sommerfield, C., & Pilgrim, H. (1999). Relatives' expressed emotion (EE) and PTSD treatment outcome. *Psychological Medicine*, 29, 801–811.
- Tracey, T. J. (1994). An examination of the complementarity of interpersonal behavior. *Journal of Social and Personality Psychology*, 67, 864–878.
- Vaughn, C. E., & Leff, J. P. (1976). The measurement of expressed emotion in the families of psychiatric patients. *British Journal of Social and Clinical Psychology*, 15, 157–165.
- Whisman, M. A. (1999). Marital dissatisfaction and psychiatric disorders: Results from the national comorbidity survey. *Journal of Abnormal Psychology*, 108, 701–706.
- Whisman, M. A., Sheldon, C. T., & Goering, P. (2000). Psychiatric disorders and dissatisfaction with social relationships: Does type of relationship matter? *Journal of Abnormal Psychology*, 109, 803–808.
- Zinbarg, R., Craske, M., & Barlow, D. H. (1994). *Therapist's guide for the mastery of your anxiety and worry program*. Albany, NY: Graywind Publishing Co.
- Zinbarg, R., Craske, M., & Barlow, D. H. (2006). *Therapist's guide for the mastery of your anxiety and worry program*. New York, NY: Oxford University Press.