



PERGAMON

Journal of Behavior Therapy  
and Experimental Psychiatry 31 (2000) 73–86

JOURNAL OF  
behavior  
therapy and  
experimental  
psychiatry

www.elsevier.com/locate/jbtp

# Psychometric properties of the credibility/ expectancy questionnaire<sup>☆</sup>

Grant J. Devilly<sup>a,\*</sup>, Thomas D. Borkovec<sup>b</sup>

<sup>a</sup>*Department of Criminology, University of Melbourne, Parkville, VIC 3010, Australia*

<sup>b</sup>*Pennsylvania State University, Stress & Anxiety Disorders Institute, University Park, Pennsylvania, USA*

Accepted 23 June 2000

---

## Abstract

The present research evaluated the psychometric properties of the credibility/expectancy questionnaire, a quick and easy-to-administer scale for measuring treatment expectancy and rationale credibility for use in clinical outcome studies. The results suggested that this questionnaire derives the two predicted factors (cognitively based credibility and relatively more affectively based expectancy) and that these factors are stable across different populations. Furthermore, the questionnaire demonstrated high internal consistency within each factor and good test–retest reliability. The expectancy factor predicted outcome on some measures, whereas the credibility factor was unrelated to outcome. The questionnaire is appended to the paper, yet the authors stress care when utilizing the scale. During the administration of the questionnaire, the participant sees two sections — one related to thinking and one related to feeling. However, the researcher needs to be aware that the 2 factors derived are not grouped into those questions. Instead credibility was found to be derived from the first three think questions and expectancy was derived from the fourth think question and the two feel questions. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Treatment; Credibility; Expectancy; Rationale; Questionnaire; Outcome; Psychometric

---

The need to assess therapy credibility and client expectancy for improvement is of paramount importance in therapy investigations (Jacobson & Baucom, 1977; Kazdin, 1979). Unless comparison conditions are equivalent on these factors in the early stages of treatment, investigators cannot rule out differential levels of credibility and/or expectancy (two potentially impacting variables contained within the nonspecific

---

<sup>☆</sup> Manuscript published under editorship of former editor L. Reyna.

\* Corresponding author. Tel.: + 61-3-8344-9440; fax: + 61-3-8349-4259.

E-mail address: dev@criminology.unimelb.edu.au (G.J. Devilly).

features of any therapy) as alternative explanations for any differences in observed outcome between the compared conditions.

The importance of assessing these two variables is highlighted by three domains of past research. First, when comparing rationales for nonspecific and other control conditions, some therapies have been shown to be significantly more credible and to generate greater expectancy among analogue participants (Borkovec & Nau, 1972; Lent, 1983). Indeed, Kazdin and Wilcoxon (1976), in their early review of the systematic desensitization literature, concluded that the absence of these assessments precluded the rejection of differential credibility/expectancy as an alternative explanation of results in the vast majority of investigations evaluating what was otherwise considered at that time to be a well-validated therapy. Second, credibility and expectancy have been demonstrated to be significantly influenced by systematic manipulation of such rationale features as length, the amount of information provided, and the kind of language used (Horvath, 1990; Kazdin & Krouse, 1983). Third, these two constructs have at times been found to significantly predict outcome. For example, expectancy has been shown to correlate with actual therapy outcome among such groups as inpatient Vietnam veterans (Collins & Hyer, 1986), social phobics (Chambless, Tran & Glass, 1997), and individuals suffering from generalized anxiety disorder (Borkovec & Costello, 1993). Although credibility has been found less often to be predictive of outcome, it has also been empirically associated with both simulated change (Nau, Caputo & Borkovec, 1974) and actual therapeutic improvement (Kirsch & Henry, 1977). In fact, it can be seen from past studies that credibility is less likely to correlate with outcome measures/assessment moments than expectancy, and this adds to the validity of examining the hypothesis that these construct are indeed two separate constructs. For example, in the Borkovec and Matthews (1988) comparative treatment study of nonphobic anxiety disorders credibility did not correlate significantly with any of the 10 outcome measures given at post and at each of two follow-up assessments (0 out of 30 possible correlations). Expectancy, on the other hand, correlated with 13 of the possible 30 measures. This phenomena was once again noticed in the Borkovec and Costello (1993) study where credibility only correlated significantly with 3 of the 11 outcome measures given at post and at each of two follow-up assessments (3 out of 33 possible correlations). Expectancy, on the other hand, correlated with 14 of the possible 33 measures/assessment moments.

Because of the above findings, many therapy outcome investigations now employ some form of credibility and/or expectancy assessment to ensure initial equivalence among compared therapy conditions. However, many different versions of assessment have been used, no psychometric properties have been established for any one of these versions, and no research has yet addressed whether credibility and expectancy scales are measuring the same or different constructs. The fact that expectancy has been found to predict outcome more often than credibility suggests that the distinction may be functionally important, even though investigators sometimes use the two terms interchangeably. This research builds upon the basic scale (Borkovec & Nau, 1972) which is the most frequently used credibility/expectancy measure in therapy research. While this research is not a complete validation of the scale, looking at its properties is thus worthwhile, independent of any future validation work.

Credibility has been defined as “how believable, convincing, and logical the treatment is,” whereas expectancy refers to “improvements that clients believe will be achieved” (Kazdin, 1979, p. 82). Both definitions use the term, “believe”. A belief, however, contains both cognitive and affective components. What a person logically thinks is the case may differ from what is felt to be the case. Therapists often confront this distinction when clients in early cognitive therapy point out that they see the logic of alternative, rational thoughts but that their older anxiety or depression-provoking thoughts simply feel more true. It was our suspicion that credibility scales tap more into the logical thought processes of our clients, whereas the expectancy scale (“How much do you expect to improve?”) is functionally more related to affective processes similar to those involved in hope or faith.

The purpose of the present study was to evaluate in three clinical samples some of the psychometric properties (factor structure, internal consistency, prediction of outcome, and test-retest reliability) of a specific credibility/expectancy questionnaire employed in prior outcome investigations (Borkovec & Costello, 1993; Devilly & Spence, 1999; Devilly, submitted for publication) and to examine whether measures of the two constructs would reveal their nature in relation to each other. To accomplish the latter purpose, a new questionnaire was developed. The first section contained the three credibility scales (logicalness, success in reducing symptoms, confidence in recommending the therapy to a friend) and the one expectancy scale from the prior questionnaire. The second section contained one of these credibility items (success in reducing symptoms) and the same expectancy scale. The two sections differed in the instructions that were provided. Introductory instructions told the client that beliefs about how well the therapy might help contains both thoughts and feelings about the therapy and that these may be the same or different. The first section of the questionnaire was to be answered in terms of what the client thought about the treatment. The second section was to be answered in terms of what the client felt about the treatment. Our expectation was that responses to the three credibility items rated in terms of what the client thought to be the case (first section) would be highly interrelated, that responses on the one credibility item rated on the basis of what the client felt (second section) and on both expectancy items (one in terms of thoughts, the other in terms of feelings) would be highly interrelated, and that factor analysis would reveal these two sets of responses to load on separate factors.

## **1. Design**

Participants in the current research came from 3 studies — one evaluating the efficacy of a lifestyle management course for Vietnam veterans and their spouses (Devilly, submitted for publication), one from an outcome study investigating the treatment of generalized anxiety disorder (Borkovec, Newman, Pincus, Lytle & Abel, in preparation), and one from a treatment outcome study investigating the relative efficacy of cognitive behavioural therapy (CBT) and Eye Movement Desensitization and Reprocessing (EMDR) in the treatment of PTSD (Devilly & Spence, 1999).

## 2. Study 1

### 2.1. Method

*Participants:* The 126 participants in study 1 included 68 male Vietnam veterans and 58 female spouses, with an overall average age of 49.88 years ( $SD = 5.10$ ). Three participants had incomplete data, leaving 123 for the analysis.

*Measures:* Only a brief outline of the measures used is presented here and the measures sections below to conserve space; readers are directed to the core references provided for further description and psychometric data. The Depression Anxiety Stress Scale - 42 (DASS-42; Lovibond & Lovibond, 1995), Novaco Anger Inventory-Short Form (NAI; Novaco, 1975), Abbreviated Dyadic Adjustment Scale (ADAS; Sharpley & Rogers, 1984), The Impact of Events Scale (IES; Horowitz, Wilner & Alvarez, 1979).

*Procedure:* Participants in study 1 were those who had attended a week long residential program, organized by the Brisbane Vietnam Veterans Counselling Service, that focussed on improving marital relationships and family functioning, managing anxiety, depression and anger, and generally increasing the quality of life for this population. All participants were administered the credibility/expectancy questionnaire (CEQ) on the first day, before any intervention but after the programme outline and rationale had been delivered. Program trainers were kept blind to the CEQ results in order to maintain confidentiality and to encourage honest responding. The CEQ for study 1 was adjusted, replacing the word ‘therapy’ with ‘course’, and also to reflect the intent of the programme (i.e. to increase general functioning).

### 2.2. Results

*Factor structure:* Descriptive statistics and tests for normality were conducted among the items to ensure no violations of technique assumptions were evident. The item inter-correlation matrix<sup>1</sup> displayed a number of significant correlations and suggested that the questionnaire may indeed have structure which could be detected by factor analysis. One hundred and twenty three cases completing the 6 CEQ items also afforded sufficient power to conduct a principal components analysis. Initial statistics indicated 2 factors with eigenvalues above 1 (see Table 1), and these 2 factors accounted for 82.46% of the total variance.

As it was theoretically consistent that treatment credibility and expectancy would not be orthogonally related, in fact the two factors derived a correlation of  $r = 0.41$ , an oblique rotation of the factors was performed. The resulting pattern matrix is presented in Table 2. As can be seen, the communalities indicated that the extracted factors accounted for between 74 and 93% of that item’s variance (e.g. about 87% of the variance from item 1 is accounted for by the 2 factors), with questions 4, 5, and

---

<sup>1</sup> Inter-correlation matrix is available from the first author upon request.

Table 1  
Factor descriptives from study 1

Factor	Eigenvalue	% of variance	Cumulative %
1	3.42	56.97	56.97
2	1.53	25.49	82.46

Table 2  
CEQ factor structure from study 1

Item	Factor 1	Factor 2	Comm	Mean	$\sigma$
1. How logical does the therapy offered to you seem	0.02	0.93	0.87	6.59	1.89
2. How successfully do you think this treatment will be in reducing your symptoms	0.34	0.79	0.74	6.21	1.68
3. How confident would you be in recommending this treatment to a friend ...	0.2	0.87	0.8	6.48	1.94
4. How much improvement in your symptoms do you think will occur	0.89	0.16	0.82	52.76	23.23
5. How much do you really <i>feel</i> that therapy will help you to reduce your symptoms	0.82	0.33	0.79	5.68	1.8
6. How much improvement in your symptoms do you really <i>feel</i> will occur	0.96	0.07	0.93	50.57	24.74

6 loading on factor 1 (expectancy) and questions 1, 2, and 3 loading on factor 2 (credibility).

*Reliability:* In order to assess internal consistency, the scales were analysed by each factor to derive Cronbach's  $\alpha$ , and each item's total correlation with the factor was examined. For the expectancy factor, question 4 derived a total item correlation of  $r = 0.83$ , question 5 correlated at  $r = 0.75$ , and question 6 at  $r = 0.85$ . As this factor contains items using different scales (see below) a standardized alpha was derived, and this equalled 0.90. The credibility factor derived Cronbach's  $\alpha = 0.86$ , and a total item correlation of  $r = 0.78$  for question 1,  $r = 0.69$  for question 2, and  $r = 0.76$  for question 3. The total scale (both factors) standardised alpha was  $r = 0.85$ .

*Predicted outcome:* The CEQ utilizes two scales during the administration (1–9, and 0–100%), and so a composite score was derived for each factor (expectancy and credibility) by first standardizing the individual items and then summing those items for each factor. As this study utilized only one condition, discriminative validity could not be addressed. However, prediction of outcome was assessed by correlating the factor scores with changes in functioning at post-intervention and at 3 month follow-up. It should be kept in mind that this intervention was not intended to be curative for any specific presentation and also that the post-intervention data were collected directly at the end of the week-long, residential course and may reflect

satisfaction with the programme as opposed to real changes in indicators of functioning. The 3 month follow-up data were collected by mail. To allow the greatest amount of latitude to discover trends and yet not commit a Type II error, Bonferroni corrections were not applied; instead, the  $\alpha$  level was set to 0.01 for significance. There were no significant correlations obtained, although expectancy displayed a trend for significance in change scores from pre- to post-intervention on the anxiety subscale of the DASS,  $r(120) = 0.20$ ,  $p < 0.04$ , and on the IES,  $r(64) = 0.26$ ,  $p < 0.05$ .

### 3. Study 2

#### 3.1. Method

*Participants:* Participants in study 2 involved 69 diagnosed GAD clients who had received 14 sessions of treatment in one of three randomly determined conditions (cognitive therapy, applied relaxation and self-control desensitization, or a combination of these therapies). There were 15 women and 8 men in each condition. Mean age for the group was 37.1 years ( $SD = 11.8$ ), and average chronicity was 12.8 years ( $SD = 12.2$ ). Data from two clients were incomplete, leaving 67 for analysis.

*Measures:* Again, only a brief outline of the measures employed have been presented here. The Hamilton Anxiety Rating Scale (HARS; Hamilton, 1959), assessor severity score of generalized anxiety disorder symptoms (0–8 pt. scale; Barlow et al., 1984), Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990), a reactions to relaxation questionnaire (RRQ; Borkovec & Costello, 1993), the trait measure of the Spielberger State-Trait Anxiety Inventory (STAI-Y2; Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983), daily diary level of anxiety severity (0–100 pt. scale; Barlow et al., 1984), Hamilton Depression Rating Scale (HDRS; Hamilton, 1964), Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock & Erbaugh, 1961).

*Procedure:* Clients in study 2 completed the CEQ at the end of each of their 14 sessions. Their completed forms were sealed in an envelope with assurances that their therapist would not see their responses. Data from the first and second sessions were used for calculating test–retest reliability for the week intervening between these two sessions. Session 1 involved presentation of the rationale for their therapy, training in self-monitoring for identifying early cues of anxiety, and (for clients in applied relaxation and combined conditions) a brief demonstration of how to generate a relaxation response using slowed, paced breathing. Session 2 proceeded with further discussion of self-monitoring and an introduction to cognitive therapy (for clients in cognitive therapy and combined conditions) and initial training in progressive relaxation (for clients in applied relaxation and combined conditions). The word “anxiety” was used instead of the word “trauma” throughout the questionnaire.

#### 3.2. Results

*Factor structure:* Again, the item inter-correlation matrix displayed a number of significant correlations and suggested that the questionnaire had structure which

Table 3  
Factor descriptives from study 2

Factor	Eigenvalue	% of variance	Cumulative %
1	3.36	55.97	55.99
2	1.06	17.7	73.69

Table 4  
Factor structure from study 2

Item	Factor 1	Factor 2	Comm	Mean	$\sigma$
1. How logical does the therapy offered to you seem	0.16	0.84 <sup>a</sup>	0.72	8.12	1.04
2. How successfully do you think this treatment will be in reducing your symptoms	0.52	0.71 <sup>a</sup>	0.78	7.31	1.45
3. How confident would you be in recommending this treatment to a friend ...	0.18	0.88 <sup>a</sup>	0.8	7.4	1.56
4. How much improvement in your symptoms do you think will occur	0.79 <sup>a</sup>	0.23	0.68	67.61	17.24
5. How much do you really <i>feel</i> that therapy will help you to reduce your symptoms	0.70 <sup>a</sup>	0.3	0.58	6.67	1.48
6. How much improvement in your symptoms do you really <i>feel</i> will occur	0.93 <sup>a</sup>	0.11	0.87	66.79	18.06

<sup>a</sup>Indicates items presumed to load on each factor.

could be detected using factor analysis. Sixty seven completed cases of the 6 CEQ items also afforded sufficient power in order to conduct a principal components analysis. Initial statistics indicated 2 factors with eigenvalues above 1 (see Table 3), and these 2 factors account for 73.69% of the total variance.

After an oblique rotation, the communalities indicated that the extracted factors accounted for between 58 and 87% of that item's variance (e.g. about 72% of the variance from item 1 is accounted for by the two factors) (see Table 4).

This factor structure appears to be very similar to the structure derived from study 1. To support this observation a confirmatory factor analysis was conducted. Expectancy and credibility are theoretically related and, therefore, the 2 factors were correlated in the model. Residual variables were also expected to correlate with items 2 and 5 conceptually connected, as well as items 4 and 6. Eleven iterations were performed to successfully converge the model, and a chi-square test displayed that there was not a significant discrepancy between the model and the data ( $\chi^2(6, N = 63) = 12.13, p < 0.06$ ). Goodness of fit indices of 0.98 (population gamma) and 0.95 (Joreskog goodness of fit index) were produced with this solution, suggesting that the model is a good approximation of the data.

*Reliability:* The same process as for study 1 was applied to study 2 to assess internal consistency. The expectancy factor derived a standardized alpha coefficient of 0.79.

Question 4 obtained a total item correlation of  $r = 0.68$ ,  $r = 0.53$  for question 5, and  $r = 0.71$  for question 6. The credibility factor derived Cronbach's  $\alpha = 0.81$ , and a total item correlation of  $r = 0.62$  for question 1,  $r = 0.71$  for question 2, and  $r = 0.71$  for question 3. The total scale (both factors) standardized  $\alpha$  equalled a similar result as study one,  $r = 0.84$ .

Test–retest reliability was assessed by correlating the factor scores at time 1 with the factor scores at time 2. The test–retest reliability of the expectancy factor was significant,  $r(67) = 0.82$ ,  $p < 0.001$ , and the credibility test–retest reliability also displayed a significant correlation,  $r(67) = 0.75$ ,  $p < 0.001$ . When these two factors were combined to produce an overall rating of the treatment rationale, the test–retest reliability of this composite index was likewise very high,  $r(67) = 0.83$ ,  $p < 0.001$ .

*Condition differences:* Time 1 factor scores of expectancy and credibility were obtained for each of the three conditions in the same manner as for study 1. A 3 (conditions)  $\times$  2 (factors) MANOVA displayed no difference between the treatment conditions on expectancy and credibility, Wilks'  $\Lambda(4, 126) = 0.98$ , ns. Thus, the combined CBT therapy package and its two-component-control conditions were equivalent in credibility and expectancy after the first therapy session.

*Prediction of outcome:* Prediction of outcome was assessed by correlating the factor scores with changes in outcome measures at post-treatment and also at 12 month follow-up. Again, to allow the greatest amount of latitude to discover trends yet not commit a Type II error, the  $\alpha$  level was set to 0.01 for significance. The only significant correlation for pre- to post-treatment change scores was with expectancy and daily diary level of anxiety severity,  $r(63) = 0.39$ ,  $p < 0.002$ . However, expectancy also displayed a trend towards significance with the HARS,  $r(67) = 0.28$ ,  $p < 0.03$ , and credibility displayed a weak trend for correlation with the daily diary,  $r(64) = 0.26$ ,  $p < 0.04$ . At 12 month follow-up, the only significant correlation was between expectancy and reactions to relaxation questionnaire,  $r(63) = 0.32$ ,  $p < 0.01$ . All other correlations, which displayed a trend towards significance, were with the expectancy factor (daily diary,  $r(63) = 0.31$ ,  $p < 0.02$ ; PSWQ,  $r(63) = 0.26$ ,  $p < 0.04$ ).

## 4. Study 3

### 4.1. Method

*Participants:* The 22 participants in study 3 included 11 clients in the CBT condition (5 males) and 11 clients in the EMDR condition (3 males). The average overall age of participants was 37.82 years ( $SD = 13.11$ ). Although the sample was too small for assessing psychometric properties, it provided an opportunity to evaluate credibility/expectancy differences between two overtly distinct therapeutic rationales.

*Measures:* The following measures were administered: The trait measure of the Spielberger State-Trait Anxiety Inventory (STAI-Y2; Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983), Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock & Erbaugh, 1961), Global Distress Scale of the Symptom Checklist-90-R (SCL-G; Derogatis, 1992), Subjective Units of Disturbance Scale (SUD; Wolpe, 1969),

Personal Problem Definition Questionnaire (PPD; Devilly, Creamer & Gournay, work in progress), Civilian Mississippi scale for PTSD (CMS; Keane, Caddell & Taylor, 1988), Impact of Events Scale (IES; Horowitz, Wilner & Alvarez, 1979), PTSD Symptom Scale-Self-Report (PSS-SR; Foa, Dancu & Rothbaum, 1993), PTSD Interview (PTSD-I; Watson, Juba, Manifold, Kucala & Anderson, 1991).

*Procedure:* Participants in study 3 were administered the CEQ after an intake interview had been completed and the treatment rationale for either CBT or EMDR had been presented. Again, the therapist was kept blind to the ratings of the participants.

#### 4.2. Results

*Condition differences:* The CBT and EMDR conditions in study 3 were compared on initial credibility and expectancy scores. As with study 1, a composite of standardized scores was used for each factor and independent *t*-tests comparing the two therapy conditions were applied to each of these measures. For expectancy, a significant difference was obtained,  $t(20) = 4.19$ ,  $p < 0.001$ , with the CBT protocol deriving a mean of 1.76 (SD = 1.78) and EMDR obtaining a mean score of  $-1.93$  (SD = 2.33). For credibility, a significant difference was also obtained,  $t(20) = 4.92$ ,  $p < 0.001$ , with CBT obtaining a mean score of 1.83 (SD = 1.13) and EMDR obtaining a mean score of  $-2.17$  (SD = 2.44).

*Prediction of outcome:* Factor scores were correlated with outcome changes, as in studies 1 and 2, again with  $\alpha = 0.01$ . Expectancy correlated with change scores from pre- to post-treatment on the SCL-G ( $r = 0.68$ ) and the IES ( $r = 0.65$ ), while credibility did not correlate with any changes in outcome. From pre-treatment to 3 month follow-up expectancy correlated with the SCL-G ( $r = 0.70$ ), BDI ( $r = 0.57$ ), Y2 ( $r = 0.60$ ), PPD ( $r = 0.71$ ), PSS-SR ( $r = 0.65$ ), CMS ( $r = 0.62$ ) and the IES ( $r = 0.65$ ), while credibility again did not correlate with any follow-up changes. However, as the CEQ discriminated between the two treatment conditions and the CBT condition evidenced a marked superiority in outcome (Deville & Spence, 1999), it was necessary to look at prediction of outcome *within* each condition. In the CBT condition expectancy correlated significantly only with pre- to post-treatment change in SCL-G ( $r = 0.74$ ) and pre-treatment to follow-up change in SCL-G ( $r = 0.77$ ). In the EMDR condition, expectancy did not correlate with any change in outcome measures at either time point. Credibility was unrelated to change in either condition at both assessment periods.

#### 5. Subsidiary analyses

Items 4 (Set I) and 2 (Set II) differ only with respect to “think” and “feel” as do items 2 (Set I) and 1 (Set II). With this in mind, a legitimate concern would be that these item pairs are measuring exactly the same construct and so an analysis of response patterns is warranted. Items 4 (Set I) and 2 (Set II) correlated at  $r = 0.83$  in study 1 and  $r = 0.68$  in study 2. However, dependent (correlated) samples *t*-tests suggest that there is

variability in the responses with a trend towards a difference in how people answered the two questions in study 1 ( $t(123) = 1.73, p < 0.09$ ), although not study 2 ( $t(67) = 0.64, ns$ ). In both studies the “feel” item was rated lower than the “think” item, and this issue of a possible repetition of measurement may warrant further investigation.

With respect to items 2 (Set I) and 1 (Set II), although they correlate in study 1 ( $r = 0.52$ ) and in study 2 ( $r = 0.55$ ), it is also evident that people answered the questions in significantly different ways in study 1 ( $t(124) = 3.43, p < 0.001$ ) and in study 2 there appeared to be a trend towards significance ( $t(68) = 1.83, p < 0.08$ ). Again, in both studies the “feel” items were rated lower than the “think” items.

## 6. Discussion

This research evaluated the psychometric properties of the credibility/expectancy questionnaire for measuring treatment expectancy and rationale credibility for use in clinical outcome studies. In accomplishing this goal three studies were presented that assessed the factor structure, internal consistency, test–retest reliability, and the scale’s ability to predict outcome and discriminate between rationales possessing unequal theoretical basis.

The results suggest that this scale derives the two predicted factors and that these factors are stable across different populations. Furthermore, this scale demonstrates high internal consistency within each factor with a standardized  $\alpha$  of between 0.79 and 0.90 for the expectancy factor, a Cronbach’s  $\alpha$  of between 0.81 and 0.86 for the credibility factor, and a standardized  $\alpha$  of between 0.84 and 0.85 for the whole scale. Inter-item correlations across studies ranged from between 0.53 and 0.85 for the items on the expectancy factor, and between 0.62 and 0.78 for items on the credibility factor. Test–retest reliability over a one-week period was also found to be good at 0.82 for expectancy and 0.75 for credibility. The CEQ also displayed the ability to differentiate between two treatment rationales in one study, one with and one without an encompassing theory, while maintaining equivalence between three rationales in another study. The scale’s ability to predict outcome still remains uncertain as two of the studies revealed expectancy to covary with *some* change scores, yet one study did not. However, the study which failed to display prediction of outcome was not one which targeted any specific symptomatic clusters but rather aimed at increasing “general functioning”. The studies which did display a correlation exhibited this on measures directly assessing the population-specific symptoms. Either way, it would appear that treatment expectancy contributes more to this concept than rationale credibility, a result congruent with past research (Borkovec & Costello, 1993).

Of particular interest is the notion that two separate, yet related, factors were derived from the questionnaire; namely treatment expectancy and rationale credibility. Such results are consistent with past research (Borkovec & Mathews, 1988; Borkovec & Costello, 1993). However, this research also initiates formalized attention

to the concerns of clinicians who see treatment credibility as not being sufficiently addressed in research (Addis, Wade & Hatgis, 1999).

It is suggested that future comparative research needs to address both credibility and expectancy when controlling for nonspecific factors. Furthermore, it would appear that asking for an affective reaction to the proposed therapy (i.e. “what do you really feel”) may obtain a better appreciation of the subjects’ expectancy. Future research may also investigate whether, in a larger questionnaire, “thinking” items on expectancy and “feeling” items on expectancy would be differentiated. It was not possible to investigate this in the current research due to only one “thinking” question on the expectancy factor. This research was conducted with archived data and so, unfortunately, no changes could be made in relation to the number and content of the questions, nor the scales used.

However, the fact that answers to the same “think” and “feel” questions (4 (Set I) and 2 (Set II); 2 (Set I) and 1 (Set II)) appear to be answered differentially further substantiates the notion that some potentially functional difference exists between “feel” and “think” credibility and expectancy, regardless of their degree of correlation.

This measure utilizes two rating scales, one from 1 to 9 and another from 0 to 100% and, therefore, the use of this measure in clinical outcome studies necessitates standardization of each item and the formation of a composite for each factor. While this may appear a complex and unusual method for deriving the two factors, such a process takes into account the different scales which were used in these studies and enables a direct comparison between two or more treatment conditions while still allowing the participant to make ecologically valid responses to the questions posed. Future research may look at changes which make the questionnaire more administrator friendly, including the rewording of items, and a full cross-over of think/feel and credibility/expectancy items. It would also be useful for a future study to investigate the actual content validity of the items. However, without the first step of presenting psychometric data as demonstrated here, and no other questionnaires with validation data, such a quest would be very difficult indeed. It was the aim of this research to pilot the CEQ and provide some psychometric data on the most commonly used expectancy/credibility questionnaire in therapy, as well as to render an informed direction for future research within this field.

## **Appendix A. Therapy evaluation form**

We would like you to indicate below how much you believe, *right now*, that the therapy you are receiving will help to reduce your anxiety. Belief usually has two aspects to it: (1) what one *thinks* will happen and (2) what one *feels* will happen. Sometimes these are similar; sometimes they are different. Please answer the questions below. In the first set, answer in terms of what you *think*. In the second set answer in terms of what you really and truly *feel*. We do not want your therapist to ever see these ratings, so please keep the sheet covered when you are done.



- 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., & Nau, S. D. (1972). Credibility of analogue therapy rationales. *Journal of Behaviour Therapy and Experimental Psychiatry*, 3, 257–260.
- Borkovec, T. D., Newman, M. G., Pincus, A., & Lytle, R. *A component analysis of cognitive behavioral therapy for generalized anxiety disorder*, (manuscript in preparation).
- Chambless, D. L., Tran, G. Q., & Glass, C. R. (1997). Predictors of response to cognitive-behavioral group therapy for social phobia. *Journal of Anxiety Disorders*, 11, 221–240.
- Collins, J. F., & Hyer, L. (1986). Treatment expectancy among psychiatric inpatients. *Journal of Clinical Psychology*, 42, 562–569.
- Derogatis, L. R. (1992). *SCL-90: Administration, scoring and procedures manual-II*. Baltimore. Clinical Psychometric Research.
- Devilly, G. J. *The psychological effects of a lifestyle management course on war veterans and their spouses*. Submitted for publication.
- Devilly, G. J., Creamer, M., & Gournay, K. *The psychometric characteristics of the Personal Problem Definition questionnaire*, (manuscript in preparation).
- Devilly, G. J., & Spence, S. H. (1999). The relative efficacy and treatment distress of EMDR and a cognitive behavior Trauma Treatment Protocol in the amelioration of post traumatic stress disorder. *Journal of Anxiety Disorders*, 13(1–2), 131–157.
- Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, 6, 459–473.
- Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32, 50–55.
- Hamilton, M. (1964). A rating scale for depressive disorders. *Psychological Reports*, 14, 914.
- Horowitz, M., Wilner, M., & Alvarez, W. (1979). Impact of event scale: A measure of subjective stress. *Psychosomatic Medicine*, 41, 209–218.
- Horvath, P. (1990). Treatment expectancy as a function of the amount of information presented in therapeutic rationales. *Journal of Clinical Psychology*, 46, 636–642.
- Jacobson, N. S., & Baucom, D. H. (1977). Design and assessment of nonspecific control groups in behavior modification research. *Behavior Therapy*, 8, 709–719.
- Kazdin, A. E. (1979). Therapy outcome questions requiring control of credibility and treatment-generated expectancies. *Behavior Therapy*, 10, 81–93.
- Kazdin, A. E., & Krouse, R. (1983). The impact of variations in treatment rationales on expectancies for therapeutic change. *Behavior Therapy*, 14, 657–671.
- Kazdin, A. E., & Wilcoxon, L. A. (1976). Systematic desensitization and nonspecific treatment effects: A methodological evaluation. *Psychological Bulletin*, 83, 729–758.
- Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi scale for combat-related PTSD: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, 56, 85–90.
- Kirsch, I., & Henry, D. (1977). Extinction versus credibility in the desensitization of speech anxiety. *Journal of Consulting and Clinical Psychology*, 45, 1052–1059.
- Lent, R. W. (1983). Perceptions of credibility of treatment and placebo by treated and quasi-control subjects. *Psychological Reports*, 52, 383–386.
- 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.
- 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.
- Nau, S. D., Caputo, J. A., & Borkovec, T. D. (1974). The relationship between credibility of therapy and simulated therapeutic effects. *Journal of Behavior Therapy and Experimental Psychiatry*, 5, 129–133.
- Novaco, R. W. (1975). *Anger control: The development and evaluation of an experimental treatment*. Lexington, MA: D.C. Heath.
- Sharpley, C. F., & Rogers, H. J. (1984). Preliminary validation of the Abbreviated Spanier Dyadic Adjustment Scale: Some psychometric data regarding a screening test of marital adjustment. *Educational and Psychological Measurement*, 44, 1045–1049.

- Spielberger, C. D., Gorsuch, R. L., Lushene, R. D., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto: Consulting Psychologists Press.
- Watson, C. G., Juba, M., Manifold, V., Kucala, T., & Anderson, P. E. D. (1991). The PTSD interview: Rationale, description, reliability, and concurrent validity of a DSM-111-based technique. *Journal of Clinical Psychology*, *47*, 179–188.
- Wolpe, J. (1969). *The practice of behavior therapy*. New York: Pergamon Press.