Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Behavior Therapy and Experimental Psychiatry 35 (2004) 319–336
JOURNAL OF behavior therapy and experimental psychiatry
www.elsevier.com/locate/jbtep

An approach to psychotherapy toleration: the Distress/Endorsement Validation Scale (DEVS) for clinical outcome studies

Grant J. Devilly*

Centre for Neuropsychology, Swinburne University, PO Box 218, Hawthorn, Vic 3122, Australia

Received 12 December 2003; received in revised form 2 June 2004; accepted 5 August 2004

Abstract

The issue of treatment tolerance within the field of psychotherapy is, at best, a nebulous construct and has been commonly evaluated via rates of subject attrition and homework compliance. This research presents the psychometric properties of a ten-item scale which endeavours to measure treatment distress and participant endorsement of therapy protocols used in clinical research. Two factors emerged and the subscales of Distress and Endorsement were derived. These subscales displayed good reliability with acceptable inter-item correlations within each subscale. The subscales were also able to differentiate the perspectives of male Vietnam veterans from their spouses on a lifestyle management course at the termination of intervention. However, this scale also displayed a cognitive behavioural trauma treatment protocol and eye movement desensitisation and reprocessing to be equivalent in treatment distress and participant endorsement in the treatment of post-traumatic stress disorder. Preliminary findings suggest that the relationship between these two subscales and outcome may, to some extent, be population specific. First evidence suggests that intervention distress ratings may be influenced by severity of presentation, whilst endorsement ratings are more influenced by symptomatic improvement over time. Suggestions for future research are presented and the full questionnaire is attached as an appendix.

Keywords: Treatment; Distress; Endorsement; Tolerance; Questionnaire; Outcome

*Tel.: +61-3-9214-5920; fax: +61-3-9214-5230.
E-mail address: gdevilly@swin.edu.au (G.J. Devilly).

0005-7916/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jbtep.2004.08.001

1. Introduction

The utility of psychological interventions for mental health disorders has predominantly been appraised through efficacy research (e.g. Chambless et al., 1998), with a growing interest in effectiveness research (see Nathan, Stuart, & Dolan, 2000) and evaluation of intervention efficiency or cost-effectiveness (e.g. see Yates, 1994). However, as intervention techniques display increasingly shared efficacy there has been a call to gauge evidence by the principles or process of change rather than the specific technique (Rosen & Davison, 2003). In this article we would like to reinforce another aspect in assessing the utility of an intervention: treatment toleration.

Historically, psycho-pharmaceutical research has been concerned with medication side-effects and other indices of participant toleration of the medications utilised (e.g. Martin, 1998; Hawley, Sivakumaran, Huber, & Ige, 1998). Specifically, research has looked at patient toleration of interventions for cancer and also subsequent treatment for these effects (e.g. Zeltzer & LeBaron, 1985). However, in recent years the issue of treatment toleration has become more prominent within the field of psychotherapy due to claims by advocates of one intervention or another that these techniques are "kinder" to clients and, therefore, better tolerated (e.g. Hyer & Brandsma, 1997). While the acceptance of the treatment process has been recently equated with compliance (Scott & Stradling, 1997) and frequently measured by participant attrition rates in past research, some have suggested that even those who do not discontinue treatment find some protocols more congenial than others (Boudewyns & Hyer, 1996; Pitman et al., 1996; Shapiro, 1989) and base this primarily on subjective interpretation of patient feedback and non-specific clinical experience.

Fear of intervention and treatment expectations have been investigated as "early in therapy" concepts and lead to the development of various scales administered at the initiation of intervention or when sampling community or specific samples (e.g. The Thoughts About Therapy Survey; Deane & Chamberlain, 1994). Other research has suggested the need to take into account participant personality variables to ensure that treatment is "comfortable" and "tolerated" (Harris & Bergman, 1987). However, to our knowledge, there are no "after-therapy" questionnaires, with known psychometric properties, available for researchers of psychological interventions to assess treatment toleration. This research presents the psychometric properties of a questionnaire which attempts to measure treatment distress and participant endorsement of the techniques used, once treatment has been delivered.

In an attempt to control non-specific treatment effects in comparative, clinical, outcome research, Borkovec and Nau (1972) developed a measure of rationale credibility and this has been further refined to include treatment expectancy (Borkovec & Costello, 1993; Devilly & Borkovec, 2000). These concepts, particularly treatment expectancy, have been found to be related to treatment outcome (Collins & Hyer, 1986; Chambless, Tran, & Glass, 1997; Kirsch & Henry, 1977) and possess the ability to differentiate between treatment rationales (Devilly & Borkovec, 2000). With a priori treatment expectancy and rationale credibility becoming a more

prominent feature in comparative research, and vigorous discourse regarding the efficacy of treatment studies, it is argued that treatment distress and participant endorsement of treatment regimens be also taken into account. However, it is suggested that this be accomplished by using a standardised assessment, with less therapist subjectivity than used in past studies and with an eye on client demand characteristics.

One of the major problems facing clinical researchers is the trade between economy of measurement (leading to participant compliance) and depth of measurement (providing a richer outcome set). Psychotherapy process measures traditionally evaluate therapist variables as well as other sessional variables such as patient psychic distress, and are typically scored by independent raters (e.g. The Vanderbilt Psychotherapy Process Scale; O'Malley, Suh, & Strupp, 1983). As well as some measures not assessing the individual's perception of the therapeutic experience, those that are available are usually lengthy and onerous to complete. With this in mind the goal of the present study was to develop and present a quick and easy-to-administer measure of treatment distress and participant endorsement. This is not to say that therapist process variables are not critical to treatment success (e.g. see Devilly & Gournay, 1995), but rather that when two treatment procedures are nearing equal efficacy, a primary factor in therapeutic selection may be participant toleration of the techniques at hand. This article describes the development of the questionnaire to its current version and presents the emerging psychometric properties. Being the first investigation of its kind, we expect this concept, and questionnaire, to evolve over time.

2. Method

2.1. Design

As with most questionnaires, this was developed and refined over time out of the need to assess treatment toleration and general satisfaction with interventions that were being provided for organisations and to individuals. As a consequence, the psychometric properties of the questionnaire encompasses two versions, which were developed over time. The first incarnation of the questionnaire (version 1, utilised in studies 1 and 2) contained only eight items. The second version (utilised in study 3) presents the properties of the full Distress/Endorsement Validation Scale (DEVS, Appendix A), which had grown to include 10 items. As will become clear, version 1 required more items due to the second subscale (Endorsement) being derived from only two items (i.e. a correlation rather than a subscale). Therefore, a second version was trialled with 10 items and the scale derived the two subscales.

2.2. Participants

Participants in the current research came from three studies: the first evaluating the efficacy of a lifestyle management course for Vietnam veterans and their spouses

(Devilly, 2002); the second from a treatment outcome study investigating the relative efficacy of Cognitive Behavioural Therapy (CBT) and Eye Movement Desensitisation and Reprocessing (EMDR) in the treatment of PTSD (Devilly & Spence, 1999); and the third a refined version of the first study using new participants.

The 209 participants in study 1 included 111 male Vietnam veterans with an average age of 51.19 years (sd = 4.13) and 98 female partners with an average age of 48.38 years (sd = 4.81). Three participants had incomplete data leaving 206 for the initial analysis.

The 23 participants in study 2 included 12 clients in the CBT condition (five males) and 11 clients in the EMDR condition (three males). The average overall age of participants was 37.96 years (sd = 12.61). Although the sample size was too small for assessing psychometric properties, it provided an opportunity to evaluate differences in ratings between two overtly distinct therapeutic rationales and somewhat different treatment approaches.

The 137 participants in study 3 included 73 male veterans and 62 female spouses. The overall age for males was 55.21 (sd = 4.25) and for females it was 52.13 (sd = 5.60).

2.3. Measures

Only a brief outline of the measures used is presented here to conserve space; readers are directed to the core references provided for further descriptions and psychometric data.

Questionnaires used in study 1: 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), Impact of Event Scale (IES; Horowitz, Wilner, & Alvarez, 1979), Credibility/Expectancy Questionnaire (CEQ; Borkovec & Costello, 1993; Devilly & Borkovec, 2000).

Questionnaires used in study 2: the trait measure of the Spielberger State-Trait Anxiety Inventory (STAI-Y2; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), Global Distress Scale of the Symptom Checklist-90-R (Derogatis, 1992), Subjective Units of Disturbance Scale (Wolpe, 1969), Personal Problem Definition Questionnaire (Devilly and Gournay, work in progress), Civilian Mississippi scale for PTSD (Keane, Caddell, & Taylor, 1988), IES (Horowitz et al., 1979), PTSD Symptom Scale—Self-Report (Foa, Riggs, Dancu, & Rothbaum, 1993), PTSD Interview (Watson, Juba, Manifold, Kucala, & Anderson, 1991), CEQ (Borkovec & Costello, 1993; Devilly & Borkovec, 2000).

Questionnaires used in study 3: the Medical Outcome Scale 36-Item Short-Form Health Survey (SF-36; Ware & Sherbourne, 1992), McMaster Family Assessment Device—General functioning subscale (McMaster; Epstein, Baldwin, & Bishop, 1983), General Health Questionnaire—28 Item (GHQ-28; Goldberg & Hillier, 1979).

2.4. Procedure

Participants in study 1 were those who had attended a week-long residential programme, organised by the Brisbane Vietnam Veterans Counselling Service. This programme focused 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 DEVS (version 1) on the last day of the course and immediately prior to departure, as part of a feedback process and post-intervention assessment. The results were kept confidential and programme trainers were specifically kept blind to the results in order to maintain confidentiality and to encourage honest responding. The DEVS for study 1 was adjusted, replacing the word "therapy" with "course". Outcome measures were obtained at post-intervention on the last day of the course and at 3-month and 6-month follow-up by postal mail.

Participants in study 2 were those who underwent an outcome study into the treatment of PTSD and were administered the DEVS (version 1) at a 2-week follow-up as part of an assessment battery by postal mail. It was hoped that by postal follow-up, the results of the whole assessment battery would be less influenced by therapist demand effects (see Devilly (2001) for a more detailed explanation of this hypothesis). Outcome measures were obtained at post-intervention and 2-week and 3-month follow-up.

Participants in study 3 were those who attended a week-long course run by the Vietnam veterans counselling service. This course was exactly similar to study 1, with the exception that different outcome measures were used. Outcome measures were obtained at 12-month follow-up. Participants were also administered the DEVS (full version 2) at this 12-month follow-up period.

Initial questions were devised using clinician suggestions and feedback. The author of this article originally developed a pool of questions pertinent to treatment toleration following experience as a clinician and with a knowledge of the research literature on clinical-outcome studies. A focus group of clinicians and researchers were then asked for more items to reflect their experience. Redundant items, or those that appeared to be asking the same question, and those that appeared to be asking about other things (e.g. therapist variables) were also removed. With the number of questionnaires administered during treatment trials increasing over the years, the questionnaire also needed to be short and easy to understand. Overall, it became clear that expert consensus appeared to concur with the author's hypothesis that treatment toleration was an amalgam of treatment distress and treatment endorsement. Thus, questions relating to these constructs were trialled. There were no further a priori reasons for item inclusion or exclusion.

3. Results

3.1. Study 1

Factor (*component*) *structure.* Descriptive statistics and tests for normality were conducted among the items to ensure no violations of technique assumptions. The

item intercorrelation matrix was considered to be "factorable" given the presence of a number of significant correlations. Two hundred and six cases completing the eight DEVS items also afforded sufficient power to conduct a principal components analysis. Initial statistics indicated two components (referred to as "factors" hereafter) with eigenvalues above 1 (4.09 and 1.33, respectively), and these two factors accounted for 67.62% of the total variance (51.06% from factor 1, 16.56% from factor 2).

These two factors appeared to be orthogonally related and a varimax rotation of the factors was performed. The resulting pattern matrix is presented in Table 1. As can be seen, the communalities indicated that the extracted factors accounted for between 46% and 81% of that item's variance (e.g. about 67% of the variance from item 1 is accounted for by the two factors), with questions 1–6 loading on "factor 1" (treatment distress) and questions 7 and 8 loading on "factor 2" (treatment endorsement). This derived second subscale can, therefore, be seen more as a correlation of two items measuring a face valid construct rather than a true subscale.

Reliability. In order to assess internal consistency, the scale was analysed by each derived subscale to assess Cronbach's alpha, and each item's total correlation with the subscale was examined. The treatment Distress subscale derived Cronbach's $\alpha = .90$, and a total item correlation of $r = .71, .80, .83, .79, .57$ and $.66$ for questions 1–6, respectively. The treatment Endorsement subscale displayed a correlation between the two items of $r(206) = .37$, $p < .0001$.

Table 1
DEVS factor structure from study 1

Item	Factor 1	Factor 2	Comm	Mean	σ
1. Distress experienced during first assessment session	.82	−.02	.67	4.54	2.44
2. Distress experienced during first treatment session	.89	.03	.79	4.83	2.52
3. After first therapy session, distress experienced for the next few hours	.89	.14	.81	4.17	2.41
4. How anxious about returning to therapy for second treatment session	.86	.10	.75	3.80	2.43
5. Overall, how intrusive was the therapy	.68	.02	.46	4.14	2.56
6. Overall, how distressing was the whole treatment	.73	.22	.57	4.25	2.28
7. If knew as much about the treatment technique as now, how inclined in still participating	−.13	−.79	.63	7.02	2.45
8. How likely in recommending this treatment to someone else…	−.01	−.85	.72	8.57	1.17

Note: Comm = communalities.

Relationship to outcome. Relationship to outcome was assessed by correlating the derived subscale scores (Distress and Endorsement) with changes in functioning from pre-intervention to post-intervention, 3-month follow-up and 6-month follow-up (i.e. Time 1–Time 2). 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- and 6-month follow-up data were collected by mail. Furthermore, these results are based upon correlation and the associations presented here are purely to elucidate the positive or negative direction of the correlation and are in no way meant to represent causative associations. To allow the greatest amount of latitude to discover trends and yet not fall prey to a Type I error, Bonferroni corrections were not applied; instead, the alpha level was set to .01 for significance. The only significant correlations ($\alpha = .01$) were with the IES (administered only to the veterans). Treatment Distress negatively correlated with IES change scores from pre- to post-intervention, suggesting the possibility that less improvement on a measure of trauma symptoms in the short term was associated ($r(105) = -.32$, $p < .01$, power = .96) with increasing distress during the intervention. Endorsement did not correlate with IES change scores from pre- to post-treatment ($r(104) = -.14$, ns). The only other significant correlation was between treatment endorsement and the IES at 6-month follow-up ($r(71) = -.34$, $p < .01$, power = .91). This may suggest that the more they improved, the more likely they were to recommend this course to others and the more inclined they would be in still participating if they knew as much about the course as they did at the end of the programme. Treatment distress did not correlate with IES change scores from pre-treatment to follow-up ($r(71) = -.15$, ns).

However, this presents an over-simplified view of the results as a more refined view emerges when gender is taken into account. There was a significant difference between males and females for both the distress ratings ($t(205) = 5.84$, $p < .01$) and the endorsement ratings ($t(204) = 2.39$, $p < .02$), with males rating higher on distress and endorsing the programme less than the females. All of the veterans were male and when only the data from these participants were analysed all of the correlations between distress and outcome (stress and IES) were negative (even those at $p < .05$; anxiety), suggesting that the less they changed for the better, the more distressing they found the course (or vice versa). It should be kept in mind that all of these correlations were at post-intervention—directly at the end of the course. However, the endorsement–outcome correlations were all positive, suggesting that the more they changed for the better over time, the more they endorsed the course. These correlations were all at the 3- and 6-month follow-ups (stress, anger and IES).

The spouses, on the other hand, had entirely negative correlations, suggesting that the more distressing and intrusive they found the course, the more they improved. These correlations were all in relation to change scores from pre-intervention to the 3- and 6-month follow-ups (notably, depression, anxiety and stress). There were no significant correlations for the females between change scores and endorsement

($p<.01$). Naturally, being correlations, we are not making any conclusions regarding direction or causation.

In order to clarify this point further, pre-intervention ratings on the outcome measures were correlated with the two subscales of Distress and Endorsement for the male veterans. It was found that measures of self-reported disturbance at intake, as quantified by ratings of depression, anxiety, stress, anger, and post-traumatic functioning, were all highly correlated with distress ratings for the intervention ($p<.01$). Only one correlation was found between endorsement and presentation—post-traumatic functioning as measured by the IES ($p<.01$).

3.2. Study 2

Relationship to outcome. Subscale scores were correlated with outcome changes, as in study 1, again with $\alpha = .01$. No significant correlations were found between treatment distress and changes in symptomatology. As in study 1, the Distress subscale correlated with presentation severity (at intake) on all the outcome measures ($p<.01$), with the exception of state anxiety as measured by the STAI ($r(23) = .45$, $p<.04$, power = .76), and trait anxiety ($r(23) = .28$, ns). No presentation severity measures correlated ($p<.01$) with the Endorsement subscale (two item composite). Neither distress nor endorsement correlated with change scores over time ($p<.01$). Due to the low degrees of freedom with such a small sample, this result is not surprising.

Condition differences. The CBT and EMDR conditions in study 2 were compared on participant ratings of treatment distress and treatment endorsement, using independent *t*-tests. There were no differences between the two conditions on participant ratings of distress ($t(21) = .25$, ns) nor on therapeutic endorsement ($t(21) = -.82$, ns).

3.3. Study 3

As mentioned above, two items measuring a hypothesised construct does not properly constitute a derived subscale. Two items were introduced into the DEVS—one that was hypothesised to load on the second factor (Endorsement subscale) and one on the first factor (Distress subscale). Item selections were based upon client feedback and consultation with a focus group of clinicians.

Factor (*component*) *structure.* Data screening displayed some skewness in what became the three items which loaded on factor two. To assess the effect of this on factor structure, these items were dichotomised into those who scored at the extreme (9; very inclined/likely/much) or scored anything else (1–8). The results of this procedure displayed these variables to possess no excessive skewness, but also demonstrated that there was no difference to the derived factor structure[1] as was obtained when such a dichotomising approach was not used (as below). As in study 1, a Varimax rotation was applied to a principal components analysis. Two factors

[1]Skewness data and the results of this factor analysis are available from the author upon request.

Table 2
DEVS factor structure from study 3

Item	Factor 1	Factor 2	Comm	Mean	σ
1. Distress experienced during first assessment session	.86	−.6	.74	4.19	2.48
2. Distress experienced during first treatment session	.86	−.01	.74	4.04	2.34
3. After first therapy session, distress experienced for the next few hours	.89	−.01	.79	3.87	2.18
4. How anxious about returning to therapy for second treatment session	.80	.13	.65	3.68	2.24
5. Overall, how intrusive was the therapy	.75	.23	.62	3.68	2.24
6. Overall, how distressing was the whole treatment	.86	.16	.77	3.95	2.13
7. Overall, how exhausting was the whole treatment	.67	.19	.48	5.11	2.35
8. If knew as much about treatment technique as now, how inclined in still participating	.10	.83	.70	6.73	2.25
9. How likely in recommending this treatment to someone else…	.05	.89	.80	7.90	1.74
10. Believe you received value for your time and money	.04	.89	.80	7.64	1.90

Note: Comm = communalities.

(components) were derived which accounted for 70.88% of the total variance (48.37% from factor 1, 22.51% from factor 2). The derived factor structure is presented in Table 2. It should be noted that an oblimin rotation displayed a correlation of only −.2 between the two components, and the pattern matrix made no difference to the interpretation of the questionnaire. Furthermore, the structure matrix showed that the items loading on Distress had small negative loadings on Endorsement, and the items loading on Endorsement had small negative loadings on Distress. This suggests that it is unlikely that there is a general (one) factor solution.

As with study 1, the communalities displayed a high percentage of the variance being accounted for by the factors on each item (48–80%), with only item seven scoring below 62%. Subscale 1, which we term "Distress", is composed of the first 7 items and subscale 2 (Endorsement) is composed of the last three items. When the two derived subscales were formed using this sample, they produced a small to medium correlation ($r(135) = -.21$, $p < .01$).

Reliability. As in study 1, in order to assess internal consistency, the scale was analysed by each factor to derive Cronbach's alpha, and each item's total correlation with the subscale was exa.mined. The treatment Distress subscale derived

Cronbach's $\alpha = .92$, and a total item correlation of $r = .77, .78, .84, .73, .69, .83$ and .60 for questions 1–7, respectively. The treatment Endorsement subscale derived Cronbach's $\alpha = .84$, and a total item correlation of $r = .67, .73$, and .74 for items 8, 9 and 10, respectively.

Relationship to outcome. Relationship to outcome was assessed by correlating the subscale scores with the outcome measures for the 12 month, Lifestyle Management, follow-up. As with study 1, these outcome measures were obtained via a postal follow-up and these correlations are not assumed to represent causation. Tests for normality suggested that the Distress subscale was normally distributed with a skewness of .38, a standard error of skewness equal to .21, and kurtosis of −.51. The Endorsement subscale, however, was not normally distributed displaying positive skew (1.53, skew standard error = .21) and kurtosis of 2.25. This was evident overall,and also within each gender. In effect, this sample greatly endorsed the lifestyle management programme, with a large number of individuals giving the course "full marks" on all three questions.

To reduce the chances of Type I errors, correlations were only conducted between the Distress and Endorsement subscales and the total scores of the various measures for the participants ($\alpha = .01$). There were no significant correlations between the Distress subscale and change scores for the outcome measures.

Endorsement, however, displayed significant correlations with changes in all the outcome measures, again with a small to medium effect size. There was a negative relationship with change scores (Time 1–Time 2) on the Mental Health Scale of the SF-36 ($r(121) = -.20$, $p < .03$, power = .72), and positive relationships with the McMasters Scale ($r(135) = .24$, $p < .01$, power = .89) and the GHQ ($r(122) = .19$, $p < .04$, power = .69). This suggests that the more the participants improved on the mental health subscale of the SF-36, the more they endorsed the intervention. Likewise, improvement on both the McMasters Family Functioning Scale and the GHQ is associated with increased endorsement of the course.

When gender was taken into account, it was found that there were no significant correlations between Distress or Endorsement and change scores on the outcome measures for females. With this in mind, it is not surprising that males, on the other hand, showed correlations for distress and endorsement on the outcome measures as above. Naturally, for males, the strength of the relationships is more evident between distress and the Mental Health Scale of the SF-36 ($r(65) = .33$, $p < .01$, power = .87) and between endorsement and the SF-36 ($r(65) = -.31$, $p < .02$, power = .83), the McMasters Scale ($r(71) = .35$, $p < .01$, power = .93), and the GHQ total ($r(67) = .26$, $p < .04$, Power = .70).

In contrast to a gender effect at post-intervention in study 1, there was no significant difference between male and female ratings of distress ($t(134) = 1.70$, ns) or endorsement ($t(137) = .18$, ns) for this study at the 12-month follow-up on the full 10-item questionnaire. To check whether this difference was more likely to be due to the different times of administration between the two studies or the addition of the two extra questions, independent *t*-tests were also applied to the Distress and Endorsement subscales comprising only the original eight items. This made no difference to the lack of statistical significance. Furthermore, unlike study 1, ratings

of disturbance, as measured by self-reported symptomatology at intake, did not correlate with the Distress subscale (or the Endorsement subscale) of the DEVS at 12-month follow-up. This was also the case when only the eight items of the original scale were used in the two subscale scores.

3.4. All studies

Overall means. The means (and standard deviations) from studies 1–3 are presented in Table 3. As can be seen, in study 1 the veterans rated all questions in subscale 1 higher than their spouses, finding the programme more distressing/intrusive (Wilks $\Lambda(6, 200) = .83$, $p < .0001$). For subscale 2 the means are more equivalent, although the females tended to endorse the programme more than the veterans (Wilks $\Lambda(2, 203) = .97$, $p < .04$). It is suggested that this may well be due to the population utilised and the veterans sensitisation to treatment programmes rather than a true gender difference or a content factor related to the course being run. However, this would suggest that the scale is sensitive to differences in samples with regard to the constructs being measured.

In study 2 it can be seen that the means for all questions were very similar for the CBT and EMDR conditions and these were not in fact significantly different ($p < .05$). There was also no significant difference when gender was treated as an independent variable for either treatment distress ($z(8, 15) = -1.03$, ns) or treatment endorsement ($z(8, 15) = -.42$).

Study 3 means and standard deviations are presented for the whole 10-item scale. There were no significant differences between male and female participants on either the distress (Wilks $\Lambda(7, 128) = .91$, ns) or endorsement (Wilks $\Lambda(3, 135) = .97$, ns) subscales. This lends weight to the hypothesis that the gender differences in study 1 were due to veterans' possible scepticism and sensitisation to treatment programmes and the Distress subscale being influenced by presentation severity when administered directly after the intervention. By the time this third study had begun, the Lifestyle Management Course had been running for at least 4 years and word had spread through the veteran community that the course was not too confrontational and was, in their opinion, a worthwhile intervention. It is our interpretation that such normalising of an intervention would also have an effect on distress appraisal. Alternatively, and possibly in interaction, the data in the third study were obtained at a 12-month follow-up and may reflect a changing pattern of distress appraisal over time.

3.5. Subsidiary analysis

A priori and post hoc perspectives. In both studies 1 and 2, the CEQ was also administered. Question 3 on the CEQ asks "How confident would you be in recommending this treatment to a friend who experiences similar problems?" and the questionnaire is administered after the intervention rationale has been explained yet before the actual treatment protocol begins. This question is very similar to question 8 on the DEVS which asks "How likely are you to recommend this program to

Table 3
Means (and standard deviations) for the DEVS in studies 1, 2 and 3

Item	Study 1 males[a]	Study 1 females[a]	Study 2 CBT[a]	Study 2 EMDR[a]	Study 3 males[b]	Study 3 females[b]
1. Distress experienced during first assessment session	5.21 (2.18)	3.78 (2.50)	5.50 (2.75)	5.91 (2.59)	4.64 (2.32)	3.65 (2.59)
2. Distress experienced during first treatment session	5.68 (2.21)	3.90 (2.53)	6.83 (2.55)	6.55 (2.77)	4.43 (2.22)	3.58 (2.41)
3. After first therapy session, distress experienced for the next few hours	5.04 (2.06)	3.20 (2.39)	5.42 (4.64)	4.64 (3.35)	4.09 (1.97)	3.61 (2.40)
4. How anxious returning to therapy for second treatment session	4.65 (2.22)	2.87 (2.34)	5.33 (2.64)	3.91 (2.88)	4.03 (2.22)	3.27 (2.21)
5. Overall, how intrusive was the therapy	4.68 (2.56)	3.50 (2.42)	4.50 (2.28)	4.73 (2.69)	4.00 (2.21)	3.31 (2.25)
6. Overall, how distressing was the whole treatment	4.72 (2.56)	3.74 (2.29)	4.92 (2.15)	5.46 (2.81)	4.05 (1.97)	3.83 (2.31)
7. Overall, how exhausting was the whole treatment	—	—	—	—	5.13 (2.36)	5.08 (2.36)
8. If knew as much about the treatment technique as now, how inclined in still participating	6.62 (2.54)	7.49 (2.26)	6.50 (1.83)	6.46 (2.88)	6.58 (2.21)	6.91 (2.30)
9. How likely in recommending this treatment to someone else…	8.51 (1.21)	8.64 (1.12)	7.67 (1.78)	6.36 (2.58)	8.01 (1.48)	7.77 (2.01)
10. Received value for time and money	—	—	—	—	7.69 (1.74)	7.57 (2.08)
Factor 1 total (treatment distress)	29.97 (10.20)	20.98 (11.95)	32.50 (10.66)	31.18 (14.25)	30.31 (11.99)	26.50 (14.11)
Factor 2 total (treatment endorsement)	15.11 (3.16)	16.12 (2.91)	14.17 (3.07)	12.82 (4.71)	22.41 (4.48)	22.25 (5.79)

[a]Comprised of only eight items.
[b]Comprised of ten items.

someone else with similar problems?''. However, these two questions did not correlate in study 1 overall ($r(202) = .008$, ns) or within just males ($r(107) = -.18$, ns), yet did show a small correlation within just females($r(95) = .22$, $p < .04$). In study 2 there was a correlation overall ($r(22) = .47$, $p < .03$) between the two questions. These results would suggest that this construct may be more a product of the treatment rationale and a coherent, logical, approach by the protocol than other aspects of the treatment experience.

4. Discussion

This research evaluated the psychometric properties of the DEVS for measuring distress and intrusiveness of a treatment protocol and participant endorsement of this protocol, for use in clinical outcome studies. This questionnaire is designed as a measure of the interventions utilised rather than a measure of therapist variables. In accomplishing this goal, three studies were presented that assessed the factor structure, internal consistency and the scale's relationship to outcome and discriminative ability between treatment protocols/populations. Studies 1 and 2 used an eight-item version of the scale, and based upon these results a 10-item version was developed which was used in study 3.

The results suggest that, when administered 12 months after intervention, this scale derives two subscales, namely treatment distress and endorsement of the protocols used. The scale demonstrates high internal consistency within the full seven-item Distress subscale (Cronbach's $\alpha = .92$ for the whole subscale) and inter-item correlations within this subscale of between .60 and .84. The Endorsement subscale originally comprised only two questions which derived a significant correlation, with a medium effect size of .37. When the full questionnaire was trialled, three items loaded on this factor, evidencing high internal consistency (Cronbach's $\alpha = .84$) and inter-item correlations of between .67 and .74. There were no cross-loadings of items on the two factors. The DEVS also displayed the ability to differentiate between two populations who underwent a Lifestyle Management course, with war-zone veterans scoring the course as more distressing and endorsing the programme less than their spouses, directly after the intervention. In explanation of this result it is suggested that veterans are more sensitised to intervention programmes and more likely to find any intervention as more harrowing than their spouses. Indeed, the six -item Distress subscale in study 1 correlated with symptomatology at intake when the DEVS was administered at the end of the course. However, in study 3, where the DEVS was administered 12 months following the end of the intervention, no relationships emerged—suggesting that the distress ratings are influenced by presentation when the DEVS is administered directly at the end of the treatment. This is further supported by the Distress subscale correlating with presentation severity in study 2, while endorsement did not correlate with presentation ($\alpha = .01$). This issue needs further clarification in future research where the relationship of neuroticism and/or negative affectivity to treatment distress needs to be further clarified. This would require measuring these indices and comparing

treatments that create such obvious different levels of distress that one could ascertain whether this derived subscale is sensitive to the treatment (as proposed here) or a by-product of neuroticism/negative affectivity.

However, although there was a difference in endorsement between the sexes in study 1, this effect was small with both populations scoring highly on endorsement. This difference was not evident, however, when ratings were obtained in study 3, 12 months after the intervention—even when only the original eight items were used. It is suggested that this result is likely to be reflective of a converging appraisal of the intervention over time between the veterans and their partners, and presentation severity becoming a more distal influence as time passes.

In comparing two treatment protocols for PTSD (an overtly CBT-based approach and EMDR), this scale did not detect any differences between the conditions for either subscale. However, at least three possibilities must be borne in mind when interpreting this result. Firstly, it could be claimed that there was a small total sample for this study. However, the effect size for any differences in means was negligible and so this seems an unlikely explanation. Secondly, and more probable, both methods utilised imaginal exposure to the traumatic experience and this could have been the component which most influenced participant ratings on both subscales. Thirdly, the ratings may have been influenced by another factor, such as a therapist variable, and hence there is the possibility of a Type III error. For example, we have yet to clarify the degree to which the therapeutic relationship affects ratings as opposed to the techniques utilised during the therapeutic encounter, as discussed in more detail below.

The relationship between treatment distress and endorsement to treatment outcome appears less obvious. From study 1, the results suggested that the veterans improved less in the short term on a measure of PTSD the more distressing they found the course (or vice versa), and the more they improved in the longer term the more likely they were to endorse the programme. However, as argued above, the short-term change and distress ratings are likely to have been influenced by their presentation severity. In study 3, when distress was measured 12 months following the end of the course, there were no significant ($\alpha = .01$) correlations with change scores over time from pre-intervention to 12-month follow-up. However, endorsement of the course at 12 months still correlated with change scores. It is suggested that people may endorse an intervention more when they get better—rather than basing it upon the "friendliness" of the intervention.

Related to this last point, the Endorsement subscale looks at evaluating the therapy (i.e. to see utility in it/to recommend it/to do it again) but is not an evaluation of the distress that they experienced during the therapy or their ability to see meaning in the actual distress. People may well experience high levels of distress during a therapy but also see meaning in it and are able to tolerate it due to this. Future research may wish to look at this issue and evaluate the merit of this possibility.

This questionnaire aimed to be an evaluation of the treatment protocol and not the therapist delivering the treatment. It is unclear whether the questionnaire has been successful in this and the co-administration of questionnaires looking at

therapist variables from the participants' perspective is still needed to clarify this question. It is also suggested that the utility of adding more questions evaluating participant endorsement to account for more variance of the construct would be a worthwhile endeavour. Also, due to the possibly changing appraisal of distress over time and the influence of presentation severity on this factor, a longitudinal study of factor structure for the full 10-item questionnaire may prove desirable. It is quite possible that directly following treatment, three factors may emerge with the recency of intervention meaning that clients are better able to differentiate between the three global questions assessing distress (questions 5–7) and the four specific questions relating to different aspects of the therapeutic encounter (questions 1–4), which all load on distress.

Furthermore, it must be stressed that this questionnaire obtains retrospective impressions of treatment (at the end of therapy), impressions which may not correlate with ratings of distress and endorsement during the actual treatment, and may also be affected by other aspects as mentioned above (e.g. protocol efficacy and therapist variables). Investigations into this would prove beneficial.

In light of the aforementioned possibilities, we would like to stress that, being the first structured assessment tool of its kind, this is an ongoing area of research and the questionnaire requires further psychometric validation with consequent refinement. For example, Devilly & Huther (in submission) recently administered the DEVS to novice raters who had read a case study of a rape victim who received either an intervention which utilised a predominantly exposure-based protocol or a cognitive-based protocol. In this study two types of outcome were stipulated for the patient in a counterbalanced design: positive or negative. It was found that raters, at least, differentiated between outcomes by their ratings on endorsement (i.e. they expected the patient to endorse either therapy more if she obtained a positive outcome from the intervention). Furthermore, the novice raters expected the exposure therapy protocol to be more distressing than she actually had rated it (the patient had, in reality, been treated with exposure therapy) and also expected the exposure protocol to be more distressing than the cognitive therapy intervention. This adds some weight to the discriminative abilities of the questionnaire—whether or not these ratings are accurate reflections of the distress and endorsement as rated by the actual patient.

It is our contention that, with claims of "patient-friendly" treatment protocols being advanced as a major reason for their use, the weight of responsibility for proving such claims rests squarely on the shoulders of the claimants. To our knowledge, this short and quickly administered questionnaire is the first of its kind within the research literature to be psychometrically evaluated and acts as a starting point for researchers generally and treatment protocol advocates specifically in backing such claims. However, of primary importance is the further development of this questionnaire which needs to be trialled further on a large research patient population other than combat veterans.

In summary, it is concluded that treatment toleration as measured in this questionnaire is comprised of at least treatment distress and treatment endorsement. It is further noted that influences on these factors may change over time with

intervention distress ratings being initially influenced by presentation severity. However, it is further concluded that participant endorsement may be more a function of improvement, a point not lost in the quest for intervention efficacy.

Acknowledgement

The author would like to thank the staff of the Queensland Vietnam Veterans Counselling Service and the participants from all studies for their assistance in this research. I would also like to thank Prof. Tom Borkovec for his comments regarding an earlier version of this article.

Appendix A. Distress/Endorsement Validation Scale

We would like you to indicate how upsetting you found the treatment which you have received. This is not a judgement of your therapist, but rather a judgement of the actual treatment and its methods.

1. How much distress did you experience during the first *assessment* session?

1	2	3	4	5	6	7	8	9
None at all			Somewhat distressed					Very distressed

2. How much distress did you experience during the first *treatment* session?

1	2	3	4	5	6	7	8	9
None at all			Somewhat distressed					Very distressed

3. On leaving the first treatment session, how much distress did you experience for the next few hours? (This is a rating of the distress caused by the "treatment" as opposed to your "normal" levels of distress).

1	2	3	4	5	6	7	8	9
None at all			Somewhat distressed					Very distressed

4. How anxious were you about returning to the second treatment session?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat anxious					Very anxious

5. Overall, how intrusive did you find the whole intervention?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat intrusive					Very intrusive

6. Overall, how distressing did you find the whole intervention?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat distressing					Very distressed

7. Overall, how exhausting did you find the whole intervention?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat exhausting					Very exhausting

8. If at the beginning of the treatment you knew as much about the intervention as you do now, how inclined would you have been in still participating?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat inclined					Very inclined

9. How likely are you to recommend this type of treatment to someone with similar problems?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat likely					Very likely

10. Do you believe that you received value for your time and money?

1	2	3	4	5	6	7	8	9
Not at all			Somewhat					Very much so

References

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

Borkovec, T. D., & 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.

Boudewyns, P. A., & Hyer, L. A. (1996). Eye movement desensitization and reprocessing (EMDR) as treatment for posttraumatic stress disorder (PTSD). *Clinical Psychology and Psychotherapy*, *3*, 185–195.

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Cristoph, P., et al. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, *51*, 3–16.

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.

Deane, F. P., & Chamberlain, K. (1994). Treatment fearfulness and distress as predictors of professional psychological help-seeking. *British Journal of Guidance and Counselling*, *22*, 207–217.

Derogatis, L.R. (1992). SCL-90: Administration, scoring and procedures manual-II. Baltimore. Clinical Psychometric Research.

Devilly, G. J. (2001). The roles of popularised distraction during exposure and researcher allegiance during outcome trials. *The Behavior Therapist*, *24*, 18–21.

Devilly, G. J. (2002). The psychological effects of a lifestyle management course on war veterans and their spouses. *Journal of Clinical Psychology*, *58*, 1119–1134.

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*, 73–86.

Devilly, G. J., & Gournay, K. (1995). Community Psychiatric Nursing With Non-Psychotic Patients: relating Process To Outcome. *The Australian and New Zealand Journal of Mental Health Nursing*, *4*, 53–60.

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.

Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy*, *9*, 171–180.

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.

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, *9*, 139–145.

Harris, M., & Bergman, H. C. (1987). Differential treatment planning for young adult chronic patients. *Hospital & Community Psychiatry*, *38*, 638–643.

Hawley, C., Sivakumaran, T., Huber, T. J., & Ige, A. K. (1998). Combination therapy with nefazodone and lithium: safety and tolerability in fourteen patients. *International Journal of Psychiatry in Clinical Practice*, *2*, 251–254.

Horowitz, M., Wilner, M., & Alvarez, W. (1979). Impact of event scale: a measure of subjective stress. *Psychosomatic Medicine*, *41*, 209–218.

Hyer, L., & Brandsma, J. (1997). EMDR minus eye movements equals good psychotherapy. *Journal of Traumatic Stress*, *10*, 515–522.

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.

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.

Martin, A. (1998). Psychopharmacology update: what's new in the treatment of depression. *Psychoanalysis and Psychotherapy*, *15*, 131–134.

Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: between Scylla and Charybdis. *Psychological Bulletin*, *126*, 964–981.

Novaco, R. W. (1975). *Anger control: the development and evaluation of an experimental treatment*. Lexington, MA: D.C. Heath.

O'Malley, S. S., Suh, C. S., & Strupp, H. H. (1983). The Vanderbilt Psychotherapy Process Scale: a report on the scale development and a process-outcome study. *Journal of Consulting and Clinical Psychology*, *51*, 581–586.

Pitman, R. K., Orr, S. P., Altman, B., Longpre, R. E., Poiré, R. E., & Macklin, M. L. (1996). Emotional processing during Eye Movement Desensitization and Reprocessing therapy of Vietnam veterans with chronic post-traumatic stress disorder. *Comprehensive Psychiatry*, *37*, 419–429.

Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification*, *27*, 300–312.

Scott, M. J., & Stradling, S. G. (1997). Client compliance with exposure treatments for posttraumatic stress disorder. *Journal of Traumatic Stress*, *10*, 523–526.

Shapiro, F. (1989). Eye movement desensitization: a new treatment for post-traumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *20*, 211–217.

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.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short form health survey (SF-36): conceptual framework and item selection. *Medical Care*, *30*, 473–483.

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.

Yates, B. T. (1994). Toward the incorporation of costs, cost-effectiveness analysis and cost-benefit analysis into clinical research. *Journal of Consulting and Clinical Psychology*, *62*, 729–736.

Zeltzer, L. K., & LeBaron, S. (1985). Does ethnicity constitute a risk factor in the psychological distress of adolescents with cancer? *Journal of Adolescent Health Care*, *6*, 8–11.