British Journal of Health Psychology (2001), 6, 179–196 *Printed in Great Britain*

The factor structure of the BDI in facial pain and other chronic pain patients: A comparison of two models using confirmatory factor analysis

Anne Miles*
Department of Epidemiology and Public Health, University College London, UK

Chris McManus
Department of Psychology, University College London, UK

Charlotte Feinmann
Joint Department of Oral and Maxillofacial Surgery, Eastman Dental Institute, London, UK

Lesley Glover
Sub-department of Clinical and Health Psychology, University College London, UK

Sheelah Harrison
Joint Department of Oral and Maxillofacial Surgery, Eastman Dental Institute, London, UK

Shirley Pearce
School of Health, University of East Anglia, UK

Objectives. 1) To compare two measurement models of the BDI in chronic pain sufferers to see which provides the better fit; 2) to assess whether model fit differs for a facial pain sample compared to a sample of pain sufferers attending a multidisciplinary pain clinic; and 3) to establish which affective and somatic sub-scales of the BDI could be used in chronic pain research.

Design. Two groups of chronic pain sufferers, a facial pain group, and a group attending a multidisciplinary pain clinic completed self-report questionnaires on pain (Multidimensional Pain Inventory), depression (BDI), and measures of anxiety and depression-related pain cognitions (the Spielberger State-Trait Anxiety Inventory and the Pain Cognitions Questionnaire). The measurement models of the BDI were tested using LISREL structural equation modelling and their construct validity examined using partial correlation analysis.

Method. A total of 173 people attending a multidisciplinary pain clinic and 157

*Requests for reprints should be addressed to Anne Miles, Department of Epidemiology and Public Health, Health Behaviour Unit, University College London, Gower Street Campus, 2–16 Torrington Place, London WC1E 6BT, UK (e-mail: annem@public-health.ucl.ac.uk).

patients attending a facial pain clinic completed self-report measures of pain and mood prior to their respective clinical consultations.

Results. The model offered by Novy *et al.* (containing one affective factor 'Negative-attitude suicide' and two somatic factors 'Performance difficulty' and 'Physiological manifestations') fitted both pain groups better than the model offered by Williams and Richardson (containing one affective factor 'Self-reproach', one somatic factor 'Somatic disturbance' and one factor with a mixture of both affective and somatic items 'Sadness about health'). However, when the factors were allowed to correlate in the latter model, both models were broadly equivalent.

Conclusions. The two measurement models adequately fitted data in both pain samples when the factors were allowed to intercorrelate in the Williams and Richardson model. Both the affective scales offered by both models could be used in future research, although the somatic factor offered by the Williams and Richardson model offered much higher levels of internal reliability than either of those offered in the Novy *et al.* model. The findings are discussed in relation to the issue of depression in chronic pain.

The Beck Depression Inventory (BDI) is widely used as both a clinical measure and a research tool (e.g. Beck, Steer, & Garbin, 1988; Romano & Turner, 1985). However, it was standardized in psychiatric patient samples, and its validity has not been established in populations with physical illness. The extent to which it measures depression in these groups has therefore been questioned as the somatic items it contains may be endorsed because of current physical illness rather than depressed mood (e.g. Kathol *et al.*, 1990; Kathol & Petty, 1981; Peterson *et al.*, 1991). This debate has extended to the usefulness of the BDI in assessing depression in chronic pain groups (Novy, Nelson, Berry, & Averill, 1995; Williams & Richardson, 1993). While the nature of the relationship between chronic pain and depression is complex (Romano & Turner, 1985; Roy, Thomas, & Matas, 1984), significant overlap between problems associated with pain and symptoms of depression (e.g. sleep disturbance and work inhibition) can exist. This overlap may confound BDI scores (Love, 1987) leading to inflated scores and the possibility of misdiagnosis. Such overlap is of theoretical as well as clinical importance, as the BDI is frequently used to test theories, notably those exploring the relationship between pain and depression (e.g. Rudy, Kerns, & Turk, 1988). Hence, determining which items confound the measurement of depression will benefit the future testing of theoretical models.

The factor structure of the BDI

In a review of the psychometric properties of the BDI, Beck *et al.* (1988) concluded that it measures one underlying general syndrome of depression that can be decomposed into three highly correlated factors: 'Negative attitudes towards the self', 'Performance impairment', and 'Somatic disturbance'. They argued that these three factors are relatively stable; having emerged across different diagnostic groups (major depressives, heroin addicts, schizophrenics, and college students) and modes of questionnaire administration (self vs. clinician). However, they also state that the explicit composition of these factors may differ across different diagnostic groups and that: 'An understanding

of the factorial composition of the BDI is crucial when the instrument is used to screen for depression in medical patients' (p. 92).

In line with this recommendation, previous research has explored the factor structure of the BDI and its construct validity in chronic pain groups (Novy *et al.*, 1995; Wesley, Gatchel, Garofalo, & Polatin, 1999; Williams & Richardson, 1993).

Williams and Richardson (1993) used an exploratory factor analysis with orthogonal rotation on data from people who met criteria for inclusion in a multidisciplinary cognitive and behavioural pain management programme and reported a three-factor structure. One factor (labelled the 'Somatic disturbance' factor) correlated positively with physical performance measures (walking distance and climbing stairs) but was unrelated to measures of anxiety and depression-related pain cognitions (the trait scale of the STAI and hopelessness and helplessness scales of the PCQ). Hence, they demonstrated that pain patients' endorsement of some somatic items seemed to be a direct result of the disabling impact of pain and not of mood.

Novy *et al.* (1995) used a confirmatory factor analysis, testing a model based on that of Tanaka and Huba (1984). The measurement model of Novy *et al.* originated from exploratory factor analysis using oblique rotation conducted on Beck's original data on consecutive admissions to both in-patient and out-patient clinics at two hospitals (Beck, 1970). Novy *et al.* (1995) demonstrated that this model adequately fitted data from a multidisciplinary pain clinic sample and that this three-factor structure still related to a common core of depressive symptomatology. Hence, Novy *et al.* (1995) partly challenged the view put forward by Williams and Richardson (1993) that a subscale of the BDI may not relate to depression. However, they offered similar recommendations regarding the scoring of the BDI to Williams and Richardson (1993): that somatic items be scored separately when the BDI is being used to measure depression in chronic pain groups and that somatic item endorsement be viewed within the broader context of the selection of cognitive and affective items of the BDI.

This work on the factor structure of the BDI in chronic pain sufferers currently poses two problems. First, two different measurement models (describing the relationship between specific items and latent factors) have been forwarded, but no direct test has yet been made of which might be a better fit. There is, therefore, no consistent message about which somatic items should be excluded when measuring depression in chronic pain and which items might form a scale measuring depression, unconfounded by the presence of pain. Second, this research has tended to focus on those attending multidisciplinary pain clinics, and such groups have not included a significant facial pain sample. While this probably reflects the fact that facial pain sufferers typically attend different clinics (Turp, Kowalski, & Stohler, 1998) facial pain sufferers differ significantly from other pain groups in that they are frequently younger, more likely to be female, and, more importantly, report lower levels of pain-related interference (the amount to which pain affects social, recreational and work activities) (Rome, Harness, & Kaplan, 1990). Facial pain sufferers are, for example, more likely to be able to continue working, despite the pain, than those suffering from pain located elsewhere (Holzberg, Robinson, Geisser, & Gremillion, 1996), and in cases where occupational disability in facial pain groups has been substantial (e.g. 50%), an inability to work has been associated with pain located outside the face and not with the presence of facial pain *per se* (Haegerstam & Allerbring, 1995). Hence, the finding of a scale related to physical

performance, and not mood, may not extend to a pain sample with relatively low levels of pain-related disability. As a result, the degree to which either of the models proposed extends to facial pain groups, and what the construct validity of the different subscales might be in this patient group, remain to be explored.

The aim of the present study was therefore to: a) test which of the two measurement models offered in previous research best fit data from both a multidisciplinary pain group (similar to those groups previously studied) and a facial pain group (different from those usually studied as they typically report relatively low levels of pain-related interference), and b) examine the construct validity of the subscales of both models in the two pain groups to see which model offers the most useful affective and somatic sub-scales.

Method

Both pain groups had chronic pain, i.e. lasting longer than 3 months, which persisted despite medical and surgical treatment. Inclusion criteria for completion of the questionnaires in both pain groups were: aged 18 or over and fluent in written and spoken English.

Facial pain sample

A total of 297 consecutive patients attending the Eastman Dental Hospital, and fulfilling the inclusion criteria noted above, completed the BDI. Of these, 140 were not included in the current analysis: 7 had missing data, 25 had no current pain problem, and 108 had chronic pain complaints in areas as well as the face, head or neck, suggesting they may have overlapped with the pain clinic group. The remaining 157 had a primary diagnosis of chronic idiopathic facial pain (i.e. atypical facial pain and/or facial arthromyalgia) with pain located only in the face (cheeks/jaw and/or mouth).

Pain clinic sample

A total of 231 consecutive patients attending pain clinics at the Whittington and Middlesex Hospitals, and fulfilling the inclusion criteria noted above, completed the BDI. Of these, 58 were not included in the current analysis: 19 had missing data, 4 had a chronic pain site in their face/head, and 35 were unable to identify primary pain sites but stated they had all-over-body pain, or pain in all joints; hence, whether or not pain was also present in the face could not be established. The breakdown according to the main pain site in the remaining 173 patients forming the pain clinic sample was as follows: 44% had back pain, a further 24% had pain located in the neck and/or upper shoulders, 13% had pain principally in the legs, and the remaining 19% were spread roughly evenly across the location sites of chest, abdomen, pelvis, and anal/genital regions.

Measures

Patients completed the Multi-dimensional Pain Inventory (MPI) (Kerns, Turk, & Rudy, 1985), the Beck Depression Inventory (BDI) (Beck, Rush, Shaw, & Emery, 1979), the Spielberger State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, & Lushene, 1970), and the Pain Cognitions Questionnaire (PCQ) (Boston, Pearce, & Richardson, 1990) as part of a battery of questionnaires prior to clinical assessment. The facial pain group were attending an appointment to assess their suitability for inclusion in a placebo-controlled trial comparing anti-depressive medication and cognitive–behavioural therapy in the treatment of chronic facial pain. The pain clinic sample were also attending an appointment to assess their suitability for inclusion in a treatment outcome study, a comparison of group vs. individual cognitive–behavioural pain

management. Both groups therefore completed these measures prior to being assessed for suitability for these programmes. The samples therefore include both those who went on to take part in the respective studies and those who did not.

The MPI was chosen because it is a widely used measure of multidimensional aspects of pain (De Gagne, Mikail, & D'Eon, 1995). It is a 60-item self-report measure with nine scales that measure aspects of pain experience such as pain severity and the amount the pain interferes with day-to-day life, as well as general measures such as perceived life control, distress, and general activity level. Five of the scales from the MPI will be reported here. Each scale ranges from 0–6, with higher scores indicating higher levels of pain severity, higher levels of pain interference, higher levels of perceived life control, higher levels of distress, and higher levels of general activity, respectively.

The BDI is a 21-item self-report measure with four response options for each item scoring 0–3. The responses to each item are summed giving an overall score, which ranges from 0 to 63.

The Spielberger State/Trait Anxiety Inventory (STAI) was chosen because it is also widely used in chronic pain research and, along with the Pain Cognitions Questionnaire, was used to assess the construct validity of the BDI in the Williams and Richardson (1993) study. The STAI measures negative affect. Both the state and trait measures contain 20 questions that are scored 1–4, with higher scores denoting higher levels of anxiety. The scores on each item are summed, and the total scores for both scales range from 20 to 80.

The PCQ was originally described as having four scales, two of which relate to depression-related pain cognitions: the hopelessness and helplessness subscales. While a more recent analysis has argued that the scale contains two, rather than four, scales (Richardson & Richardson, 1999), the original scales will be used in the present study as these were used in the Williams and Richardson (1993) paper; hence, the results of the present study can be more usefully compared with previous research.

Statistical analyses

Confirmatory factor analysis (*CFA*). This uses an *a priori* theoretical model and determines whether a given data set is consistent with it, i.e. whether the model adequately fits the data. This model can be based on prior theoretical and/or empirical work. This is in contrast to an exploratory factor analysis, which uses *data* to derive several possible models and where no model is specified prior to the analysis.

The two measurement models offered by Novy *et al.* and Williams and Richardson are shown in Figs 1 and 2, respectively.

Both of the measurement models were made up of three primary constructs. Consistent with the testing of measurement models of the BDI in previous work (e.g. Tanaka & Huba, 1984), the initial models were kept relatively simple. Hence, the initial specifications for both models did not permit any correlated uniquenesses (correlated errors indicating terms that have variance in common not attributable to a primary construct) or item complexity, whereby one item loads on more than one factor.

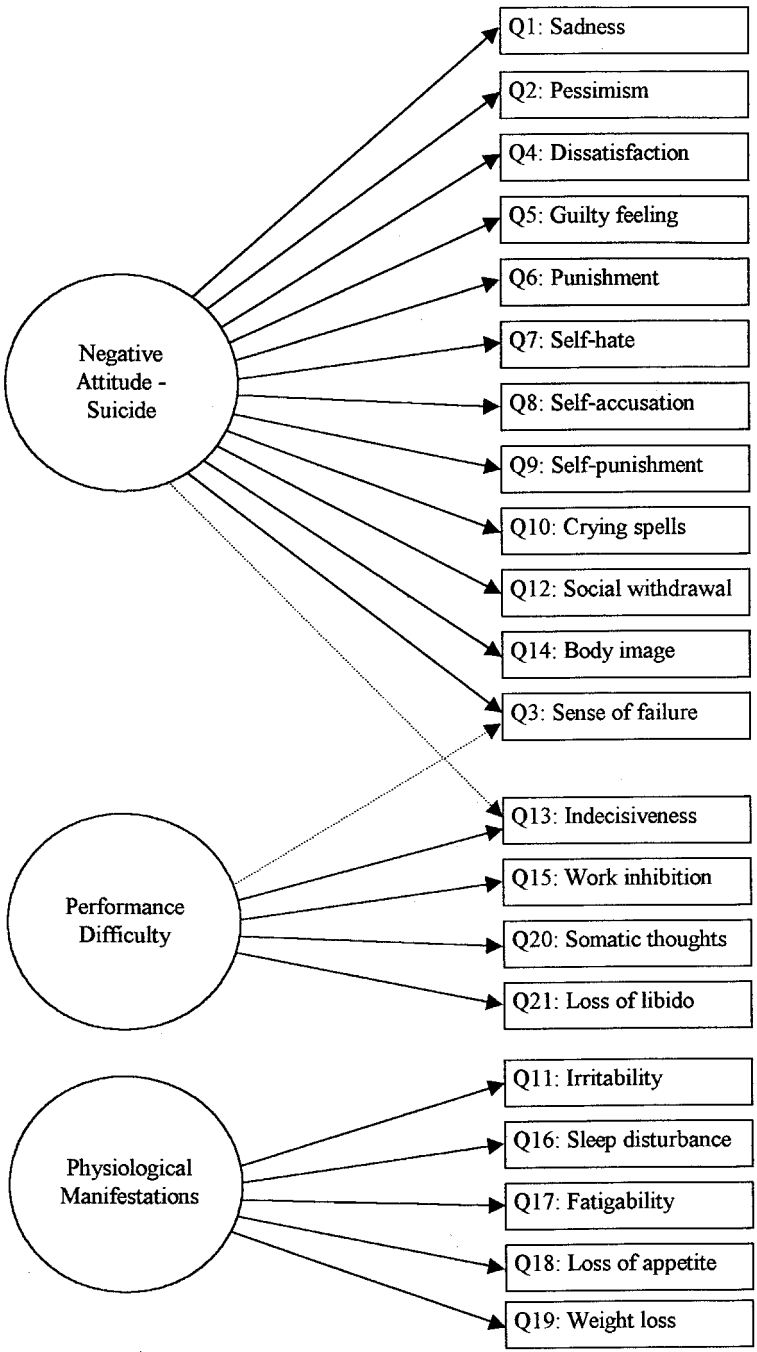

Figure 1. Novy *et al.* measurement model. This diagram shows the factor loadings but not the factor covariances of the Novy *et al.* measurement model.

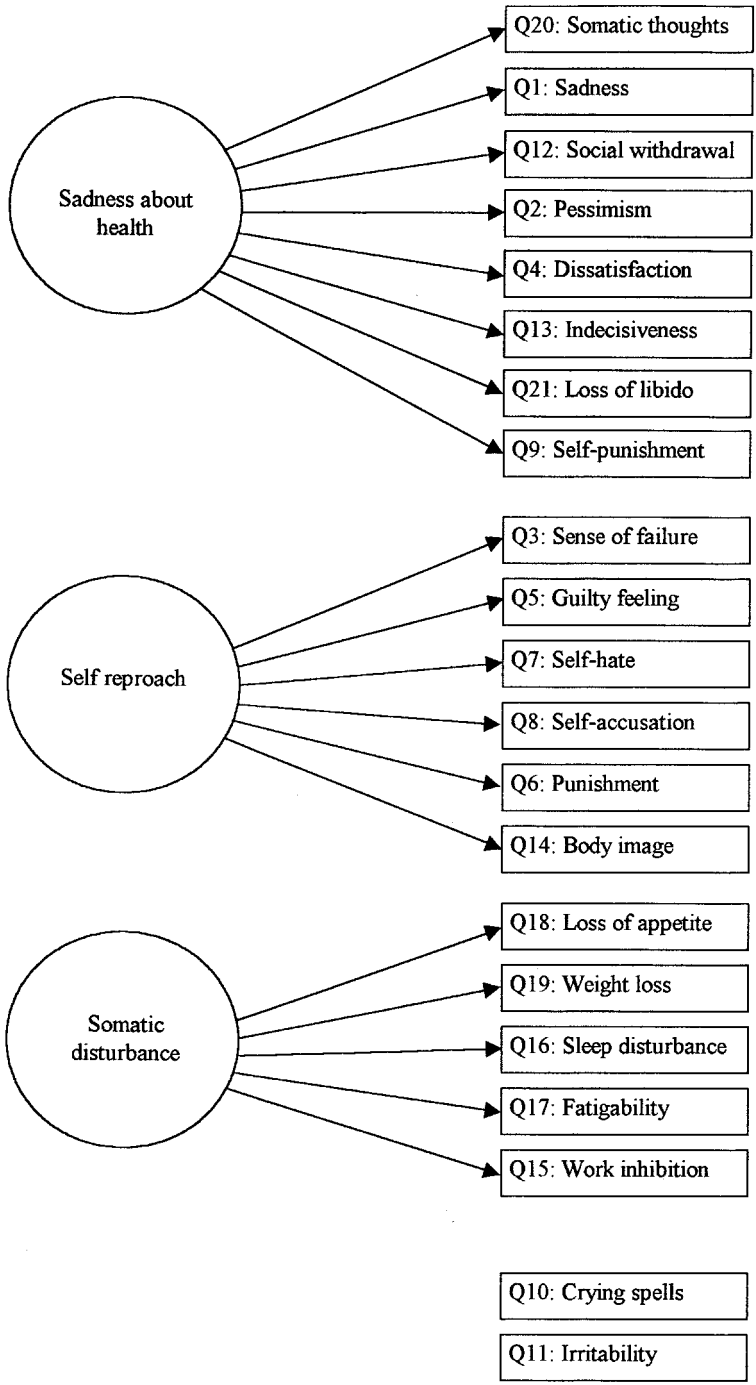

Figure 2. Williams and Richardson measurement model.

The estimation method used to test both of these models in both pain groups was maximum likelihood. This is the most frequently used estimation method and is recommended for small sample sizes. It is also robust against moderate departures from normality (Joreskog & Sorbom, 1989; p. 24). The analysis begins by estimating parameters, and then the maximum likelihood function attempts to find a value for the fitting function that cannot be made smaller by making further minor adjustments to the parameter estimates.

Missing data were dealt with using listwise deletion: 26 cases were excluded. Although this method of dealing with missing data has the disadvantage of not maximizing the data available, it was chosen so that there would not be any different parts of the correlation matrix that were generated by different data, which would have resulted in the possibility that the data matrix was not internally consistent.

Assessing the adequacy of the measurement models. There are a number of different fit indices that can be used in CFA, although no widespread agreement currently exists about which is the best (Maruyama, 1998). In a recent paper testing the performance of a variety of fit indices, Hu and Bentler (1999) suggested using a two-index presentation strategy for evaluating model fit. They found such a strategy minimized Type I and Type II error rates because different fit indices are sensitive to different aspects of model misspecification. They argued that one of the indices should be the standardized root mean square residual (SRMR) and that this should be supplemented by an index such as the root mean square error of approximation (RMSEA). Furthermore, they presented evidence that conventional cut-off criteria of, for example, 0.90 for GFI and 0.05 for RMSEA, result in the over-rejection of adequately specified models and hence were too strict.

The fit indices and cut-off criteria of Hu and Bentler were therefore used in the present study, but because Novy *et al.* used the Chi-square statistic, the goodness-of-fit index (GFI) and the adjusted goodness-of-fit index (AGFI), the two measurement models were compared on the five following indices: Chi-square, the goodness-of-fit index, the adjusted goodness-of-fit index, the standardized root mean square residual and the root mean square error of approximation. This allowed the criteria of Hu and Bentler to be used and comparisons to be made with the results of Novy *et al.*

The Chi-square test is typically viewed as a goodness-of-fit statistic and, alongside other indices, is used to decide whether or not the model adequately fits the data. As a result, it is not viewed as testing a specific hypothesis, so the model may be accepted even though there is a significant Chi-square (indicating that there is a significant gap between the predicted model and the data).

The following goodness-of-fit indices are all called absolute (as opposed to incremental) fit indices as they give an indication of how well the *a priori* model reproduces the data sample (rather than examining the improvement in model fit by comparing one model with a baseline model).

The GFI index indicates the relative amount of variance and covariance explained by the model as a whole. Values vary between 0 and 1, with larger numbers indicating a better-fitting model.

The AGFI is similar to the GFI but takes into account degrees of freedom, and hence, simpler models yield better fit indices.

The SRMR illustrates how much error there is between the estimated parameters

and the actual parameters taken from the data. Scores vary between 0 and 1, with smaller values indicating a better-fitting model.

The RMSEA is also a measure of discrepancy of fit but takes into account number of degrees of freedom but, as noted above, is sensitive to different aspects of model misspecification compared to the SRMR.

There are currently no fixed guidelines as to acceptable cut-off points for the above indices. While cut-offs of 0.95 for GFI and 0.90 AGFI are often cited, some set more stringent (e.g. Hayduk suggests 0.95 for AGFI) (Hayduk, 1996) and some less stringent cut-offs. For example, Novy *et al.* set themselves the revised criteria of 0.90 for GFI and 0.80 for AGFI, and Tanaka and Huba (1984) described a GFI of 0.78 as 'quite acceptable' (p. 627). The cut-off values for SRMR and RMSEA are typically values less than or equal to 0.05 (Maruyama, 1998), although Hu and Bentler suggest optimal cut-offs of close to 0.08 for SRMR and close to 0.06 for RMSEA, with values of 0.11 and 0.08, respectively, being acceptable. For the present purpose, the RMSEA is more valuable than the SRMR. This is because the amount of error in a model typically decreases as the number of parameters in the model increases, and the two models being compared have a different number of parameters. The Novy *et al.* model contains more parameters than the Williams and Richardson model, as it uses all 21 questions of the BDI, whereas the Williams and Richardson model uses 19. Hence the RMSEA offers a measure that is more comparable across models than the root mean square residual.

However, it should be noted that the main aim of CFA is to see whether the error in predicting the sample correlation matrix from the specified model is small enough to make the model theoretically worthwhile. Hence, any goodness-of-fit indices need to be considered within this broader theoretical context.

Results

Participant characteristics

Differences between the two pain groups on demographic and psychological measures are shown in Table 1.

The two pain groups differed on measures of mood, pain interference and demographic variables. Consistent with previous research into facial pain, this group were younger, a larger proportion were female, and they scored lower on measures of pain-related disability (the pain interference subscale of the MPI).

Confirmatory factor analysis

Two models were tested initially; the Novy *et al.* model and the Williams and Richardson model.

Novy *et al.* model: The three factors were 'Negative attitude—suicide', 'Performance difficulty', and 'Physiological manifestations', and the measurement model is shown in Fig. 1. The factors were allowed to be oblique (i.e. each factor was allowed to correlate with each of the others).

Table 1. Comparisons across the two pain groups on demographic and psychological measures

	Facial pain sample ($N = 157$)	Pain clinic sample ($N = 173$)	
Age (years)	37.3 (14.1)	48.1 (12.3)	$t = -7.83$*
Sex (percentage female)	80	64	χ^2 (1) = 11.21*
Pain chronicity (years)	4.5 (7.4)	8.2 (7.6)	$t = -4.42$*
Pain severity (MPIsev)	2.87 (1.40)	4.22 (1.12)	$t = -9.82$*
Interference in daily activities caused by pain (MPInt)	1.77 (1.55)	4.29 (1.16)	$t = -16.99$*
Perceived life control (MPIcont)	3.45 (1.25)	3.07 (1.37)	$t = 2.72$
Distress (MPIdist)	2.93 (1.35)	3.70 (1.29)	$t = -5.48$*
General activity level (MPIgen)	3.24 (0.98)	2.67 (1.04)	$t = 5.30$*
BDI	8.41 (7.84)	16.46 (9.22)	$t = -8.65$*
Spielberger State Anxiety	40.08 (11.07)	44.49 (12.17)	$t = -3.42$*
Spielberger Trait Anxiety	41.59 (12.24)	47.72 (11.16)	$t = -4.68$*
Pain-related hopeless cognitions (PCQhope)	1.64 (0.59)	2.18 (0.67)	$t = -7.74$*
Pain-related helpless cognitions (PCQhelp)	1.37 (0.42)	1.68 (0.59)	$t = -5.43$*

*Significant at the 0.05 level following Bonferroni corrections.

Williams and Richardson model: The three factors were 'Sadness about health', 'Self-reproach', and 'Somatic disturbance', and the measurement model is shown in Fig. 2. In line with the exploratory factor analysis on which this model was based, the three factors were orthogonal (i.e. were not allowed to correlate with each other).

For both models, a 21 ×21 correlation matrix of the BDI scores for each of the pain groups were used. The fit indices for these models are shown in Table 2.

Novy et al. model. For the facial pain group, the goodness-of-fit was reasonable (e.g. GFI = 0.83; RMSEA = 0.07), although there was a highly significant overall Chi-square, indicating a less than complete fit between the model and the data. The fit for the pain clinic pain group was slightly better (e.g. GFI = 0.85; RMSEA= 0.08), although, again, there was a highly significant overall Chi-square. Table 3 shows the loadings of each of the items on the three factors. This demonstrates that in all cases but one, the loadings were significant and in the expected directions.

The only item that did not fit with $p < .001$ was item q19 (Loss of weight) in the facial pain group, where the loading was only 0.02 ($t = 0.22$, n.s.). Apart from that one exception, it is noteworthy that the loadings were numerically very similar in the two groups in most cases.

As might be expected, the three factors were substantially intercorrelated in both the facial pain and pain clinic groups (see Table 4). All are significant with $p < .001$.

This model meets fit criteria forwarded by Hu and Bentler (1999) but not more conventional cut-off criteria of 0.05 for RMSEA, for example. Novy *et al.* (1995) also found that this model did not meet more conventional fit criteria and they went on to use the automatic modification features present in LISREL, which gives information on how the model can be modified in order to improve its fit. On the basis of this

Table 2. Fit indices for the three measurement models across both pain groups

	Facial pain sample	Multidisciplinary pain clinic sample
Novy *et al.* model		
Chi-square (d.f.)	351.99 (186)	393.26 (186)
Chi-square/d.f.	1.89	2.11
Root mean square residual standardized	0.07	0.07
Root mean square error of approximation (RMSEA)	0.07	0.08
GFI	0.83	0.85
AGFI	0.79	0.81
Williams and Richardson model (with uncorrelated factors)		
Chi-square (d.f.)	613.96 (191)	552.35 (191)
Chi-square/d.f.	3.21	2.89
Root mean square residual standardized	0.25	0.20
Root mean square error of approximation (RMSEA)	0.12	0.10
GFI	0.72	0.78
AGFI	0.66	0.74
Williams and Richardson model (with correlated factors)		
Chi-square (d.f.)	397.14 (188)	405.89 (188)
Chi-square/d.f.	2.11	2.15
Root mean square residual standardized	0.12	0.10
Root mean square error of approximation (RMSEA)	0.08	0.08
GFI	0.80	0.85
AGFI	0.75	0.81

information, Novy *et al.* allowed two of the items (q3 and q13) to load on both Factors 1 and 2. However, the largest modification indices calculated for the present study did not include those freed by Novy *et al.* Fitting a revised model in which q3 and q13 were allowed to load on Factors 1 and 2 resulted in a significant improvement in the Chi-square goodness-of-fit statistic for the multidisciplinary pain group ($\Delta\chi^2 = 8.51$, Δd.f. $= 2$, $p < .025$), (the loadings on Factor 1 for items 3 and 13 were 0.87 and 0.54, respectively, and the loadings on Factor 2 for these items were -0.26 and -0.02, respectively). However, this small improvement in fit should be contrasted with the change in Chi-square of 111.84 reported by Novy *et al.* (1995). Furthermore, the model did not fit the facial pain sample, resulting in impossible parameters for the correlation between factors of over 1. It seems, therefore, that the improvement in fit found by Novy *et al.* in their modified measurement model may not be a replicable measurement structure.

Williams and Richardson model. For the facial pain group, the model did not fit well (GFI $= 0.72$; RMSEA $= 0.12$), and there was a highly significant overall Chi-square of 613.96 with 191 d.f. The fit for the pain clinic group was marginally better (GFI $= 0.78$, RMSEA $= 0.10$), and again, there was a highly significant overall Chi-square of 552.35

Table 3. Item loadings for the Novy *et al.* model on the three factors for the separate pain groups

Factor	Facial pain sample		Pain clinic sample	
	Loading	*t* (sig)	Loading	*t* (sig)
Negative attitude—suicide				
Q1: Sadness	.66	8.94***	.67	8.95***
Q2: Pessimism	.71	9.93***	.61	7.97***
Q3: Sense of failure	.66	8.93***	.70	9.53***
Q4: Lack of satisfaction	.67	9.11***	.65	8.60***
Q5: Guilty feeling	.57	7.49***	.67	9.00***
Q6: Sense of punishment	.56	7.37***	.55	6.98***
Q7: Self-hate	.71	9.84***	.64	8.57***
Q8: Self-accusation	.70	9.62***	.67	9.09***
Q9: Self-punishment	.64	8.58***	.49	6.19***
Q10: Crying spells	.32	3.94***	.46	5.71***
Q12: Social withdrawal	.63	8.41***	.55	7.11***
Q14: Body image	.41	5.11***	.44	5.40***
Performance difficulty				
Q13: Indecisiveness	.70	9.60***	.57	6.82***
Q15: Work inhibition	.73	10.09***	.52	6.12***
Q20: Somatic preoccupation	.48	6.03***	.57	6.83***
Q21: Loss of libido	.63	8.29***	.48	5.70***
Physiological manifestations				
Q11: Irritability	.44	5.16***	.34	3.71***
Q16: Sleep disturbance	.57	6.89***	.37	4.01***
Q17: Fatigability	.74	9.26***	.57	6.35***
Q18: Loss of appetite	.37	4.30***	.48	5.30***
Q19: Involuntary weight loss	.02	0.22	.32	3.45***

***$p < 0.001$.

Table 4. Correlation matrix for the factors of both measurement models for both pain groups

Pain clinic sample above diagonal/facial pain below diagonal	I	II	III
Novy *et al.* model			
I	–	.81	.70
II	.91	–	.92
III	.70	.92	–
Williams and Richardson model			
I	–	0.77	0.74
II	0.87	–	0.59
III	0.90	0.73	–

with 191 d.f., indicating a less than complete fit between the model and the data. More importantly, these indices did not meet Hu and Bentler's criteria. The modifications indices indicated overwhelmingly that the model would be improved by allowing the factors to correlate. Doing this significantly improved the fit of the model in both pain

groups (facial pain groups: $\Delta\chi^2 = 206.82$; Δd.f. $= 3$; $p < 0.001$; multidisciplinary pain group $\Delta\chi^2 = 146.46$; Δd.f. $= 3$; $p < 0.001$). The fit indices, shown in Table 2, indicate that this model fitted the multidisciplinary pain group better than the facial pain group, according to some indices (e.g. GFI $= 0.85$ vs GFI $= 0.80$) but not others, such as the index of remaining error adjusted for degrees of freedom, where the models emerged as equivalent (RMSEA $= 0.08$ vs. 0.08). In this revised form, the fit indices for the Williams and Richardson model and the Novy *et al.* model were similar for the multidisciplinary pain clinic sample, although the Novy model emerged as a marginally better fit for the facial pain sample. Table 5 shows the loadings of each of the items on the three factors. Again, the only item that did not fit with $p < .001$ was item Q19 (Loss of weight) in the facial pain group, where the loading was only 0.02 ($t = 0.22$, n.s.).

Table 5. Item loadings for the Williams and Richardson model on the three factors for the separate pain groups

	Facial pain sample		Pain clinic sample	
Factor	Loading	*t* (sig)	Loading	*t* (sig)
Sadness about health				
Q20: Somatic preoccupation	0.44	5.57***	0.59	8.01***
Q1: Sadness	0.65	8.72***	0.66	9.30***
Q12: Social withdrawal	0.67	9.07***	0.56	7.59***
Q2: Pessimism	0.70	9.64***	0.70	9.91***
Q4: Lack of satisfaction	0.68	9.37***	0.65	9.00***
Q13: Indecisiveness	0.72	9.96***	0.52	6.96***
Q21: Loss of libido	0.61	8.01***	0.48	6.35***
Q9: Self-punishment	0.62	8.21***	0.51	6.68***
Self-reproach				
Q3: Sense of failure	0.70	9.41***	0.75	10.98***
Q5: Guilty feeling	0.61	7.85***	0.71	10.20***
Q7: Self-hate	0.76	10.59***	0.65	9.04***
Q8: Self-accusation	0.72	9.79***	0.69	9.75***
Q6: Sense of punishment	0.58	7.37***	0.58	7.80***
Q14: Body image	0.45	5.51***	0.43	5.57***
Somatic disturbance				
Q18: Loss of appetite	0.37	4.34***	0.46	5.37***
Q19: Involuntary weight loss	0.03	0.36	0.29	3.29***
Q16: Sleep disturbance	0.48	5.84***	0.38	4.35***
Q17: Fatigability	0.67	8.74***	0.57	6.71***
Q15: Work inhibition	0.76	10.17***	0.58	6.81***
Q10: Crying spells				
Q11: Irritability				

***$p < 0.001$.

Construct validity of the different factors across the two main groups

In order to explore the construct validity of the factors of the Novy *et al.* model in the two pain groups, simple scales were computed for each factor based on the relevant items (see

Fig. 1), although item 19 was omitted from the factor 'Physiological manifestations'. Scales were created by simple averaging of the score on each of the items included. Alpha coefficients for reliability for 'Negative attitude—suicide' were 0.86 (facial pain) and 0.85 (pain clinic); for 'Performance difficulty' 0.70 (facial pain) and 0.60 (pain clinic); and for 'Physiological manifestations' .53 (facial pain) and .46 (pain clinic). These alpha scores are broadly similar to those reported by Novy *et al.* of 0.89, 0.71 and 0.58, respectively. The internal reliability of 'Negative attitude—suicide' was good for both pain groups, but the internal reliability of 'Performance difficulty', while good for the facial pain group, was quite poor for the pain clinic group. The internal reliability of the 'Physiological manifestations' was poor for both pain groups. However, it is worth noting that the alpha for the BDI scale as a whole was high, 0.89 (facial pain) and 0.87 (pain clinic).

Simple scales were also computed based on the factors of the Williams and Richardson model (see Fig. 2), again omitting q19 from the 'Somatic disturbance' factor. Alpha coefficients for reliability for the three scales in the Williams and Richardson measurement model were: 'Sadness about health', 0.82 (facial pain) 0.80 (pain clinic); 'Self-reproach' 0.79 (facial pain) 0.79 (pain clinic); and 'Somatic disturbance' 0.60 (facial pain) 0.55 (pain clinic).

Partial correlation coefficients were computed using three measures broadly relating to negative affect (anxiety, and hopeless and helpless pain-related cognitions) and three measures relating to pain and pain-related disability (pain severity, pain interference, and general activity levels). Age and sex were partialled out (see Tables 6 and 7).

Table 6. Partial correlations between the factors in the Novy *et al.* model and other psychological and pain measures

	Negative attitude—suicide (1)		Performance difficulty (2)		Physiological manifestations (3)	
	FP	MPC	FP	MPC	FP	MPC
MPIsev	0.02	0.06	0.06	0.05	0.14	0.13
MPIint	0.09	0.06	**0.19***	**0.18***	**0.24****	0.06
MPIgen	0.01	−0.01	−0.14	**−0.30*****	−0.10	**−0.18***
Trait anxiety	**0.52*****	**0.53*****	**0.41*****	**0.26*****	**0.32*****	0.07
PCQhope	0.08	**0.36*****	−0.03	**0.22****	−0.06	**0.18***
PCQhelp	**0.38*****	**0.27*****	0.10	0.01	0.01	0.09

Note. FP = facial pain group; MPC = pain clinic group; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

As can be seen from Tables 6 and 7, each measurement model contains one scale that correlates with other measures of negative affect and depression-related cognitions but not with pain in both pain groups ('Negative attitude—suicide' and 'Self-reproach'). The finding of Williams and Richardson that one factor containing somatic items appeared to relate to physical performance measures and not other measures of negative affect or depression-related pain cognitions was not replicated in this study. Where there were significant partial correlations between measures of pain and factors containing somatic items ('Performance difficulty' and 'Physiological manifestations' from the Novy *et al.* model and 'Somatic disturbance' from the Williams and Richardson model), these factors also correlated with other measures of mood.

Table 7. Partial correlations between the factors in the Williams and Richardson model and other psychological and pain measures

	Sadness about health (1)		Self-reproach (2)		Somatic disturbance (3)	
	FP	MPC	FP	MPC	FP	MPC
MPIsev	0.10	0.06	−0.10	0.08	**0.17***	0.09
MPIint	0.08	0.12	0.05	0.02	**0.28*****	0.12
MPIgen	−0.06	**−0.23****	0.10	0.04	−0.10	**−0.21****
Trait anxiety	**0.56*****	**0.51*****	**0.43*****	**0.40*****	**0.32*****	−0.01
PCQhope	0.05	**0.38*****	0.02	**0.23****	−0.04	**0.26*****
PCQhelp	**0.19***	0.10	**0.45*****	**0.27*****	0.03	**0.16***

Note. FP = facial pain group; MPC = pain clinic group; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Discussion

Novy *et al.*'s measurement model fitted data from both pain groups better than the original Williams and Richardson measurement model (with uncorrelated factors). However, the models were broadly equivalent when the constraint of orthogonal factors in the Williams and Richardson model was relaxed. In addition, both models adequately fit data from the two different pain samples, suggesting that the factor structure of the BDI is the same for facial pain sufferers as for other pain groups, despite the former's lower levels of pain-related disability. It is noteworthy, though, that the Novy *et al.* model was a marginally better fit in the facial pain sample than the pain clinic sample, and the Williams and Richardson model was a better fit in the pain clinic sample than the facial pain sample. This pattern of results probably reflects the fact that the pain clinic sample were more similar to the sample from which the original Williams and Richardson (1993) model was derived.

The results of the CFAs suggest that either model can be taken as a basis for measuring different elements of depression. However, there are potential limitations in both.

The strength of the Williams and Richardson model lies in the fact that it was derived from a chronic pain sample (in contrast to the model tested by Novy *et al.*). However, it was done so using orthogonal rotation, and, as noted above, it was only when the factors were allowed to intercorrelate that the fit indices became acceptable. Allowing the factors to correlate potentially questions the empirical basis of the model, as it is unclear whether these factors would have emerged from the EFA had an oblique rotation been used. While Williams and Richardson did attempt an oblique rotation, the factor loadings were not reported.

The model used by Novy *et al.* has the advantage of being a robust model that has emerged across different samples, using oblique rotation. It has also been found to fit a pain sample. However, it has been derived from non-pain samples, and it may not necessarily be an optimum model of BDI responses in a chronic pain group.

Given the broadly equivalent levels of fits, to what extent do these two models offer substantially different factors? Inspection of the partial correlations between factors and other measures of cognitive, affective, and physical functioning shows that both models contain factors that relate to negative affect but not to measures of pain, activity, and

pain-related interference ('Negative attitude—suicide' and 'Self-reproach') and that both models also contain factors dominated by somatic content ('Physiological manifestations' and 'Somatic disturbance'). It is noteworthy that the Williams and Richardson measurement model contains two factors that broadly correspond to those identified by Beck and Lester (1973) as robust features of the BDI ('Self-reproach' and 'Somatic disturbance', respectively). Two key differences, however, between the two models are: 1) that the 'Somatic disturbance' factor contains all the items that Beck and Lester (1973) identify as definitive of the two factors, 'Performance difficulty' and 'Physiological manifestations'; and 2) the factor in the Williams and Richardson model 'Sadness about health' represents a departure from other empirical findings and from the factors laid out by Beck and Lester (1973). The latter factor had high levels of internal reliability in both pain groups, despite the fact that it emerged from an EFA that included few facial pain sufferers. Whether this represents a useful construct for understanding depression in chronic pain cannot be addressed in the current study. CFA requires an *a priori* model, whether that is based on previous empirical or theoretical work. As noted by Williams (1998) in her discussion of depression and chronic pain, there has been a lack of specific theorising about depression in pain. The model tested by Novy *et al.* was based on the theoretical model of depression developed by Beck, whereby depression was conceived as having three components: physiological, cognitive, and motivational. Furthermore, the item content of the BDI was developed from clinical work with depressed people. The marginally superior fit of the Novy model probably reflects the closeness of the match between the development of the items of the BDI and Beck's three-component model noted above. The actual item content of a depression scale may differ if similar qualitative empirical work was done in the area of depression in chronic pain. However, little of this type of work has been conducted. It is conceivable that there may be different elements to the experience of depression in pain groups, and 'Sadness about health' may be one of these.

In summary, the results of this study indicate that either of the two factors 'Negative attitude—suicide' and 'Self-reproach' could be used as a measure of depression in chronic pain groups. Both have high levels of internal reliability (though the Novy *et al.* factor is slightly higher), and both correlate uniquely with other measures of negative affect but not measures of pain. However, while both measurement models offer a factor containing predominantly somatic content, the 'Somatic disturbance' factor from the Williams and Richardson model has higher levels of internal reliability, suggesting that it may form a better somatic measure. It is noteworthy that, in contrast to the original findings regarding this scale, there was no evidence in the present study that it was a construct related to physical performance but unrelated to other measures of negative mood. Hence, the finding of somatic item endorsement, which is independent of mood, may be one that only emerges in pain samples with relatively high levels of pain-related disability, and not in other pain groups.

Where does this leave recommendations for scoring the BDI in chronic pain populations? As noted earlier, there has been sufficient concern about somatic items being endorsed because of pain and not mood to caution against using a total BDI score. However, the high correlations between the factors in the facial pain group suggest that this may not be an issue in this sample, provided the pain is restricted to the face, head and/or neck (as those reporting pain elsewhere in addition to facial pain were not

included in the analysis). However, for the pain clinic sample, the correlations between factors containing affective items and those containing somatic items were relatively low. While some researchers have recommended that somatic items be scored and looked at alongside scores on affective and cognitive items separately (Novy *et al.*, 1995; Williams & Richardson, 1993), others have recommended using a higher cut-off point (i.e. 21) by which to classify chronic pain sufferers as depressed (Geisser, Roth, & Robinson, 1997). This latter research also showed that including somatic items does not affect the BDI's ability to discriminate depressed chronic pain patients from non-depressed patients when patients were classified as depressed or not depressed on the basis of interviews with clinical psychologists. However, the criteria used to diagnose depression in the latter study included somatic items, such as weight loss. This result may therefore reflect the similarity of the items of the BDI and the classification criteria used by the psychologists in the diagnostic interview.

There seems little doubt that certain items on the BDI are likely to relate to chronic pain, such as weight loss, sleep disturbance, and work inhibition (Love, 1987; Wesley *et al.*, 1999) and that these may confound the measurement of depression. Hence, if there is particular concern that the presence of pain may be affecting somatic item endorsement, and the aforementioned items in particular, one of the two subscales relating to mood ('Negative attitude—suicide' or 'Self-reproach') could be used rather than the total score.

Acknowledgements

This research was supported by grants from the Department of Health (a placebo-controlled trial comparing anti-depressive medication and cognitive–behavioural therapy in the treatment of chronic facial pain) and the Medical Research Council (a comparison of group vs. individual cognitive–behavioural pain management).

References

Beck, A. T. (1970). *Depression: Causes and treatment.* Philadelphia, PA: University of Pennsylvania Press.

Beck, A. T., & Lester, D. (1973). Components of depression in attempted suicide. *The Journal of Psychology, 85*, 257–260.

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

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

Boston, K., Pearce, S., & Richardson, P. H. (1990). The Pain Cognitions Questionnaire. *Journal of Psychosomatic Research, 34*, 103–109.

De Gagne, T. A., Mikail, S. F., & D'Eon, J. L. (1995). Confirmatory factor analysis of a 4-factor model of chronic pain evaluation. *Pain, 60*, 195–202.

Geisser, M. E., Roth, R. S., & Robinson, M. E. (1997). Assessing depression among persons with chronic pain using the Centre for Epidemiological Studies-Depression Scale and the Beck Depression Inventory: A comparative analysis. *Clinical Journal of Pain, 13*, 163–170.

Haegerstam, G., & Allerbring, M. (1995). Lack of disability in patients with chronic orofacial pain: A retrospective study. *Acta Odontology Scandinavia, 53*, 345–348.

Hayduk, L. A. (1996). *LISREL: Issues, debates and strategies.* Baltimore, MD: John Hopkins University Press.

Holzberg, A. D., Robinson, M. E., Geisser, M. E., & Gremillion, H. A. (1996). The effects of depression and chronic pain on psychosocial and physical functioning. *Clinical Journal of Pain, 12*, 118–125.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit-indexes in covariance structure analysis: conventional criteria vs. new alternatives. *Structural Equation Modeling, 6*, 1–55.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7: User's reference guide*. Chicago SSI Scientific Software International.

Kathol, R. G., Noyes, R., Williams, J., Mutgi, A., Carroll, B., & Perry, P. (1990). Diagnosing depression in patients with medical illness. *Psychosomatics*, *31*, 434–440.

Kathol, R. G., & Petty, F. (1981). Relationship of depression to medical illness. *Journal of Affective Disorders*, *3*, 111–121.

Kerns, R. D., Turk, D. C., & Rudy, T. E. (1985). The West-Haven Yale Multi-dimensional Pain Inventory (WHYMPI). *Pain*, *23*, 245–256.

Love, A. W. (1987). Depression in chronic low back pain patients: Diagnostic efficiency of three self-report questionnaires. *Journal of Clinical Psychology*, *43*, 84–89.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Beverley Hills, CA: Sage.

Novy, D. M., Nelson, D. V., Berry, L. A., & Averill, P. M. (1995). What does the Beck Depression Inventory measure in chronic pain? A reappraisal. *Pain*, *61*, 261–270.

Peterson, R. A., Mesquita, M. L., Kimmel, P. L., Simmens, S. J., Sacks, C. R., & Reiss, D. (1991). Depression, perception of illness and mortality in patients with end-stage renal disease. *International Journal of Psychiatry in Medicine*, *21*, 343–354.

Richardson, I. H., & Richardson, P. H. (1999). Does cognitive change predict the outcome of cognitive–behavioural pain management? *Psychology, Health and Medicine*, *4*, 27–44.

Romano, J. M., & Turner, J. A. (1985). Chronic pain and depression: Does the evidence support a relationship? *Psychological Bulletin*, *97*, 18–34.

Rome, H. P., Harness, D. M., & Kaplan, H. J. (1990). Psychological and behavioural aspects of chronic facial pain. In: Jacobson, A. L., & Donlon, W. C. (Eds.), *Headache and facial pain*. New York: Raven Press.

Roy, R., Thomas, M., & Matas, M. (1984). Chronic pain and depression: A review. *Comprehensive Psychiatry*, *25*, 96–105.

Rudy, T. E., Kerns, R. D., & Turk, D. C. (1988). Chronic pain and depression: Toward a cognitive–behavioural mediation model. *Pain*, *35*, 129–140.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The State-trait anxiety inventory*. Palto Alto, CA: Consulting Psychologists Press.

Tanaka, J. S., & Huba, G. J. (1984). Confirmatory hierarchical factor analyses of psychological distress measures. *Journal of Personality and Social Psychology*, *46*, 621–635.

Turp, J. C., Kowalski, C. J., & Stohler, C. S. (1998). Treatment seeking patterns of facial pain patients: Many possibilities, limited satisfaction. *Journal of Orofacial Pain*, *12*, 61–66.

Wesley, A. L., Gatchel, R. J., Garofalo, J. P., & Polatin, P. B. (1999). Toward more accurate use of the Beck Depression Inventory with chronic back pain patients. *Clinical Journal of Pain*, *15*, 117–121.

Williams, A. C. (1998). Depression in chronic pain: Mistaken models, missed opportunities. *Scandinavian Journal of Behaviour Therapy*, *27*, 61–80.

Williams, A. C., & Richardson, P. H. (1993). What does the BDI measure in chronic pain? *Pain*, *55*, 259–266.

Received 26 January 1999; revised version received 23 May 2000